

Many things are seen over the surface of the earth. Most of them such as landforms, rivers, minerals, vegetation, animals are created by Nature, etc. Moreover Man himself also resides on this earth. Man has created many things through his acumen. These include population, dwellings, agriculture, industries, means and types of transport etc. These natural as well as man-made things are known as Natural and Cultural elements respectively.

These elements are spread unevenly over the surface of the earth.

Technically speaking, their distribution is uneven. Information can be collected about these elements. Such information is called **"Geographic Information."** It can be in **numerical form** as well, and in that case it is called **'Data'**. These data convey some characteristics about the distribution of any elements, its number or characteristics. So, such data can be **qualitative** as well as **quantitative** in nature.

Need for Information (Data)

It is important to know as to how these natural as well as cultural elements are distributed on the surface of the earth. It is not enough to know only about their location, but it is also necessary to know how these elements are situated and located reciprocally. For a researcher it is necessary to know whether there is any relation in their arrangement. When the geographic information is collected and interpreted, a shocking conclusion may occur once a while. Such interpretation are useful and important in geographical studies. Such interpretation may be useful and important to other subjects also. Thus the knowledge about the distribution of natural and cultural elements on the surface of the earth has become a necessity for any study.

Presentation of Information

Only the information of the elements on the surface of the earth does not serve any purpose. It is also necessary to know how to present the collected information. As a matter of fact this is more important. This presentation can be made in various forms like numerical, tabular or in a visual pictorial form. This form of the presentation depends on the fact whether the data is quantitative or qualitative in nature, because it is difficult to show quantitative information of the dataset in the quantitative format.

Data sources and Types

Basically there are two methods to collect statistical data : (i) Direct method and (ii) Indirect method. A person or a group of persons goes in the field 'physically' or 'in person' to collect some information for general or special purpose. This information is recorded systematically. For the ground verification and for succeeding in his objectives, information about the region is recorded in the form of sketches, photographs, videography etc. This information, i.e. the Data, which the person collects by visiting the area physically is called **'Direct'** method and is called **'Primary Data'**.

Primary data is that data which is not collected and/or published by any other person earlier. Such data is indigenous and collected completely by the person himself. The process of visiting the area for collecting such data by an individual is technically known as **'Field Work'**. When a public contact is made for such field work, some questions are asked to the pre-decided number of

persons. These questions are pre-prepared, which is called a '**Questionnaire**' or a '**Schedule**'. In a questionnaire, the answers are to be collected by asking the questions to the person personally, while in a schedule, the **answers** are collected through mails, telephones, E-mails or through other indirect methods.

When a person uses a set of data collected by other person/s or by any institution or by any other published literature for his study, it is called '**Secondary data**' and it is the indirect method. Total population, structure of population, data from Government or semi- government and private enterprises can be included in this category. These data are not restricted. Any person can use these data for his/her personal study. It is mandatory to mention credit and the source of the data.

Data Compilation and Presentation

It is necessary to keep a record of all data because the data collected at the primary level for any specific purpose may be useful to some other persons in future. That is why the data should be preserved. Earlier, the data (information) was preserved only through the writings, occasionally through diagrams or through pictures. The write-ups of early travellers justify this fact. The physical and cultural data of various places of earlier periods can be known even today. The details of which type of vegetation and animals existed at various place, how were the life style and physical appearance of the people were in those days etc. are information compiled in these write-ups.

With the passage of time the forms to compile the data changed. The data was preserved in the form of pictures, and later on by black & white photographs. Today the geographic information (i.e. data) can be stored very comfortably through the coloured photographs, cinemotography, videography, mobile (cell phones) and internet. A noteworthy contribution has come from satellites. Among all these, the data compiled and given in maps is very important. Many maps are published by Government and private publishers.

Certified and authentic maps giving information about India are published by **Survey of India (SOI - Dehra Dun, Uttarakhand State)** and by **National Atlas and Thematic Mapping Organization (NATMO, Kolkata, West Bengal)** on behalf of Government of India. These institutions have also published atlases with different purposes. Some private institutions also publish atlases giving information about India and foreign countries. Institutions which are associated with tourism include maps giving information about the tourists places of various countries in their brochures. Nowadays we also get video CDs and DVDs giving geographic information about various places.

