

PRACTICAL WORK in Geography

PART II

TEXTBOOK FOR CLASS XII



12101



एन सी ई आर टी
NCERT

राष्ट्रीय शैक्षिक अनुसंधान और प्रशिक्षण परिषद्
NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING

First Edition

February 2007 Magha 1928

Reprinted

February 2008 Magha 1929

January 2009 Magha 1930

January 2010 Magha 1931

December 2010 Pausa 1932

January 2014 Magha 1935

February 2015 Magha 1936

December 2015 Agrahayana 1937

February 2017 Magha 1938

December 2017 Pausa 1939

December 2018 Agrahayana 1940

August 2019 Shravana 1941

January 2021 Pausa 1942

December 2021 Agrahayana 1943

PD 10T RSP© National Council of Educational
Research and Training, 2007

₹ 75.00

Printed on 80 GSM paper with NCERT
watermarkPublished at the Publication Division by the
Secretary, National Council of Educational
Research and Training, Sri Aurobindo
Marg, New Delhi 110 016 and printed at
Raj Printers, A-9, Sector B-2, Tronica City
Industrial Area, Loni, District Ghaziabad
201 102 (U.P.)**ALL RIGHTS RESERVED**

- No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the publisher.
- This book is sold subject to the condition that it shall not, by way of trade, be lent, re-sold, hired out or otherwise disposed of without the publisher's consent, in any form of binding or cover other than that in which it is published.
- The correct price of this publication is the price printed on this page. Any revised price indicated by a rubber stamp or by a sticker or by any other means is incorrect and should be unacceptable.

**OFFICES OF THE PUBLICATION
DIVISION, NCERT**NCERT Campus
Sri Aurobindo Marg
New Delhi 110 016

Phone : 011-26562708

108, 100 Feet Road
Hosdakere Halli Extension
Banashankari III Stage
Bengaluru 560 085
26725740

Phone : 080-

Navjivan Trust Building
P.O.Navjivan
Ahmedabad 380 014

Phone : 079-27541446

CWC Campus
Opp. Dhankal Bus Stop
Panihati
Kolkata 700 114

Phone : 033-25530454

CWC Complex
Maligaon
Guwahati 781 021

Phone : 0361-2674869

Publication Team

Head, Publication Division : Anup Kumar Rajput

Chief Editor : Shveta Uppal

Chief Production Officer : Arun Chitkara

Chief Business Manager : Vipin Dewan

Editor : M.G. Bhagat

Production Assistant : Prakash Veer Singh

Cover and Layout

Blue Fish

Cartography

Cartographic Design Agency

Foreword

The National Curriculum Framework (NCF), 2005, recommends that children's life at school must be linked to their life outside the school. This principle marks a departure from the legacy of bookish learning which continues to shape our system and causes a gap between the school, home and community. The syllabi and textbooks developed on the basis of NCF signify an attempt to implement this basic idea. They also attempt to discourage rote learning and the maintenance of sharp boundaries between different subject areas. We hope these measures will take us significantly further in the direction of a child-centred system of education outlined in the National Policy on Education (1986).

The success of this effort depends on the steps that school principals and teachers will take to encourage children to reflect on their own learning and to pursue imaginative activities and questions. We must recognise that, given space, time and freedom, children generate new knowledge by engaging with the information passed on to them by adults. Treating the prescribed textbook as the sole basis of examination is one of the key reasons why other resources and sites of learning are ignored. Inculcating creativity and initiative is possible if we perceive and treat children as participants in learning, not as receivers of a fixed body of knowledge.

These aims imply considerable change in school routines and mode of functioning. Flexibility in the daily time-table is as necessary as rigour in implementing the annual calendar so that the required number of teaching days are actually devoted to teaching. The methods used for teaching and evaluation will also determine how effective this textbook proves for making children's life at school a happy experience, rather than a source of stress or boredom. Syllabus designers have tried to address the problem of curricular burden by restructuring and reorienting knowledge at different stages with greater consideration for child psychology and the time available for teaching. The textbook attempts to enhance this endeavour by giving higher priority and space to opportunities for contemplation and wondering, discussion in small groups, and activities requiring hands-on experience.

The National Council of Educational Research and Training (NCERT) appreciates the hard work done by the textbook development committee responsible for this book. We wish to thank the Chairperson of the advisory committee for textbooks in Social Sciences, at the higher secondary level, Professor Hari Vasudevan and the Chief Advisor for this book, Professor M.H. Qureshi for guiding the work of this committee. Several teachers contributed to the development of this textbook; we are grateful to their principals for making this possible. We are indebted to the institutions and organisations which have

generously permitted us to draw upon their resources, material and personnel. We are especially grateful to the members of the National Monitoring Committee, appointed by the Department of Secondary and Higher Education, Ministry of Human Resource Development under the Chairpersonship of Professor Mrinal Miri and Professor G.P. Deshpande, for their valuable time and contribution. As an organisation committed to systemic reform and continuous improvement in the quality of its products, NCERT welcomes comments and suggestions which will enable us to undertake further revision and refinement.

New Delhi
20 November 2006

Director
National Council of Educational
Research and Training

© NCERT
not to be republished

Textbook Development Committee

CHAIRPERSON, ADVISORY COMMITTEE FOR TEXTBOOKS IN SOCIAL SCIENCES AT THE HIGHER SECONDARY LEVEL

Hari Vasudevan, *Professor*, Department of History, University of Calcutta, Kolkata

CHIEF ADVISOR

M. H. Qureshi, *Professor*, Centre for the Study of Regional Development, Jawaharlal Nehru University, New Delhi

ADVISOR

S. M. Rashid, *Professor*, Jamia Millia Islamia, New Delhi

MEMBERS

K. K. Sharma, *Principal (Retd.)*, Lohia College, Churu

M. H. Quasmi, *Lecturer*, IASE, Jamia Millia Islamia, New Delhi

R. N. Vyas, *Professor*, CSSH, Mohanlal Sukhadia University, Udaipur

Shahab Fazal, *Reader*, Aligarh Muslim University, Aligarh

Sucharita Sen, *Associate Professor*, CSRD, Jawaharlal Nehru University, New Delhi

MEMBER-COORDINATOR

Tannu Malik, *Lecturer*, Department of Education in Social Sciences and Humanities, NCERT, New Delhi



Nirmalya Chakraborty, College of Art, New Delhi

Acknowledgements

The National Council of Educational Research and Training acknowledges the contributions of H. Ramachandran, *Professor and Head*, Delhi School of Economics, Delhi University; B. S. Butola, *Professor*, CSRD, JNU; Odilia Coutinho, *Reader*, R.P.D. College, Belgaum; Anup Saikia, *Reader*, Gauhati University, Guwahati; Abdul Shaban, *Asstt. Professor*, Tata Institute of Social Sciences, Mumbai and Rupa Das, *PGT*, DPS, R.K. Puram, New Delhi in the development of this textbook.

Special thanks are due to Savita Sinha, *Professor and Head*, Department of Education in Social Sciences and Humanities for her valuable support at every stage of preparation of this textbook.

The Council is thankful to the Survey of India for certification of maps given in the textbook. It also gratefully acknowledges the support of individuals and organisations as listed below for providing various photographs and illustrations used in this textbook:

S.M. Rashid, *Professor*, Jamia Millia Islamia, New Delhi for fig. 1.2, 1.3 and 1.4; M.H. Quasmi, *Lecturer*, IASE, Jamia Millia Islamia, New Delhi for fig. 3.9, 3.10, 3.11 and 3.12; R.N. Vyas, *Professor*, CSSH, Mohan Lal Sukhadia University, Udaipur for fig. 5.1, 5.2 and 5.3; Odilia Coutinho, *Reader*, R.P.D. College, Belgaum for fig. 5.4 and 5.5 and Shahab Fazal, *Reader*, Aligarh Muslim University, Aligarh for fig. 6.8, 6.9, 6.10, 6.12 and 6.13.

The Council also gratefully acknowledges the contribution of Anil Sharma and Ishwar Singh *DTP Operators*; Ajay Singh, *Copy Editor*, Aarati Baloni, *Proof Reader* and Dinesh Kumar, *Computer Incharge* who have helped in giving a final shape to this book. The contribution of the Publication Department, NCERT in bringing out this textbook is also duly acknowledged.

The following are applicable to all the maps of India used in this textbook

1. © Government of India, Copyright 2006
2. The responsibility for the correctness of internal details rests with the publisher.
3. The territorial waters of India extend into the sea to a distance of twelve nautical miles measured from the appropriate base line.
4. The administrative headquarters of Chandigarh, Haryana and Punjab are at Chandigarh.
5. The interstate boundaries amongst Arunachal Pradesh, Assam and Meghalaya shown on this map are as interpreted from the “North-Eastern Areas (Reorganisation) Act.1971,” but have yet to be verified.
6. The external boundaries and coastlines of India agree with the Record/Master Copy certified by Survey of India.
7. The state boundaries between Uttaranchal and Uttar Pradesh, Bihar and Jharkhand, Chhattisgarh and Madhya Pradesh have not been verified by the Governments concerned.
8. The spellings of names in this map, have been taken from various sources.

Contents

FOREWORD	<i>iii</i>
CHAPTER 1	
Data – Its Source and Compilation	1 – 12
CHAPTER 2	
Data Processing	13 – 31
CHAPTER 3	
Graphical Representation of Data	32– 54
CHAPTER 4	
Use of Computer in Data Processing and Mapping	55 – 70
CHAPTER 5	
Field Surveys	71 – 84
CHAPTER 6	
Spatial Information Technology	85 – 100
ANNEXURE	101 – 105
GLOSSARY	106

THE CONSTITUTION OF INDIA

PREAMBLE

WE, THE PEOPLE OF INDIA, having solemnly resolved to constitute India into a ¹**[SOVEREIGN SOCIALIST SECULAR DEMOCRATIC REPUBLIC]** and to secure to all its citizens :

JUSTICE, social, economic and political;

LIBERTY of thought, expression, belief, faith and worship;

EQUALITY of status and of opportunity; and to promote among them all

FRATERNITY assuring the dignity of the individual and the ²[unity and integrity of the Nation];

IN OUR CONSTITUENT ASSEMBLY this twenty-sixth day of November, 1949 do **HEREBY ADOPT, ENACT AND GIVE TO OURSELVES THIS CONSTITUTION.**

1. Subs. by the Constitution (Forty-second Amendment) Act, 1976, Sec.2, for "Sovereign Democratic Republic" (w.e.f. 3.1.1977)
2. Subs. by the Constitution (Forty-second Amendment) Act, 1976, Sec.2, for "Unity of the Nation" (w.e.f. 3.1.1977)



12101CH01

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 interrelationship 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 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 phenomenon, 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

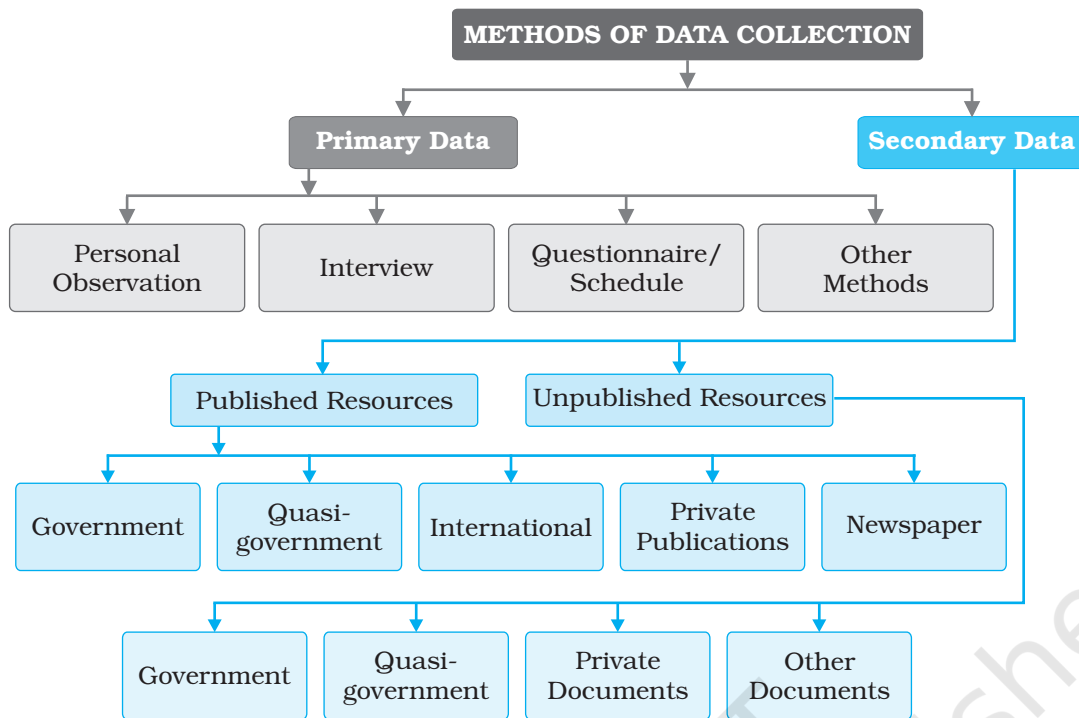


Fig. 1.1 : Methods of Data Collection

carrying 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 the 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 the 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 scientists 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 sources of secondary data.