Data Interpretation

The statistical data which is collected is divided into two categories for interpretation. These are :

- (1) Collection of data and
- (2) Classification of data.

When the data is being collected all details pertaining to them should be noted down very carefully. Inadequate and insufficient record of the data does not give correct result. Such data may lead to incorrect conclusions about the geographical study. So it is very important that the data is correct.

Once the informative data is collected, it is to be edited. This process is known as '**Editing**'. It is possible that much information has been collected but afterwards some of the data may not be useful. Such information is deleted from the dataset and only that information which may prove to be useful is retained. This process is known as '**Editing**'. The size of the data can be reduced by editing and its structure may be made more compact. Then the data is grouped under different headings and is presented in its ultimate format. While classifying the data, points like the range of the data, its regionality, quality, extent etc. all to be considered. This is known as **Data Presentation**. Data can be represented through maps and statistical methods.

Tabulation : Geographical and statistical data are collected for various purposes. It is possible that only certain information from it is useful for a study. So the information which is necessary out of the total information is to be sorted out and is rearranged into a special format. This process is known as '**Tabulation**'. A table consists of rows and columns. The table is prepared with necessary number of vertical and horizontal strips (or cells).The strip which runs horizontally from west to east is called **Row** and the strip which is drawn vertically north to south is called **Column**. Statistical information is entered into these cells made by rows and columns.

In any table, there is a heading which indicates the nature of the data. There can be sub-headings also. With this, a general out-lines of **Rows** and **Columns** is given below. It is an accepted tradition that the source of the data is mentioned at the bottom of the table.

Title					
<div>Headings of Rows</div> <div>Headings of Columns</div>	Main Heading				Total
	Sub Headings		Sub-Headings		
	Headings of the columns		Headings of the columns		
1	2	3	4	5	6

11.1 General Structure of rows and columns

Source :

This is a total basic out-line of a table. Necessary changes can be made in the design of the table to accommodate the collected data by adding more columns and rows. Let us take an

example to understand this point. Here, the information is given first in the form of a written text and then it is transformed into a table.

Example

A total of 377 candidates have applied for an admission to the post graduate class in Geography. On verification of their admission forms, the following data was derived.

There are 377 applications for getting an admission. Out of these 253 are boys and 124 are girls. There are 153 male candidates from Gujarat and the rest 100 belong to other states. Among the female candidates, 90 are from Gujarat State and 34 are from other states.

Among the applicants for the admission, 105 male students from Gujarat and 80 from other states have applied for hostel accommodation. Among the girls, 25 from Gujarat and 34 from other states have applied for hostel accommodation. (These figures can be shown in percentage as well).

The picture about the admission seekers does not become very clear by ONLY reading these details. It becomes a little bit difficult to compare and interpret them. So, if this data is transformed into a table, as shown below, the picture about the admission becomes clearer.

11.2 Geography : Application for the Post-Graduate classes in Geography

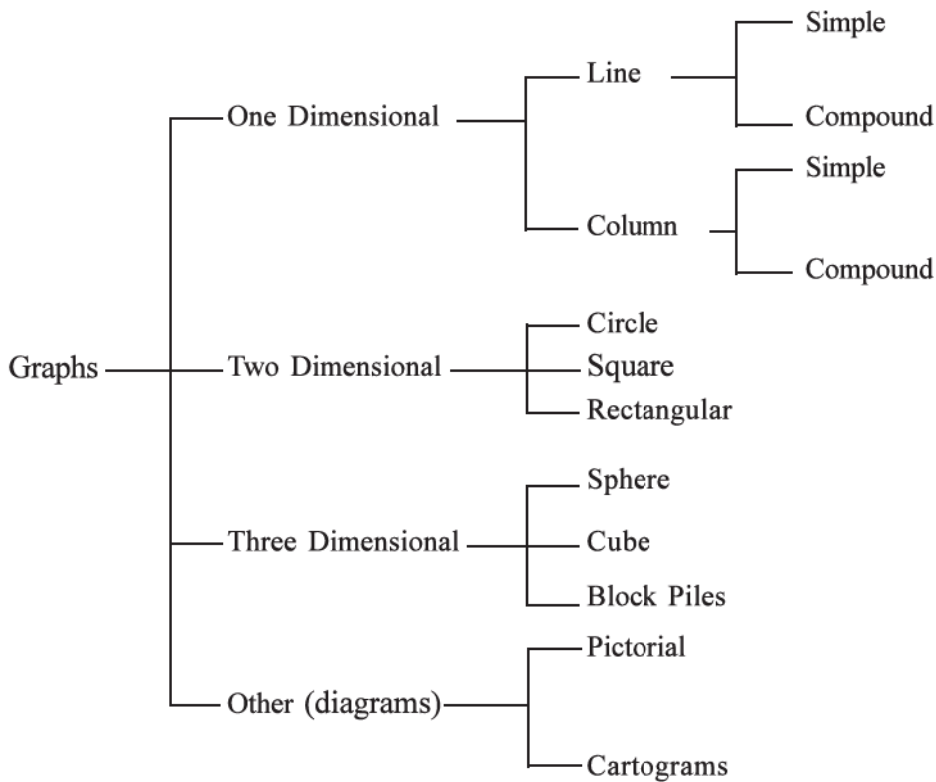
Details	From Gujarat State		From other States		Total Candidates		
	Boys	Girls	Boys	Girls	Boys	Girls	Total
	2	3	4	5	6 (2+4)	7 (3+5)	8 (6+7)
Total Applications	153	90	100	34	253	124	377
Applications for hostel accommodation	105	25	80	34	185	59	244

When the data is entered into the tables as shown here, a clear picture emerges about the details of the admissions. These figures arranged in this way also help us to compare them. Here, the admission forms were first scrutinised and then the details were transferred into a tabular form. This process in which the statistical data is transformed into a table is called '**Tabulation**'.

Graphs

The figures given in the table give information about something. It can be interpreted with the help of rows and columns. It may happen that the size of the table may be very large and may create adversity in its interpretation. Instead of figures only, if the data is arranged into a table it makes the interpretation easier. If this work is to be made still simpler, then these figures should be transformed into some diagramatic form, which is called 'Graph '. The data transformed into a pictorial form are understood much easily. The data can be transformed into different forms depending on their nature, and can be classified accordingly. This is shown in fig. 11.3.

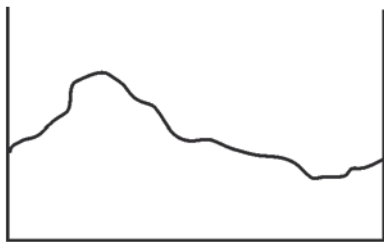
11.3 Classification of graphs



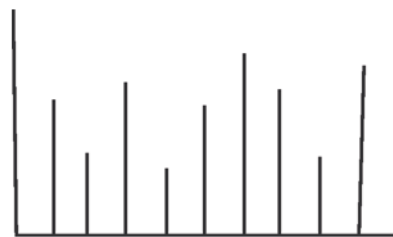
(1) One dimensional graph : In this type of graph that geometrical symbol is used which shows only one dimension to represent the data. A curved or a straight line can be used for it (fig 11.4).

Example :