6. Electronic Media

The electronic media, specially the 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 *patwari* 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 residents' 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, 2011

State/ UT Code	India/State/ Union Territory	Total Population		
		Persons	Males	Females
1	2	3	4	5
	INDIA¹	1,21,05,69,573	62,31,21,843	58,74,47,730
1.	Jammu and Kashmir ²	1,25,41,302	66,40,662	59,00,640
2.	Himachal Pradesh	68,64,602	34,81,873	33,82,729
3.	Punjab	2,77,43,338	1,46,39,465	1,31,03,873
4.	Chandigarh ³	10,55,450	5,80,663	4,74,787
5.	Uttarakhand	1,00,86,292	51,37,773	49,48,519
6.	Haryana	2,53,51,462	1,34,94,734	1,18,56,728
7.	National Capital Territory of Delhi	1,67,87,941	89,87,326	78,00,615
8.	Rajasthan	6,85,48,437	3,55,50,997	3,29,97,440
9.	Uttar Pradesh	19,98,12,341	10,44,80,510	9,53,31,831
10.	Bihar	10,40,99,452	5,42,78,157	4,98,21,295

¹ inclusive of all territorial boundary of India

² excluding PoK

³ Union Territory

Source : Census, 2011

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 :

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

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:

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

$\sum q_1$ = 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	$\frac{32.5}{32.5} \times 100$	100
1980-81	42.2	$\frac{42.2}{32.5} \times 100$	130
1990-91	53.7	$\frac{53.7}{32.5} \times 100$	165
2000-01	67.4	$\frac{67.4}{32.5} \times 100$	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 1.2 : Literacy Rate : 1951 – 2011

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
2011	73.0	80.9	64.6

Source: Census, 2011

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	#####	5
20-30	22,28,26,22,28	#####	5
30-40	39,32,37,36,35,37,31	#####	7
40-50	47,46,44,43,42,41	#####	6
50-60	53,57,50,51,58, 59,56,58,56,59	#####	10
60-70	64,62,67,65, 63,64,68,64	#####	8
70-80	74,73,75,70,74,71	#####	6
80-90	89,84,84,80,83	#####	5
90-100	96,92,90,90	#####	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 number of individuals (places in the fourth column of *Table 1.5*) is 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**.

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
	$\sum f = N = 60$	

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).

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 under 100

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 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 the 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 two or more than two frequency distributions (*Fig. 1.5*). The two frequencies are shown using a bar diagram and a line graph respectively.

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 ojee). 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 subtract 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*.

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*.

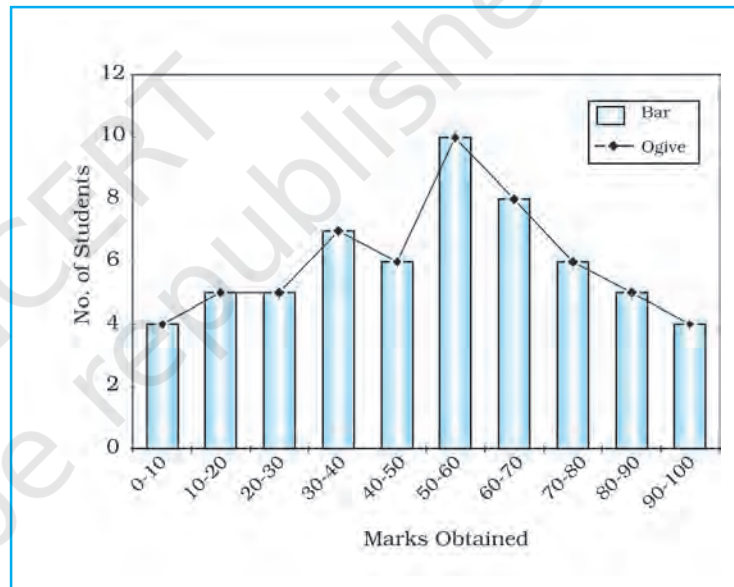


Fig. 1.5 : Frequency Distribution Polygon

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

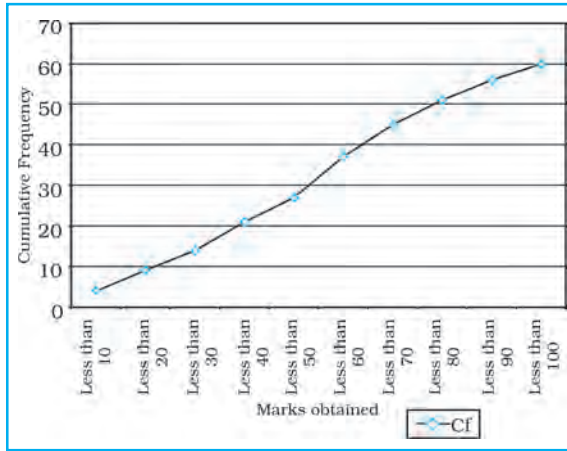


Fig. 1.6 : Less than Ogive

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

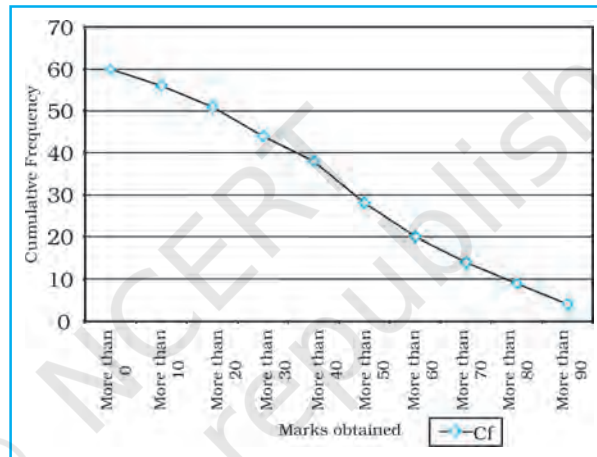


Fig. 1.7 : More than Ogive

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
40 - 50	27	38
50 - 60	37	28
60 - 70	45	20
70 - 80	51	14
80 - 90	56	9
90 - 100	60	4

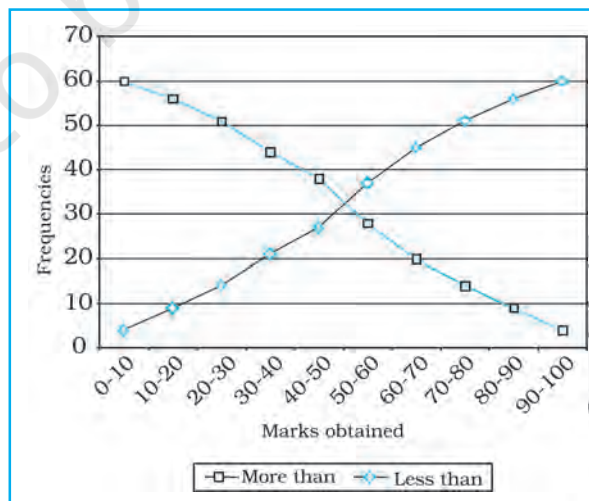


Fig. 1.8 : Less than and more than Ogive

Exercises

- Choose the right answer from the four alternatives given below:
 - A number or character which represents measurement is called
(a) Digit (b) Data (c) Number (d) Character
 - A single datum is a single measurement from the
(a) Table (b) Frequency (c) Real world (d) Information
 - 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
 - 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
 - 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
- Answer the following questions in about 30 words:
 - Differentiate between data and information.
 - What do you mean by data processing?
 - What is the advantage of foot note in a table?
 - What do you mean by primary sources of data?
 - Enumerate five sources of secondary data.
- Answer the following questions in about 125 words:
 - Discuss the national and international agencies where from secondary data may be collected.
 - What is the importance of an index number? Taking an example examine the process of calculating an index number and show the changes.

Activity

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



12101CH02

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:

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

Where,

\bar{X} = Mean

\sum = Sum of a series of measures

x = A raw score in a series of measures

$\sum x$ = The sum of all the measures

N = Number of measures

Table 2.1 : Calculation of Mean Rainfall

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

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*:

* 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:

$$\begin{aligned}\bar{X} &= \frac{\sum x}{N} \\ &= \frac{6,484}{7} \\ &= 926.29\end{aligned}$$

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 a 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:

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

Where,

A = Subtracted constant

$\sum d$ = 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 :

$$\begin{aligned}\bar{X} &= 800 + \frac{884}{7} \\ &= 800 + \frac{884}{7} \\ \bar{X} &= 926.29 \text{ mm}\end{aligned}$$

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 $\sum fx$ that is finally divided by the number of observations i. e., N . Hence, mean is calculated using the following formula :

$$\bar{X} = \frac{\sum fx}{N}$$

Where :

\bar{X} = Mean

f = Frequencies

x = Midpoints of class intervals

N = Number of observations (it may also be defined as $\sum f$)

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	$U = \frac{(x-100)}{20}$	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
$\sum fx$ and $\sum fx$	$\sum f = 99$		$\sum fx =$ 10,160		$\sum fd =$ 260		$\sum fu =$ 13

Where $N = \sum f = 99$

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 ($\sum fx$) divided by N .

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

$$\begin{aligned}\bar{X} &= \frac{\sum fx}{N} \\ &= \frac{10,160}{99} \\ &= 102.6\end{aligned}$$

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

$$\bar{x} = A \pm \frac{\sum fd}{N}$$

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 $\sum f$

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 ($\sum fd$). Note that the sign attached to $\sum fd$ is replaced in the formula following A , where \pm is given.

The mean using indirect method is computed as under :

$$\begin{aligned}\bar{x} &= A \pm \frac{\sum fd}{N} \\ &= 100 + \frac{260}{99} \\ &= 100 + 2.6 \\ &= 102.6\end{aligned}$$

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 :

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

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 :

$$\begin{aligned} &\text{Value of } \left(\frac{N+1}{2} \right) \text{ th item} \\ &= \left(\frac{7+1}{2} \right) \text{ th item} \\ &= \left(\frac{8}{2} \right) \text{ th item} \end{aligned}$$

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,

$$M = 8,172 \text{ m}$$

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 :

$$M = l + \frac{i}{f} \left(\frac{N}{2} - c \right)$$

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	$M = \frac{N}{2}$ $= \frac{50}{2}$ $= 25$
60-70	7	10	
70-80	11	21	
80-90 (median group)	16	37	
90-100	8	45	
100-110	5	50	
	$\sum f$ or N = 50		

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 :

$$M = l + \frac{i}{f}(m - c)$$

$$\begin{aligned}
 &= 80 + \frac{10}{16} (25 - 21) \\
 &= 80 + \frac{5}{8} \times 4 \\
 &= 80 + \frac{5}{2} \\
 &= 80 + 2.5 \\
 M &= 82.5
 \end{aligned}$$

Mode

The value that occurs most frequently in a distribution is referred to as **mode**. It is symbolised as **Z** or **M_o**. 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.