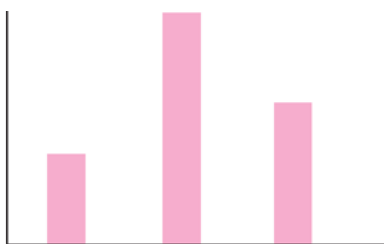
One dimensional graph



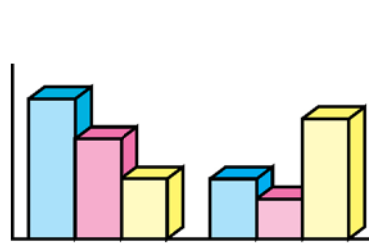
Frequency curve



Simple line graph



Bar



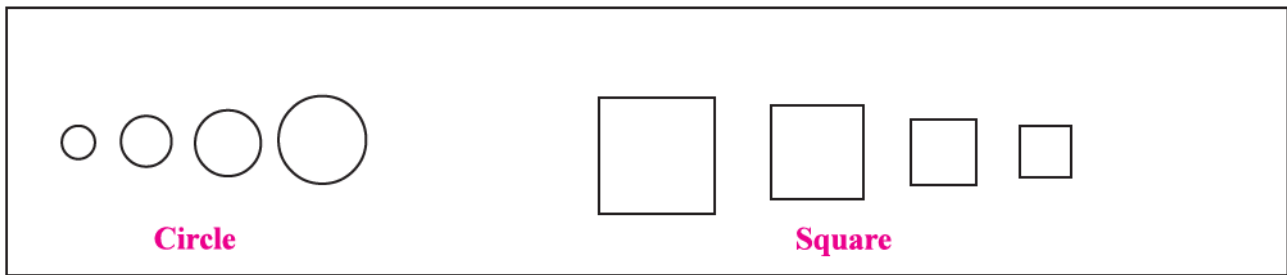
Three dimensional compound bars



11.4 One dimensional symbols

(2) Two dimensional graph : Those geometric symbols which show two dimensions simultaneously are used here to prepare the graph. These symbols include circles and squares, because these symbols convey two dimensions simultaneously (fig. 11.5).

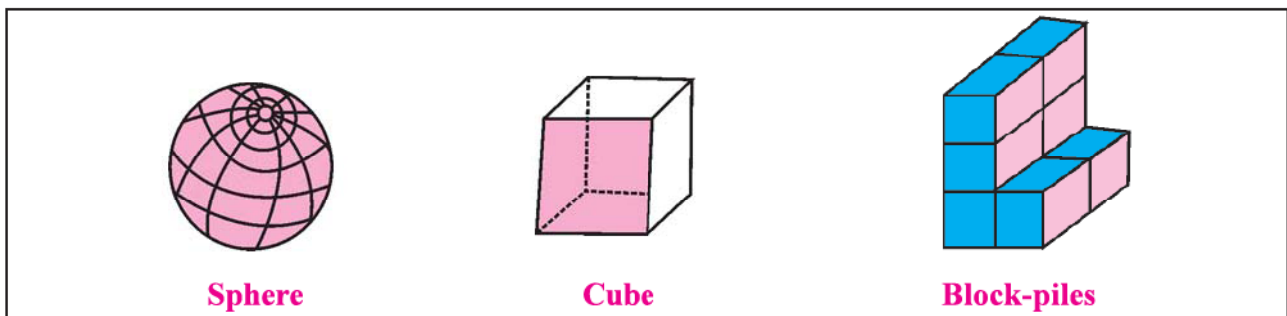
Two dimensional graph







11.5 Two dimensional diagrams

(3) Three dimensional graph : Where the size of the data is very large it can be transformed into a pictorial form by using three dimensional geometric symbols such as a sphere, cube or by block pile method etc. More data can be accommodated by using three dimensional symbols (fig 11.6).

Three dimensional graph



11.6 Three dimensional diagrams

(4) Pictorial diagrams : Here the data is represented by a symbol which would resemble to the information given in the data, e.g. a symbol of human figure  can be used to show demographic information. Moreover every symbol can be attributed its value, e.g. each  = 10,000 persons. Similarly a leaf  can be used to indicate crop production or the symbol of a tree  may be used to give information about forests along with their values. Interpretation can be made from such data also.

Statistical Methods : In the next phase testing of the data takes place. Some statistical methods (techniques) are used for this purpose. These techniques are selected according to the purpose of the data testing. There are many techniques to test a data but three fundamental techniques are used more. These are :

- (1) Measures of Central Tendency, (2) Dispersion (3) Correlation

Now we shall learn more about the Central Tendency.

Measures of Central Tendency : A set of primary and/or secondary data is collected for

any study. The data is collected under different categories or units. The group of all units is called the '**Population**' of the study. Here the figures taken for study is called '**Sample**'. With the help of questionnaire the data. Here the word 'population' means the **figures or the numerical** considered for the study. It is not to be misunderstood for any demographic unit. The collected data is tabulated. The data in the tables show certain attributes. Few statistical techniques are used to test these attributes.

These tests are expected to reveal the central value of the data. There are three major techniques to find out the values which lie somewhere within the data. These are : **(1) Mean (2) Median and (3) Mode**. With the help of these three different techniques, the central tendencies are calculated. These techniques are described below :

(1) Mean : This is the most widely used technique to measure a central tendency of any data. It can be defined as : "**The mean is a value which is derived by summation of all values and then dividing it by the number of observations**". It is written as **X** and is pronounced as '**X-Bar**'. every frequency is identified by the sign x. According to the number of frequencies, these are written as X_1, X_2, X_3, \dots and so on. The sum total of all frequencies is mentioned as **'Sigma'**. For the total number of frequencies, '**n**' is written. Let us take an example to understand these facts.

Example : Given below are the heights in centimetres of seven students. Find out the Mean of this data.

Height : 116, 123, 110, 114, 122, 130, 125

There are seven frequencies in this data. It will be written as follows :

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$$

Now if we write the values, it will be like this. More over, its summation is also to be calculated. So it is written as follows :

$$116 + 123 + 110 + 114 + 122 + 130 + 125$$

Formula : $\bar{m} = \frac{\sum x}{n}$

Where \bar{m} is the mean,

$$\bar{m} = \frac{116 + 123 + 110 + 114 + 122 + 130 + 125}{7}$$

$$= \frac{840}{7}$$

$$= 120$$

Thus the mean of the heights of these seven students would be 120. Statistical data are of different nature, so accordingly, the techniques to find out their means are also different.

Median : This is another technique to calculate a central tendency. Its general meaning is '**one which is in the centre**' or '**one which is located in the middle**'. It is located in the middle among all the frequencies of the data. Perhaps this is the reason why it is called a '**median**'.

A median is the value which divides the frequencies into two equal parts. The number of frequencies having more value than the median is the same as the number of frequencies having less value. However it is not so always. It is possible only when the frequencies are in odd numbers. Under this situation the frequencies are arranged into either ascending or in descending order, and the frequency which is in the middle of it is considered as the '**median**' of the data. It can be calculated with the help of the following formula.

$$\text{Formula : } M = \frac{N+1}{2}, \text{ Value of that ranked frequency}$$

Here M is the median and n is the number of frequencies. Let us take an example.

Example : The market rates (in rupees) of the share of a Company for 11 days were as follows :

Rates of the share : 148, 185, 240, 251, 255, 243, 205, 190, 210, 225, 237

Let us arrange this rates, i.e. frequencies into an ascending order, as follows :

148, 185, 190, 205, 210, 225, 237, 240, 243, 251, 255

There are a total of 11 frequencies which is an **odd number**. Now when we replace these figures into the formula, it will be as under :

$$M = \frac{N+1}{2}; M = \frac{11+1}{2} = \frac{12}{2} = 6 \text{ (i.e. the value of the sixth ranked frequency)}$$

Here the M will be the value of the sixth ranked frequency in the data. This value is 225. Thus the median of the market rates of the shares for 11 days will be 225 (rupees). This way it becomes easier to find out the median of the data having odd number of frequencies. But if the number of the frequencies is in even, there is a different method to find out its median.

If the frequencies are in even numbers, then the mean of the centrally located two frequencies is taken as the median of the data. Let us take an example.

Example : There are 10 employees in one institution. Following are figures of their age. Find out the median of this data :

Age : 42, 45, 39, 52, 55, 48, 50, 53, 36, 33

If the total number of the frequencies, i.e. n, is even, then the following formula will be applicable. The formula is : $m = \frac{N}{2}$ where the small m represents the serial number of the frequency. In such data, the value of the median is calculated with help of the following formula :

$$\text{Formula : } M = \frac{\text{Value of 'm' th frequency} + \text{value of m+1th frequency}}{2}$$

Now we arrange these frequencies in ascending order which will be,

33, 36, 39, 42, 45, 48, 50, 52, 53, 55 There a total of 10 frequencies, so $n = 10$. The frequencies are even numbers, so the combined value of 5th and 6th frequencies, $m + (m + 1)$ will be 93 (45 + 48). Now using the formula, it will be

$$M = \frac{\text{Value of the 5th frequency} + \text{Value of the } (5 + 1)\text{th frequency}}{2}$$

$$\frac{45 + 48}{2} = \frac{93}{2} = 46.5$$

Thus the median of this data with even number of frequencies will be 46.5.