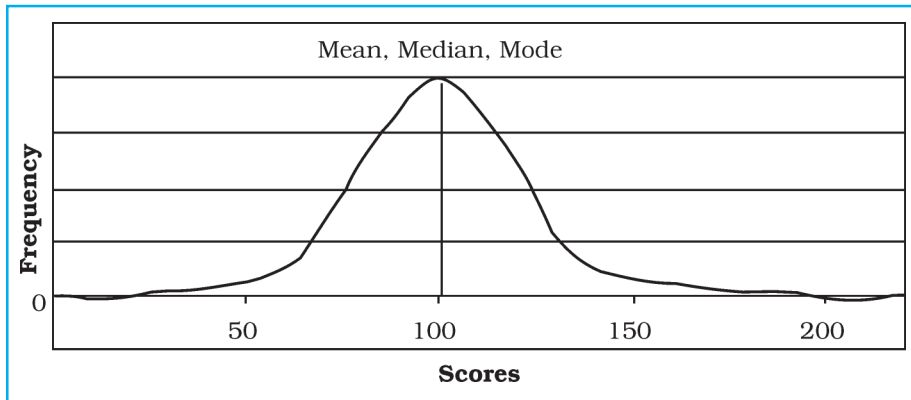


Fig. 2.3 : Normal Distribution Curve

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).

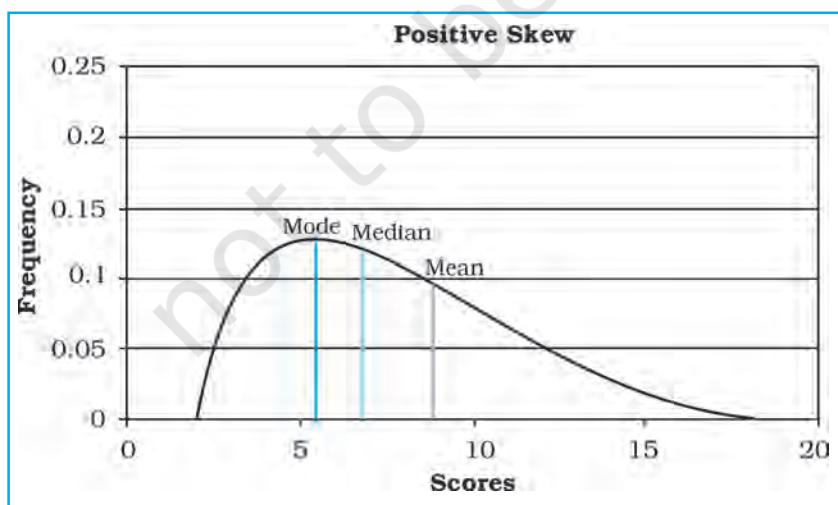


Fig. 2.4 : Positive Skew

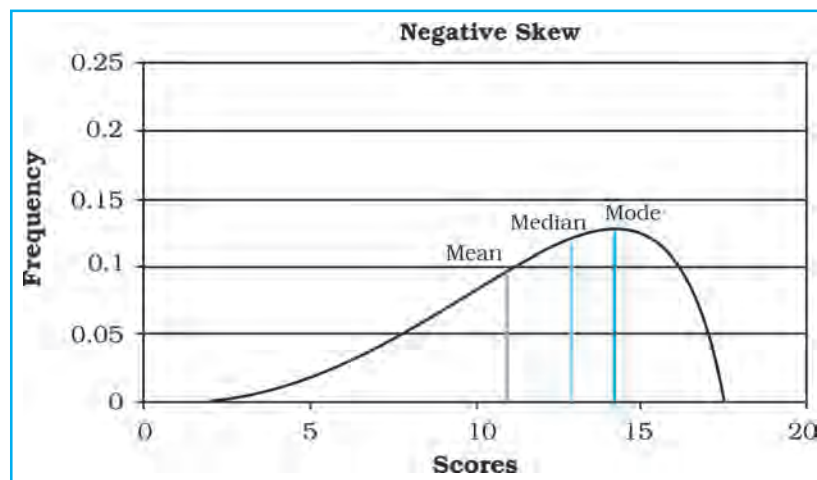


Fig. 2.5 : Negative Skew

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

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

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 Coefficient of Variation (CV)
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 Coefficient 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 :

$$R = L - S$$

Where

'R' is Range,

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

Hence,

$$\begin{aligned} \mathbf{R} &= L - S \\ &= 100 - 40 = \mathbf{60} \end{aligned}$$

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 σ denotes it.

To obtain SD, deviation of each score from the mean (\bar{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 - $\sum x^2$ (care should be taken that these are **not** summed first and then squared). This sum of the squared deviations ($\sum x^2$) 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 :

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

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

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum x^2}{N}} \\ &= \sqrt{\frac{40}{5}} \\ &= \sqrt{8} = 2.828 \quad \sqrt{8} \quad 2.828 \\ &\mathbf{2.83}\end{aligned}$$

Table 2.7 : Computation of Standard Deviation

X	$x(X - \bar{X})$	x^2
1	-4	16
3	-2/-6	4
5	0	0
7	2	4
9	6-Apr	16
$\sum X = 25$		
$N = 5$		
$\therefore = 5$		

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 x^2 s 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
<i>f</i>	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) <i>f</i>	(3) <i>x'</i>	(4) <i>fx'</i>	(5) <i>fx'²</i>
120 - 130	2	-3	-6	18
130 - 140	4	-2	-8	16
140 - 150	6	-1	-6	6
150 - 160	12	0	0	0
160 - 170	10	1	10	10
170 - 180	6	2	12	24
	N=40		$\sum fx' = 2$	$\sum fx'^2 = 74$

The following formula is used to calculate the Standard Deviation :

$$SD = \sqrt{\frac{\sum fx'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

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 :

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

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

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

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

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

$$CV = \mathbf{56\%}$$

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 a 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.

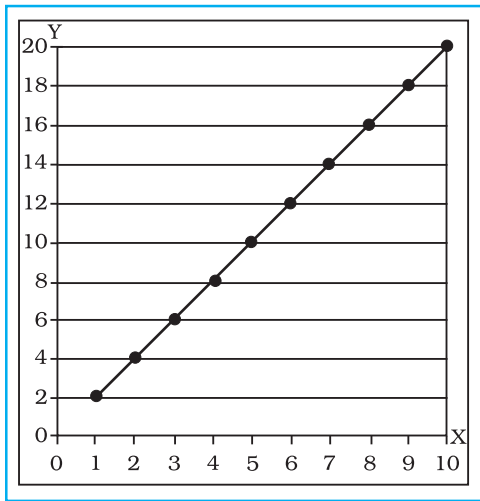


Fig. 2.6 : Perfect Positive Correlation

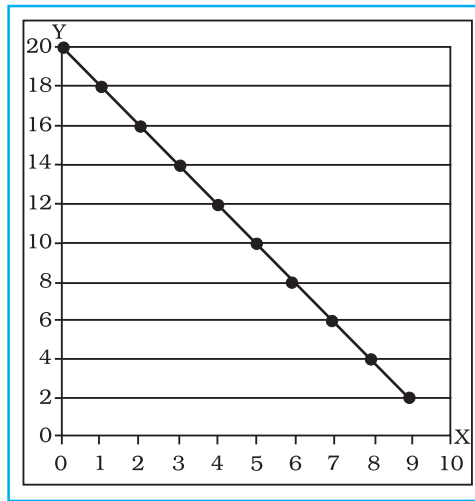


Fig. 2.7 : Perfect Negative Correlation

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 up to **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.

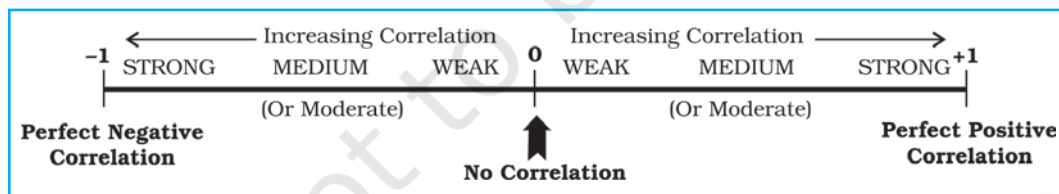


Fig. 2.8 : Spread of Direction and Degree of Correlation

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. Scatter plot A shows no correlation when Y does not respond to changes in X. Similarly, zero correlation occurs in Scatter plot B when X does not respond to changes in Y.

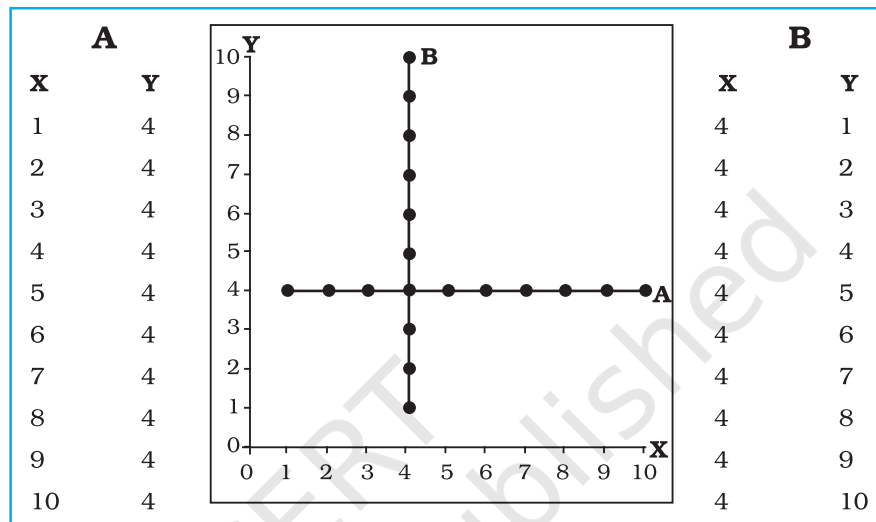


Fig. 2.9 : Scatter plot showing No Correlation

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).

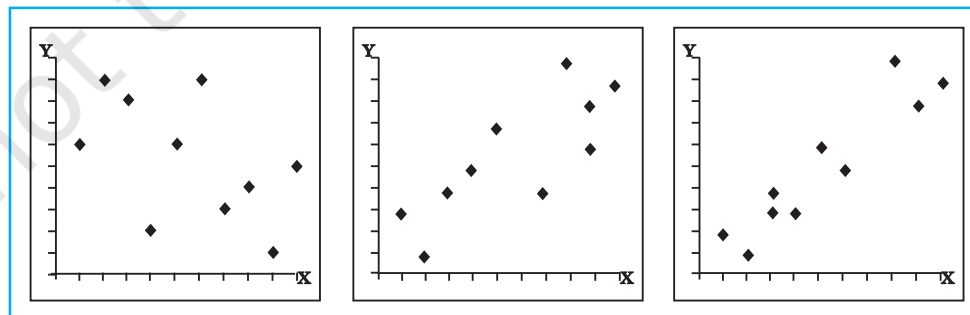


Fig. 2.10 : Weak Negative Correlation

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:

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Where,

ρ = rank correlation

$\sum D^2$ = 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

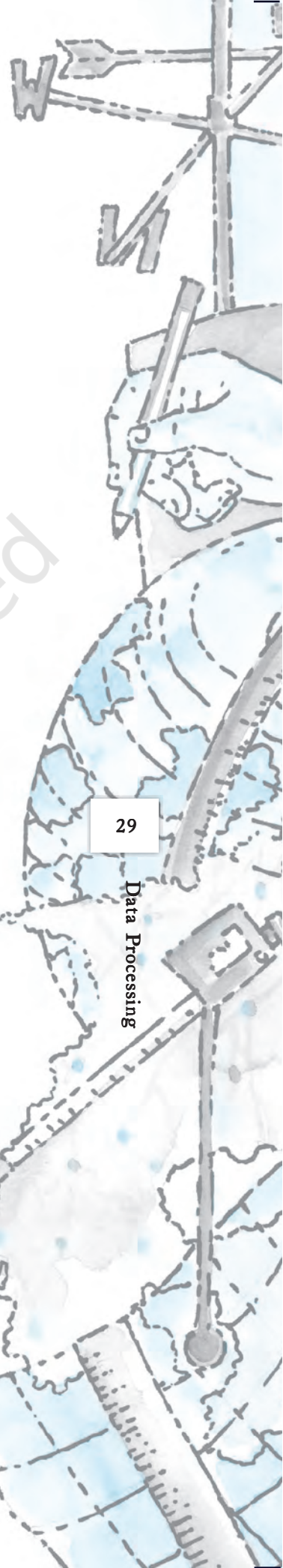


Table 2.8 : Computation of Spearman's Rank Correlation

(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

$$\begin{aligned}
 \rho &= 1 - \frac{6 \sum D^2}{N(N^2 - 1)} \\
 &= 1 - \frac{6 \times 8}{10(10^2 - 1)} \\
 &= 1 - \frac{48}{10(100 - 1)} \\
 &= 1 - \frac{48}{10(99)} \\
 &= 1 - \frac{48}{(990)} \\
 &= 1 - 0.05 \\
 &= 0.95
 \end{aligned}$$

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.

Exercises

- Choose the correct answer from the four alternatives given below:
 - The measure of central tendency that does not get affected by extreme values:
 - Mean
 - Mean and Mode
 - Mode
 - Median
 - The measure of central tendency always coinciding with the hump of any distribution is:
 - Median
 - Median and Mode
 - Mean
 - Mode
 - A scatter plot represents negative correlation if the plotted values run from:
 - Upper left to lower right
 - Lower left to upper right
 - Left to right
 - Upper right to lower left
- Answer the following questions in about 30 words:
 - Define the mean.
 - What are the advantages of using mode ?
 - What is dispersion ?
 - Define correlation.
 - What is perfect correlation ?
 - What is the maximum extent of correlation?
- Answer the following questions in about 125 words:
 - Explain relative positions of mean, median and mode in a normal distribution and skewed distribution with the help of diagrams.
 - Comment on the applicability of mean, median and mode (*hint: from their merits and demerits*).
 - Explain the process of computing Standard Deviation with the help of an imaginary example.
 - Which measure of dispersion is the most unstable statistic and why?
 - Write a detailed note on the degree of correlation.
 - What are various steps for the calculation of rank order correlation?

Activity

- Take an imaginary example applicable to geographical analysis and explain direct and indirect methods of calculating mean from ungrouped data.
- 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 °C) on Y axis.
- (c) Choose an appropriate scale and label it on Y-axis. If the data involve a negative figure, then the selected scale should also show it as shown in *Fig. 3.1*.

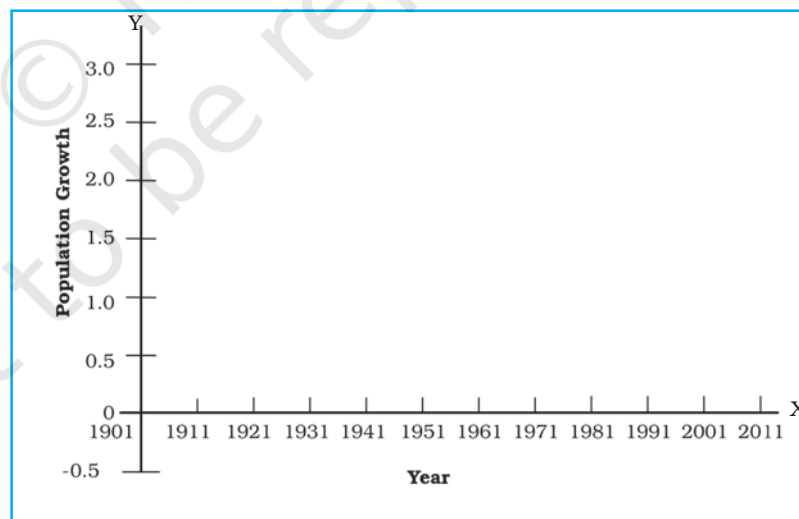


Fig. 3.1 : Construction of a Line Graph

- (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 2011

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
2011	1.79

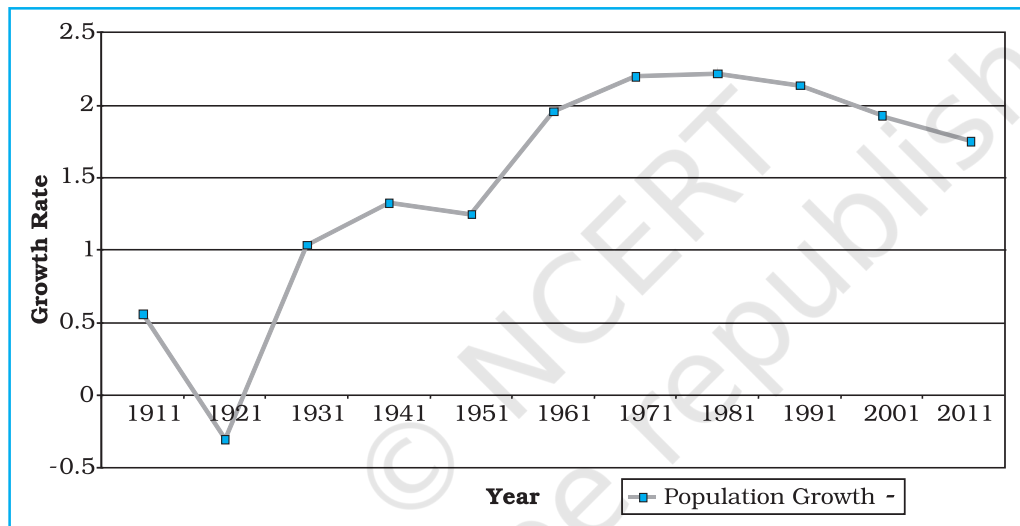


Fig. 3.2 : Annual Growth of Population in India 1901-2011

Activity

Find out the reasons for sudden change in population between 1911 and 1921 as 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 – 1961-2011

States/UT	1961	1971	1981	1991	2001	2011
Delhi	785	801	808	827	821	866
Haryana	868	867	870	860	846	877
Uttar Pradesh	907	876	882	876	898	908

Source : Census, 2011

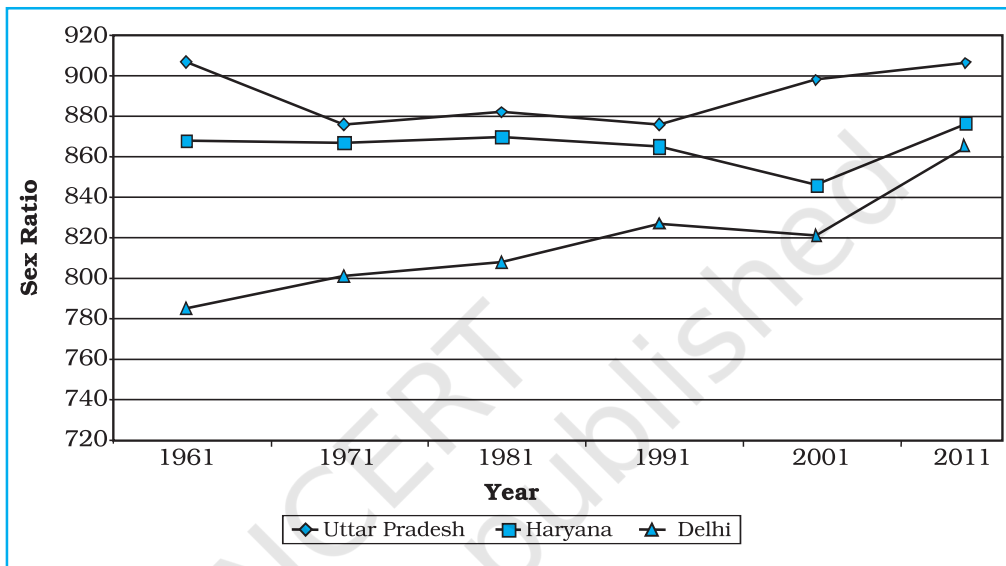


Fig. 3.3 : Sex-Ratio of Selected States 1961-2011

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:

- The width of all the bars or columns should be similar.
- All the bars should be placed on equal intervals/distance.
- 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 Thiruvananthapuram as given in Table 3.3 :

Table 3.3 : Average Monthly Rainfall of Thiruvananthapuram

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.

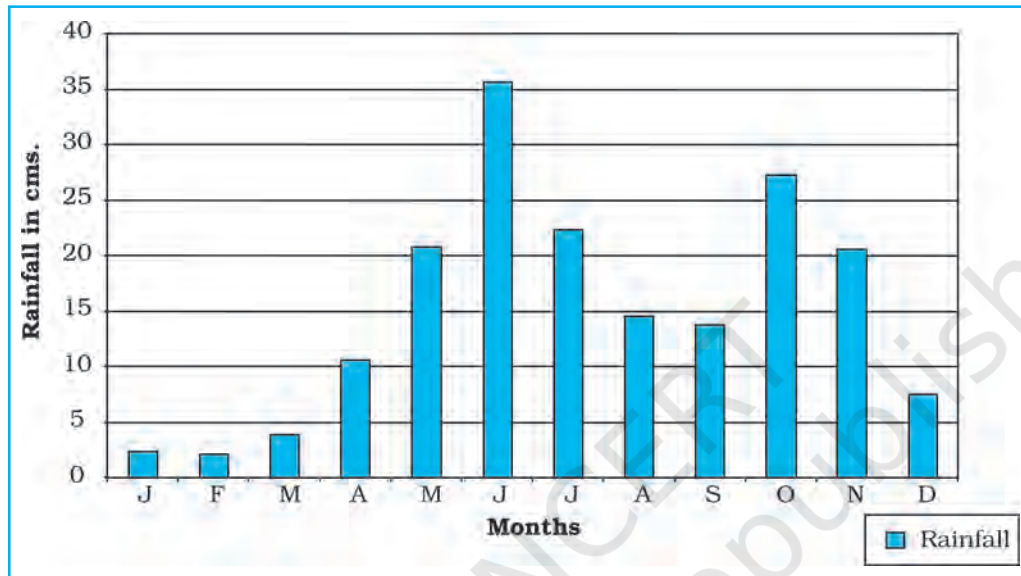


Fig. 3.4 : Average Monthly Rainfall of Thiruvananthapuram

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.

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

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 :

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 bar diagram as shown in Fig. 3.5.

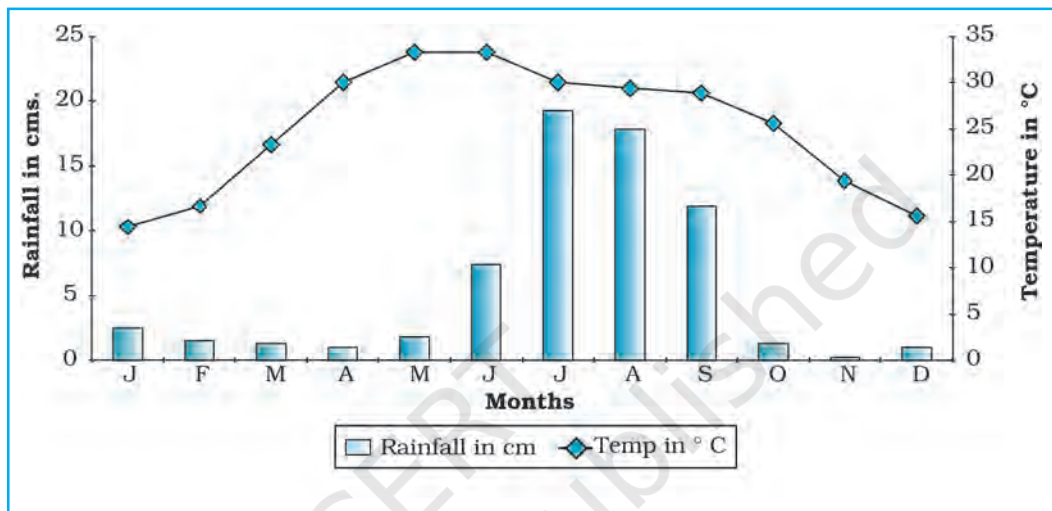


Fig. 3.5 : Temperature and Rainfall in Delhi

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 1951–2011 as given in Table 3.5 :

Table 3.5 : Literacy Rate in India, 1951–2011 (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
2011	73.0	80.9	64.6

Source : Census, 2011

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).

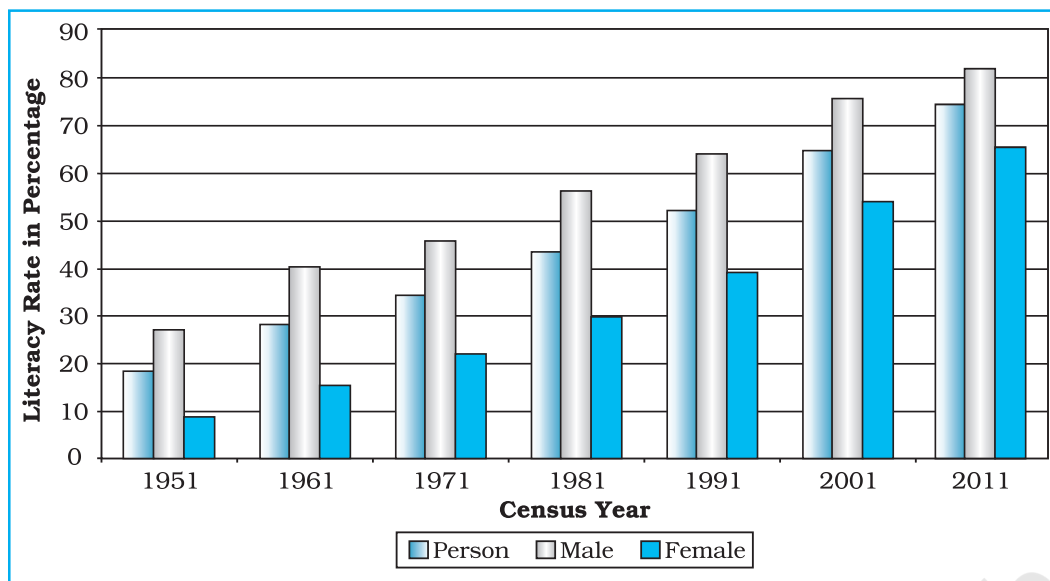


Fig. 3.6 : Literacy Rate in India, 1951-2011

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
2008-09	616.2	110.1	14.9	741.2
2009-10	677.1	104.1	18.6	799.8
2010-11	704.3	114.2	26.3	844.8

Source: Economic Survey, 2011-12

Construction

- Arrange the data in ascending or descending order.
- 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 **Divided Circle Diagram**.

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

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

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

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

For example, a pie diagram may be drawn to show the 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.

Calculation of Angles

- Arrange the data on percentages of Indian exports in an ascending order.
- 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. $360/100$.

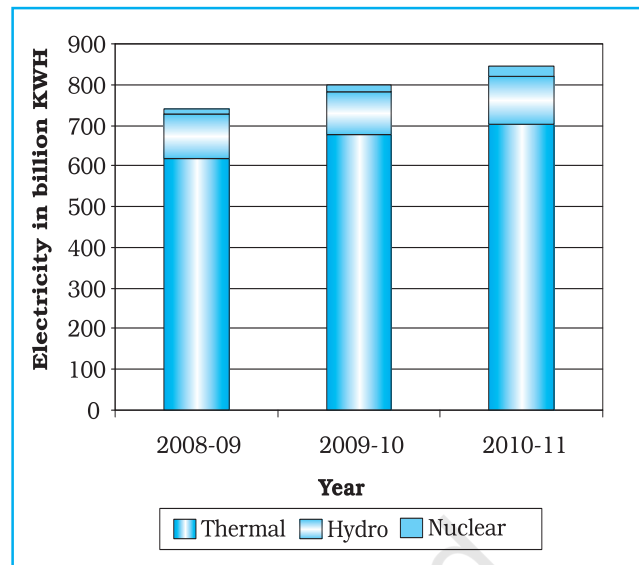


Fig. 3.7 : Gross Electricity Generation in India

Table 3.7 (a) : India's Export to Major Regions of the World in 2010-11

Unit/Region	% of Indian Export
Europe	20.2
Africa	6.5
America	14.8
Asia and ASEAN	56.2
Others	2.3
Total	100

Source : *Economic Survey 2011-12*

- (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 2010-11

Countries	%	Calculation	Degree
Europe	20.2	$20.2 \times 3.6 = 72.72$	73°
Africa	6.5	$6.5 \times 3.6 = 23.4$	23°
America	14.8	$14.8 \times 3.6 = 53.28$	53°
Asia and ASEAN	56.2	$56.2 \times 3.6 = 202.32$	203°
Others	2.3	$2.3 \times 3.6 = 8.28$	8°
Total	100		360°

Construction

- 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.
- Draw a line from the centre of the circle to the arc as a radius.
- Measure the angles from the arc of the circle for each category of vehicles in an ascending order clock-wise, starting with smaller angle.
- 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

- The circle should neither be too big to fit in the space nor too small to be illegible.
- Starting with bigger angle will lead to accumulation of error leading to the plot of the smaller angle difficult.

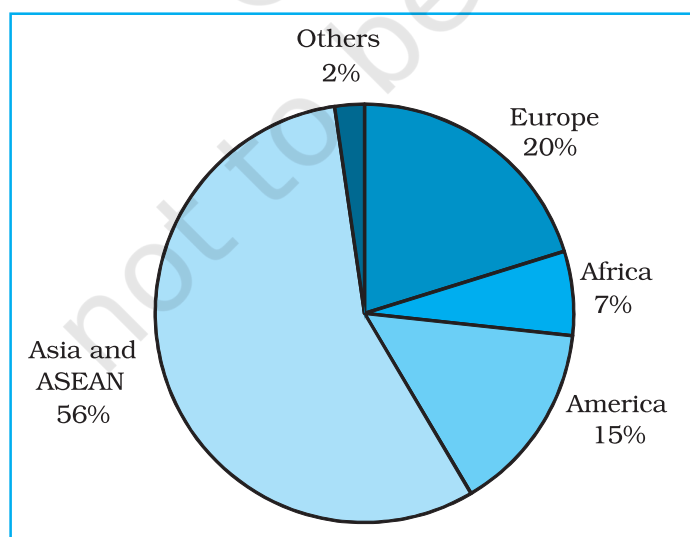


Fig. 3.8 : Direction of Indian Exports 2010-11

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 **Dynamic Map**. Transport map, which shows the 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.

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

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*.

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.

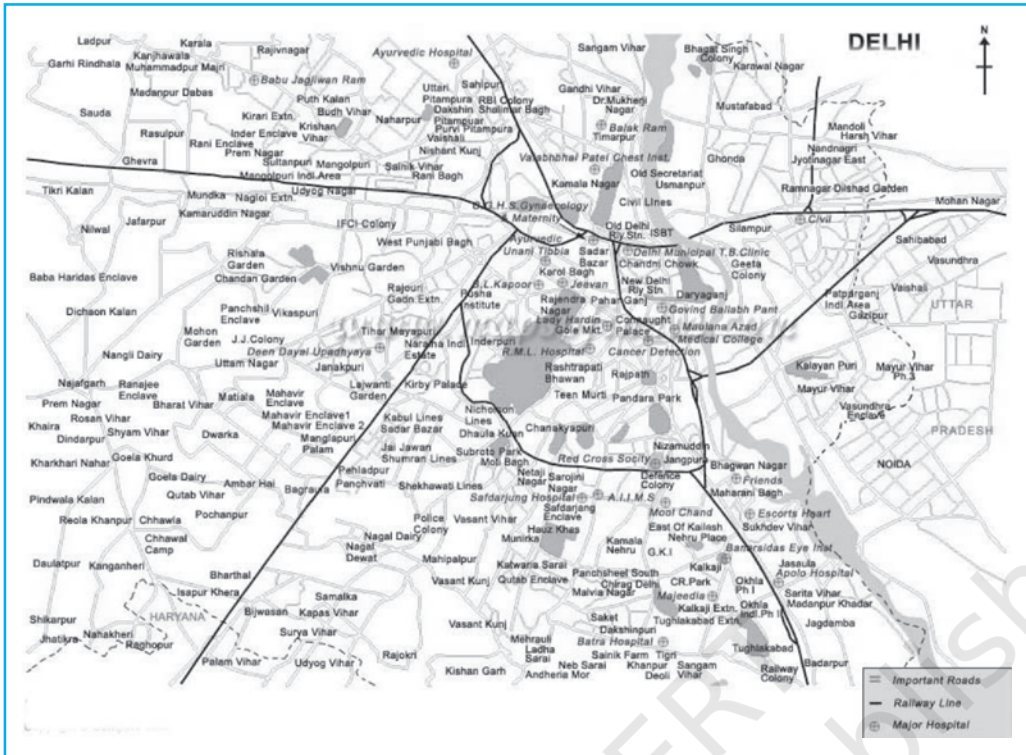


Fig. 3.9 : Map of Delhi

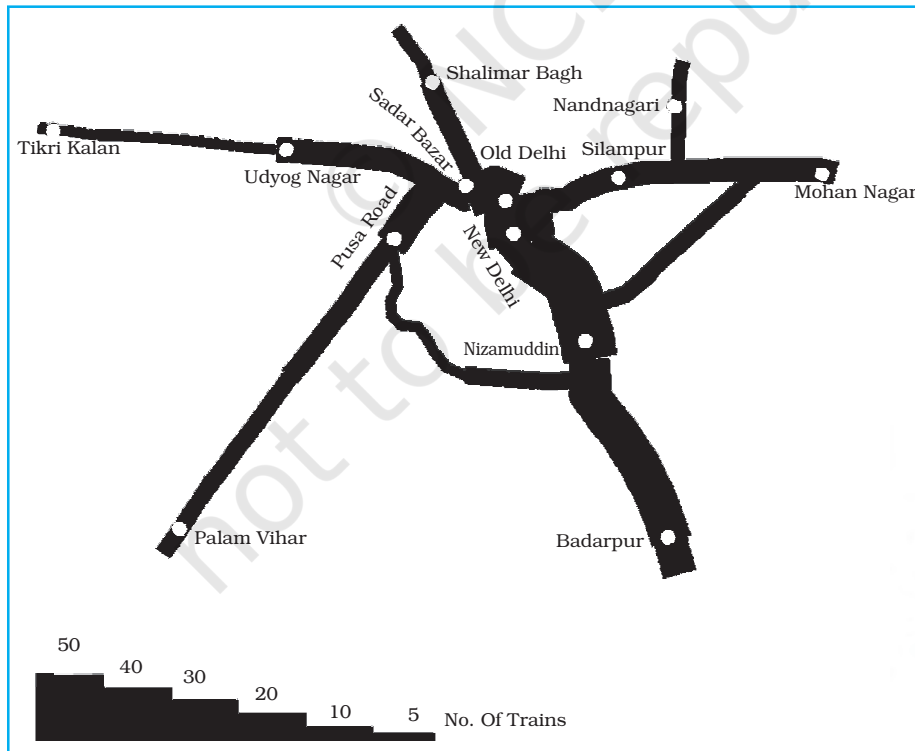


Fig. 3.10 : Traffic (Railway) Flow Map of Delhi

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

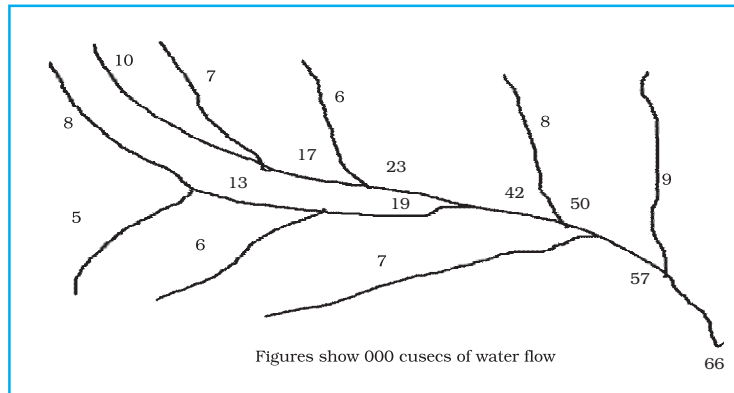


Fig. 3.11 : Ganga Basin

Construction

- Take a scale as a strip of 1cm width = 50,000 cusecs of water.
- Make the diagram as shown in Fig. 3.12.

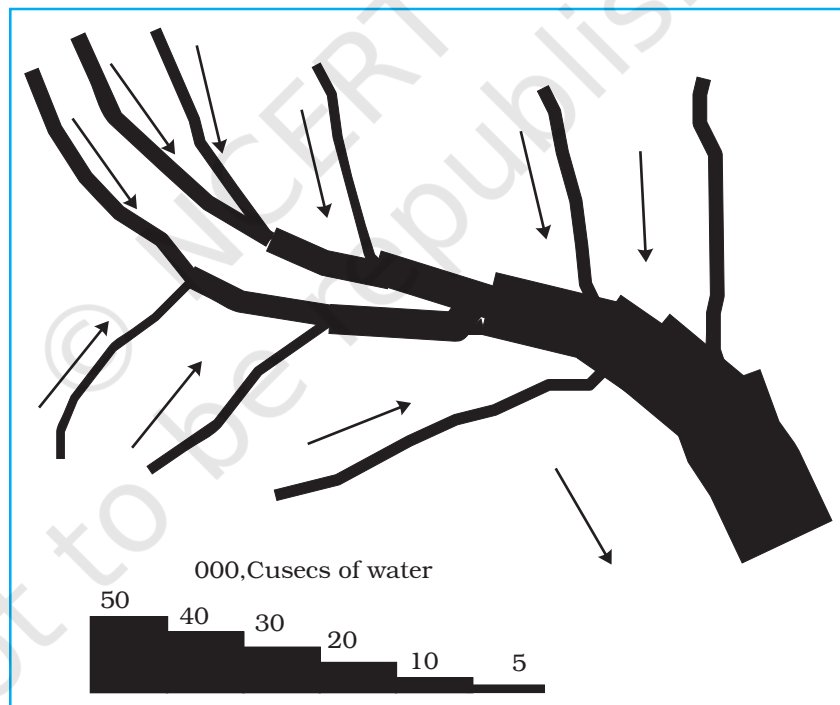


Fig. 3.12 : Construction of a Flow Map

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 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 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 :

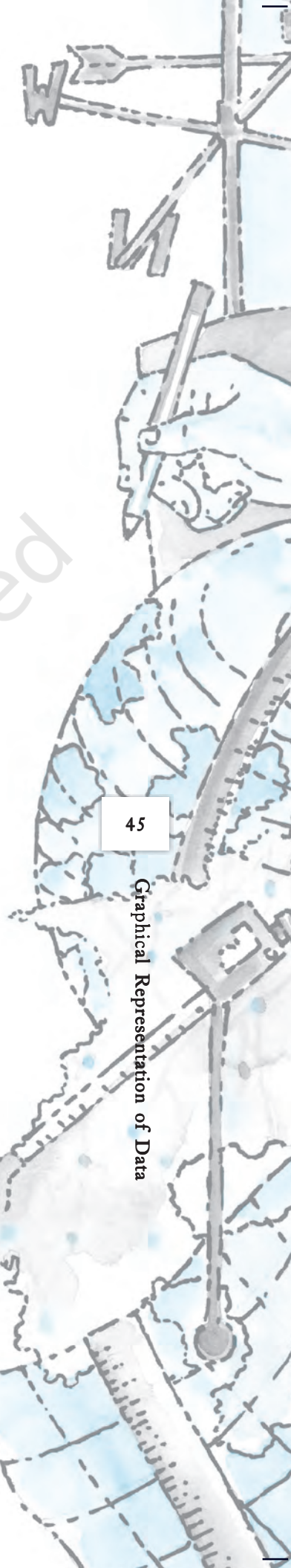
- (a) Dot maps
- (b) Choropleth maps
- (c) Isopleth maps

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 of 2001 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.	Uttarakhand	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.	Odisha	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

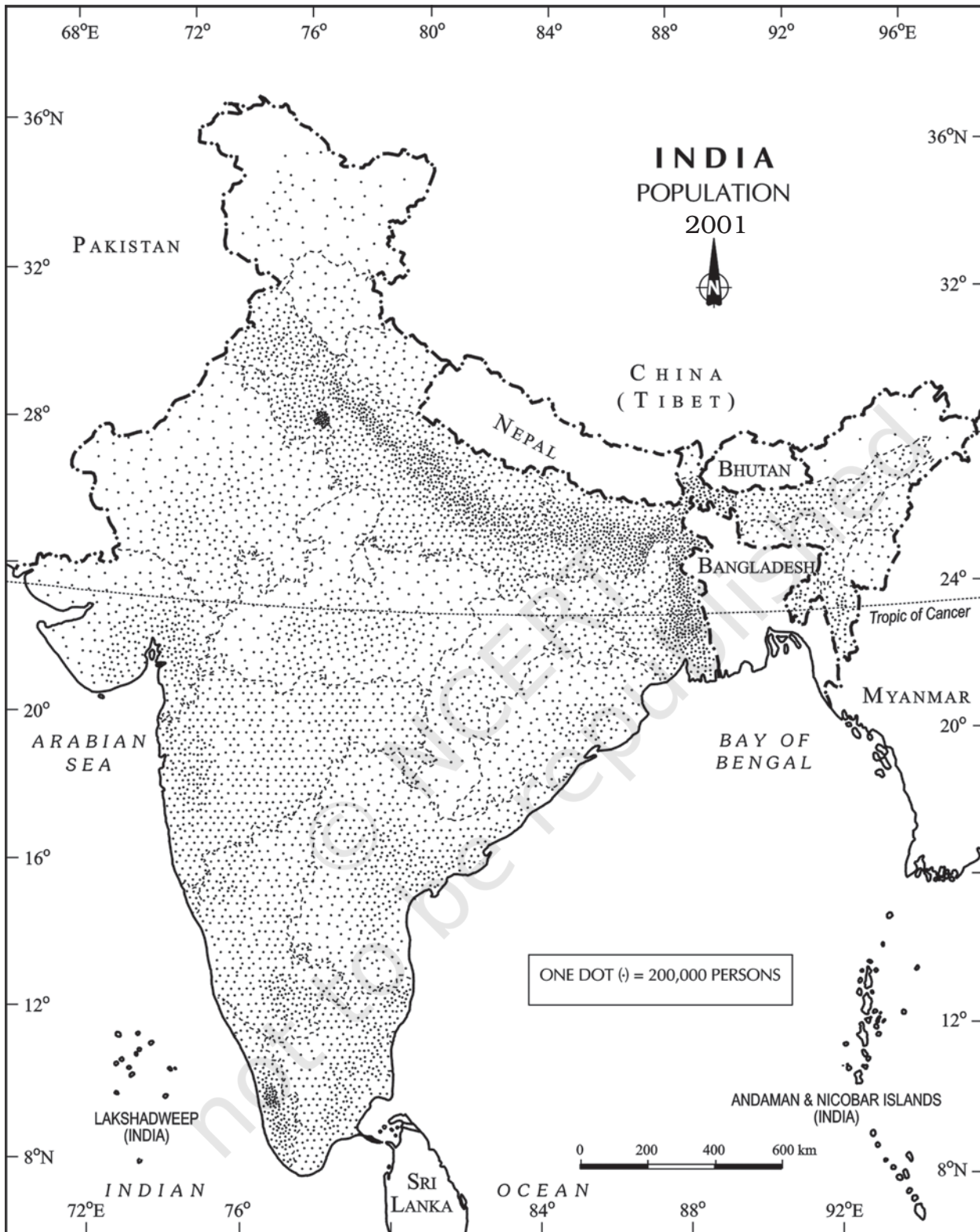


Fig. 3.13 : Population of India, 2001

Construction

- Select the size and value of a dot.
- 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.
- Place the dots in each state as per the determined number in all states.
- 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 map of the area depicting different administrative units.
- Appropriate statistical data according to administrative units.

Steps to be followed

- Arrange the data in ascending or descending order.
- Group the data into 5 categories to represent very high, high, medium, low and very low concentrations.
- The interval between the categories may be identified on the following formulae i.e., $\text{Range}/5$ and $\text{Range} = \text{maximum value} - \text{minimum value}$.
- 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 in 2001 as given in Table 3.10.

Construction

- Arrange the data in ascending order as shown above.
- Identify the range within the data. In the present case, the states recording the lowest and highest literacy rates are Bihar (47%) and Kerala (90.9%), respectively. Hence, the range would be $91.0 - 47.0 = 44.0$
- 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., to 9.0
- Determine the number of the categories alongwith the 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, Odisha, Assam, Madhya Pradesh, Chhattisgarh)

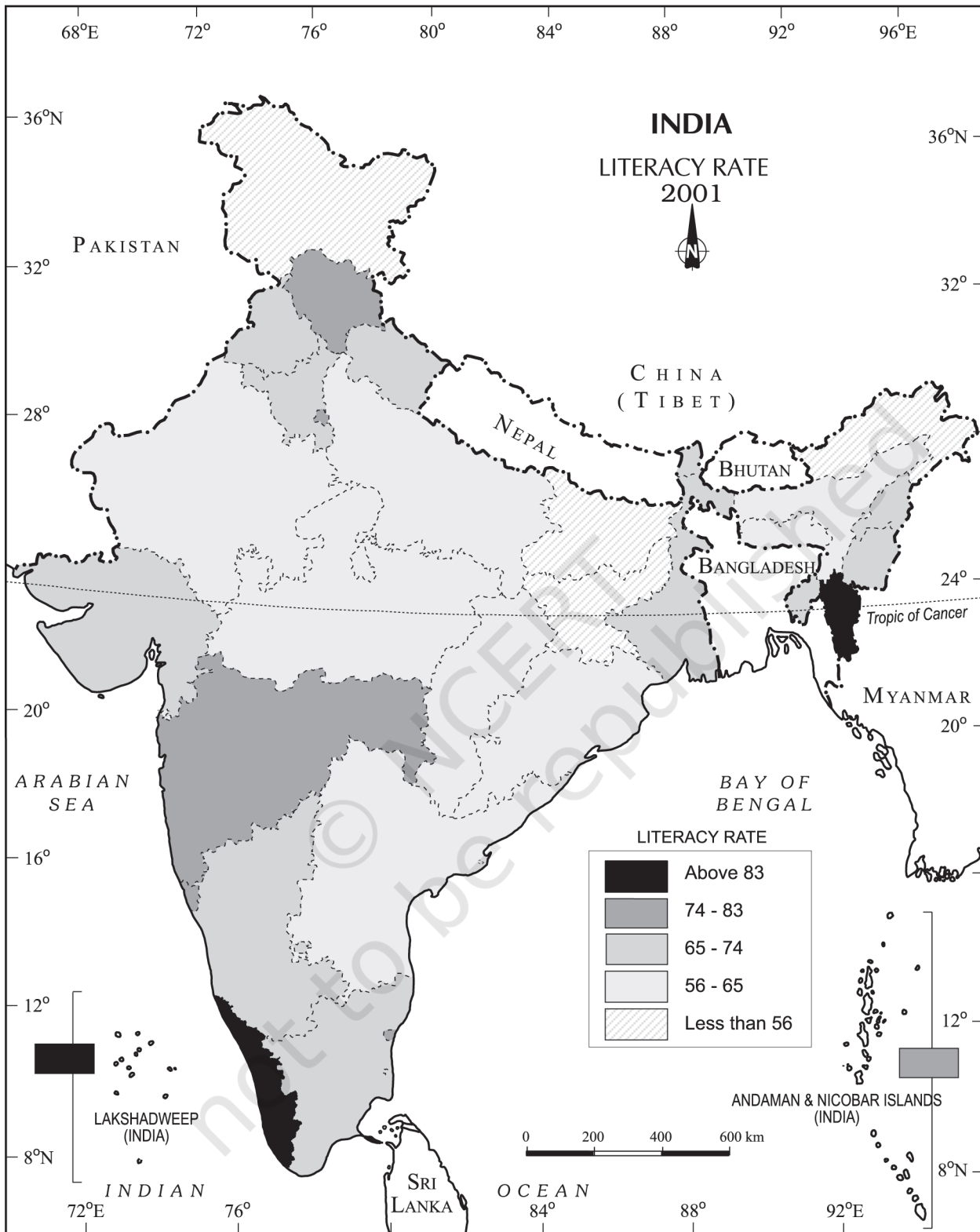


Fig. 3.14 : Literacy Rate, 2001

Table 3.10 : Literacy Rate in India, 2001

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

50

- 65 – 74 Medium (Nagaland, Karnataka, Haryana, West Bengal, Sikkim, Gujarat, Punjab, Manipur, Uttarakhand, 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), Isoneph (equal cloudiness), Isohels (equal sunshine), contours (equal heights), Isobaths (equal depths), Isohaline (equal salinity), etc.

Requirement

- Base line map depicting point location of different places.
- Appropriate data of temperature, pressure, rainfall, etc. over a definite period of time.
- Drawing instrument specially French Curve, etc.

Rules to be observed

- An equal interval of values be selected.
- Interval of 5, 10, or 20 is supposed to be ideal.
- 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:

- Firstly, determine the minimum and maximum values given on the map.
- Calculate the range of value i.e. Range = maximum value – minimum value.
- 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.

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

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 28 °C and 33 °C and you want to draw 30 °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 30 °C will be plotted 4mm away from 28 °C or 6mm ahead of 33 °C.

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

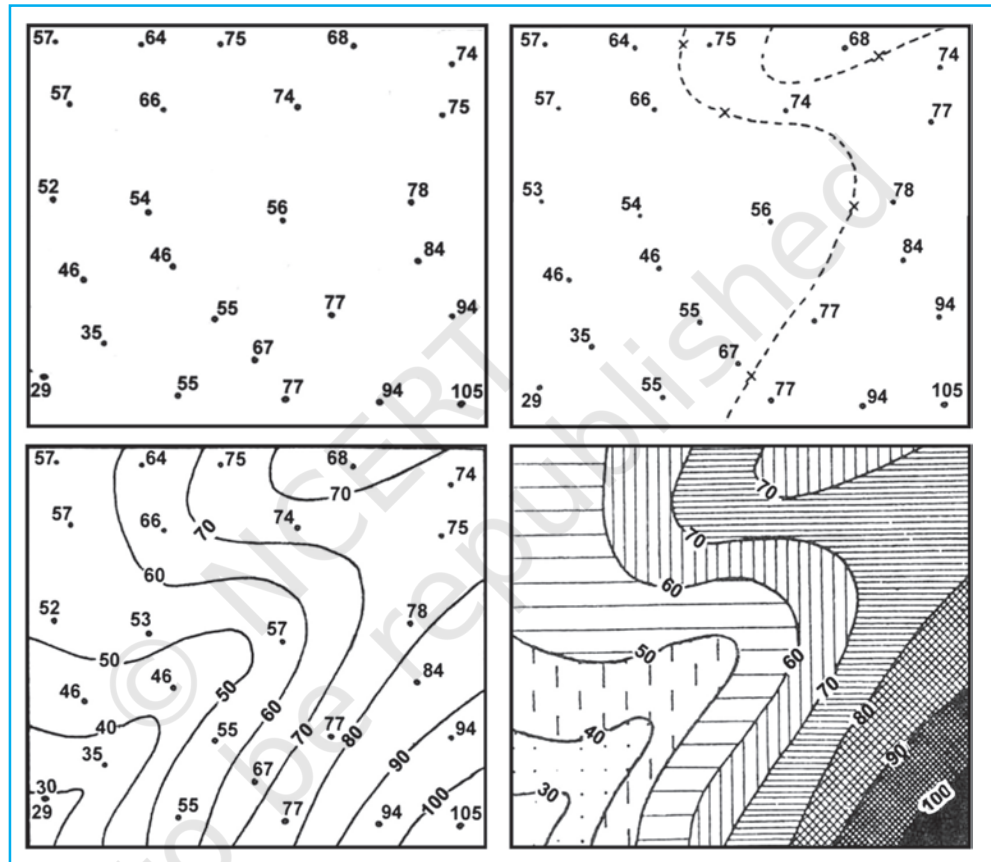


Fig. 3.15 : Drawing of Isoleths

Exercises

- Choose the right answer from the four alternatives given below:
 - Which one of the following map shows the population distribution:
 - Choropleth maps
 - Isoleth maps
 - Dot maps
 - Square root maps
 - Which one of the following is best suited to represent the decadal growth of population?
 - Line graph
 - Bar diagram
 - Circle diagram
 - 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:

- What is a thematic map?
- Differentiate between multiple bar diagram and compound bar diagram.
- What are the requirements to construct a dot map?
- Describe the method of constructing a traffic flow map.
- What is an Isopleth map ? How an interpolation is carried out?
- Describe and illustrate important steps to be followed in preparing a choropleth map.
- 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

- Construct a multiple bar diagram to show area under rice in each State.
- Construct a pie-diagram to show the percentage of area under rice in each State.
- Construct a dot map to show the production of rice in each State.
- 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



12101CH04

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 representation 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 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 a 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 computation and data processing.
2. It can handle huge volume of 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 hardware and software. The hardware configurations comprise storage, display, and input and output sub-systems, whereas software are the programs that are made 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 equipment. 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 a typewriter. It has various keys for different purposes. While working on a Personal Computer (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.

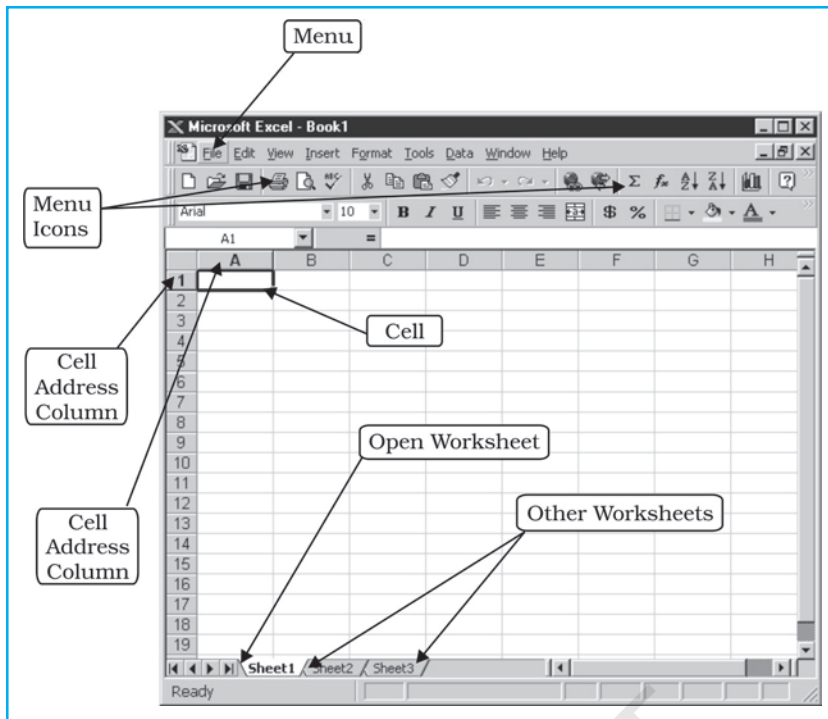


Fig. 4.1 : MS Excel Workbook

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.

	A	B	C	D	E
1	States	Urban Population	Total population	Percent of urban population	
2	Andhra Pradesh	20503597	75727541	=B2/C2*100	
3	Arunachal Pradesh	222688	1091117		
4	Assam	3389413	26638407		
5	Bihar	8679200	82878796		
6	Chhattisgarh	4175329	20795956		
7	Gujarat	18899377	50596992		
8	Haryana	6114139	21082989		

Fig. 4.2 : Cell Operation in MS Excel

	A	B	C	D
1	States	Urban Population	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 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 the 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

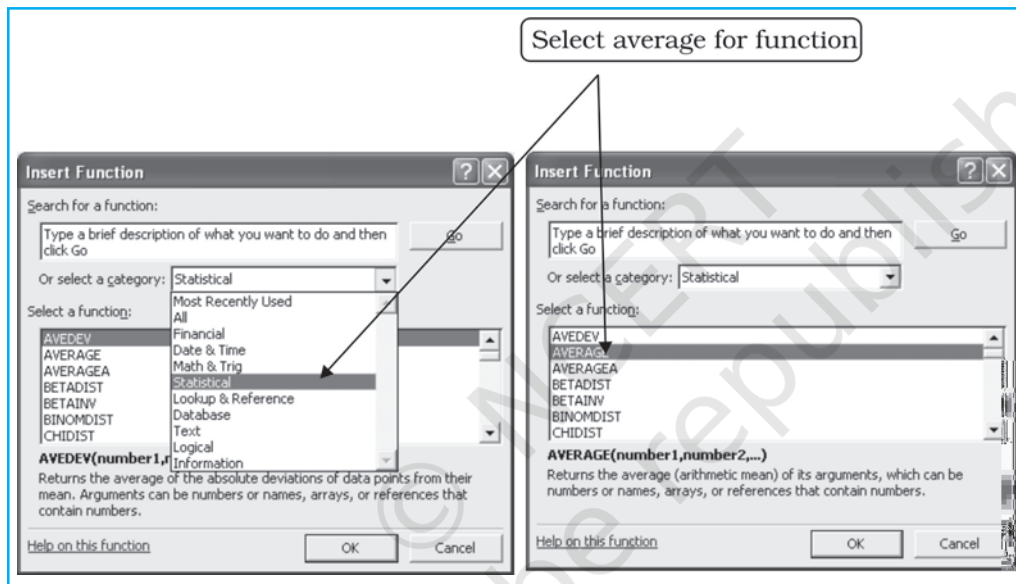


Fig. 4.5 : Selection of Statistical Function

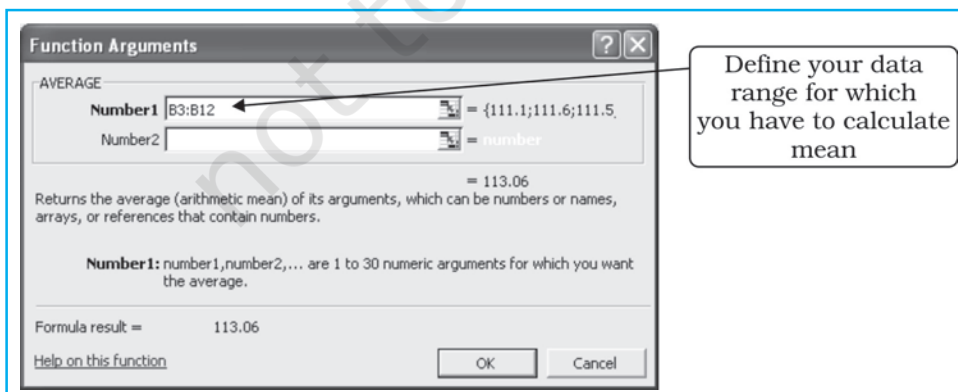


Fig. 4.6 : Defining Range in Function Arguments dialog box

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*.

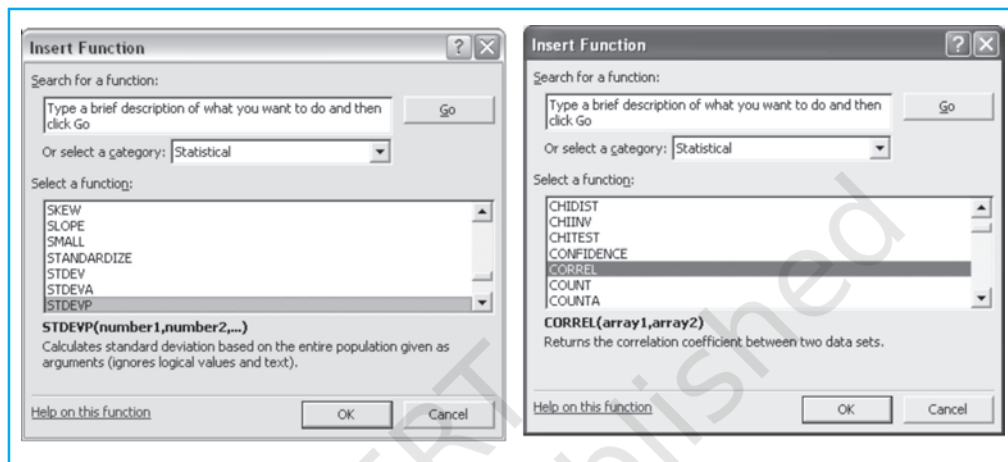


Fig. 4.7 : Function for Standard Deviation

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 50 years are presented in a 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 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.

	A	B	C	D	E	F	G
		Cultivators	Agricultural Labourers	Household Labourers	Other workers		Chart Wizard
4							
5	1981	43.66	19.84	3.17	33.33		
6	2001	31.34	20.82	3.02	44.82		
7							

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.

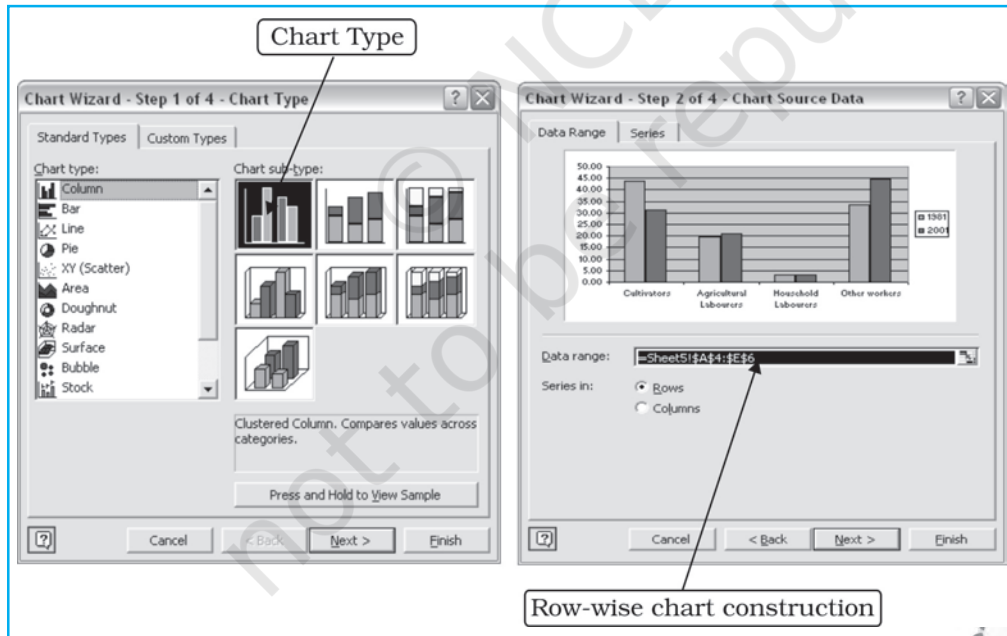


Fig. 4.10 : Step 1 to 4 of Chart Wizard **Fig. 4.11 :** Step 2 of 4 of Chart Wizard

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*.

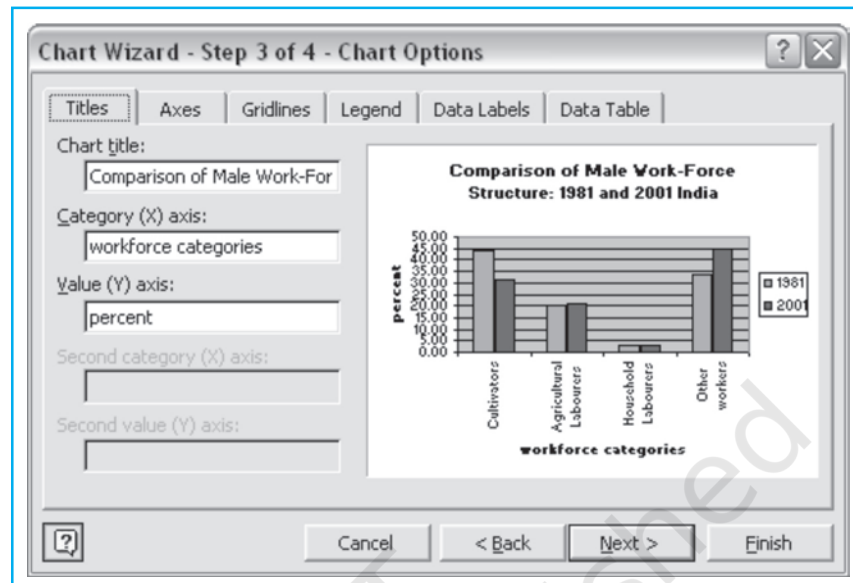


Fig. 4.12 : Entering names of Axes

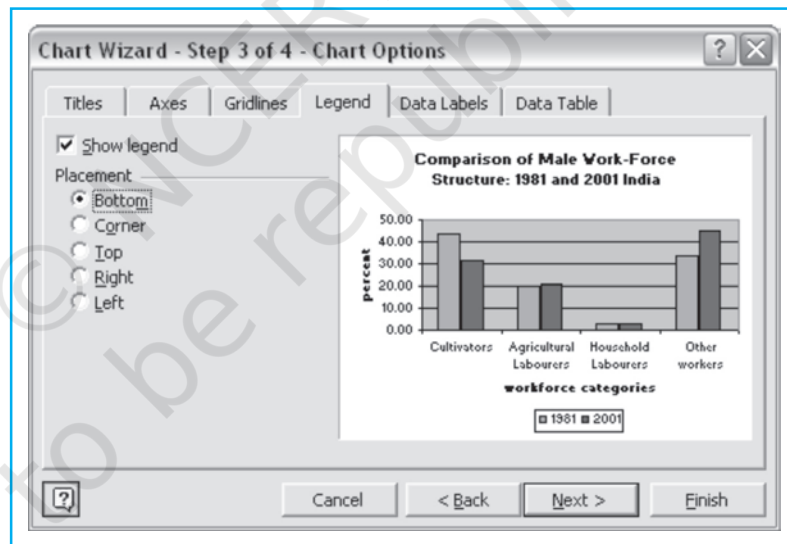


Fig. 4.13 : Choosing Location of Legend

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 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.

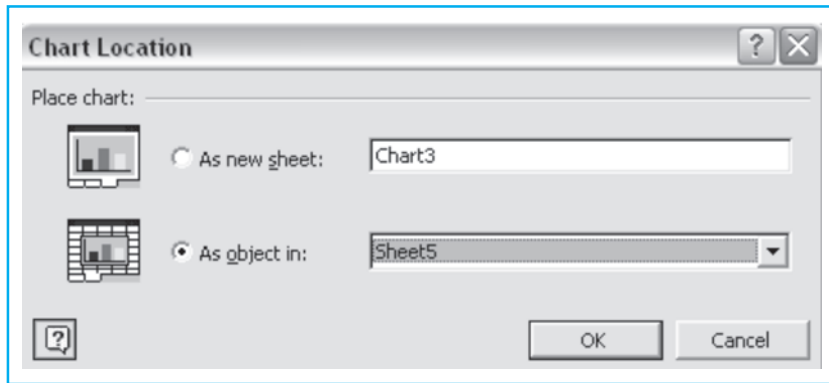


Fig. 4.14 : Choosing Location of Chart

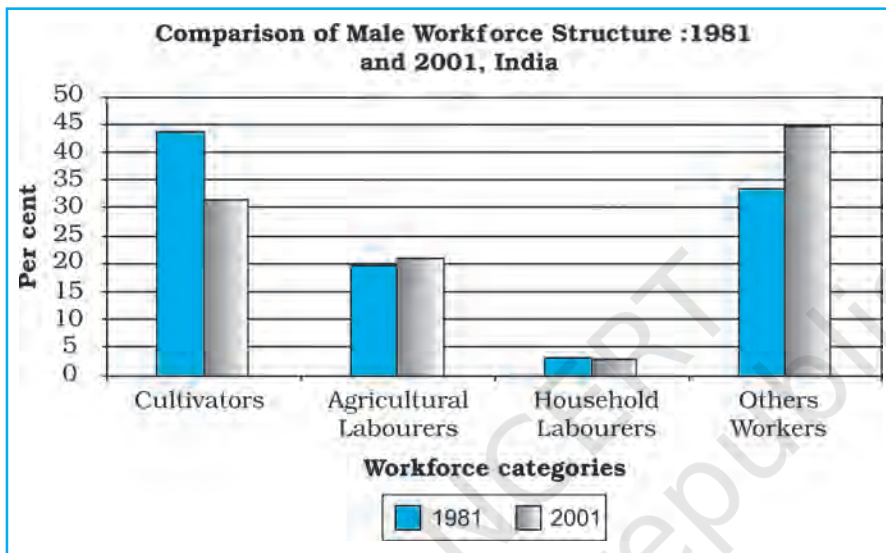


Fig. 4.15 : The Complete Bar Diagram

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 along with its integration with attribute or non-spatial data. It further involves the 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 database 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 polygons. The point data represent positional characteristics of 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 the 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 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 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, equipment, 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
Odisha	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