Mode : In statistical data the frequency which appears repetitively for maximum number of times is called '**Mode**'. It is indicated by a symbol resembling to the English letter '**Z**' or as **Mo**. In mean and median only one frequency becomes the answer while here, more than one frequency can be the answer. Let us take an example to understand this.

Example : Find out the mode of the following data :

In a footwear shop, shoes having the following numbers were sold on one day.

Shoe number : 6, 6, 9, 8, 7, 7, 7, 9, 8

Here a total of nine frequencies are given. Let us rearrange these frequencies in ascending order. It will be :

Ascending order : 6, 6, 7, 7, 7, 8, 8, 9, 9

In this data, the frequency showing the value of 7 appears repeatedly for three times which is the maximum. This frequency, i.e. 7 is the 'mode' or $Z = 7$ of this data set.

There is only one value which reappears maximum number of times. Compared to that, other frequencies are less in number. But sometimes more than one frequency may appear repeatedly for more number of times. In such cases, all such frequencies are considered to be the mode of the data. This will be clear from the following example.

Example : Given below are the ages of 15 research scholars working in a research institution. Find the mode of this data :

Age : 23, 27, 23, 30, 36, 32, 28, 23, 25, 23, 30, 35, 30, 30, 28

These frequencies are rearranged in ascending order, as follows :

23, 23, 23, 23, 25, 27, 28, 28, 30, 30, 30, 30, 32, 35, 36

In this data, the frequency value of 23 appears a maximum number of four times. Similarly the value of 30 also appears four times. Other frequencies show lesser values. These two values, i.e. 23 and 30 have appeared for maximum times. These are identified respectively as $Z = 23$ and $Z = 30$ Modes.

The data where there is only one mode is known as **Unimodal Data** and if there are two modes, it is called **Bimodal Data**.

There can be more than two modes in a dataset and it may also happen that a mode cannot be decided in a dataset.

Examples : Following are the temperature data of a city for 7 days.

Temperature in $^{\circ}\text{C}$: 26, 29, 33, 36, 32, 31, 28

In this dataset, not a single frequency is repeated, so there is no mode in this dataset. The mean and the median of this dataset can be calculated but not the mode.

This way, a mean, a median and a mode of any dataset can be calculated. All these three values are known as the **Central Tendencies** of any dataset.

Exercise

1. Answer the following questions in details :

- (1) Classify the graphs and explain them.
- (2) Discuss the types of Central Tendencies.

2. Write to-the-point answer of the following questions :

- (1) Mention the sources to collect the data.
- (2) State the institution publishing official maps of India.
- (3) Define mean and Median and explain their importance.

3. Answer the following questions in brief :

- (1) Give brief information about tabulation.
- (2) What is meant by Two dimensional graph ?
- (3) Explain 'Population' giving illustrations.

4. Answer the following questions in one-two sentences :

- (1) Where is the Survey of India institute located ?
- (2) Write the full form of NATMO.
- (3) What is a Unimodal data ?

5. Select the correct option from the options given for the following questions and write answer :

- (1) Which of the following is a natural element ?
(a) Railway (b) River dam (c) Ahmedabad (d) Pavagadh
- (2) Which of the following is not a cultural element ?
(a) Granite (b) Settlement (c) State Highway (d) Metro Rail
- (3) Which of the following generates primary data ?
(a) Answers of the questions (b) Census Report
(c) Industrial data (d) Longest land route
- (4) What is the secondary data ?
(a) Data collected by the researcher himself (b) Data collected through personal visits
(c) Data taken from the published material (d) Unpublished data belonging to others

Activity

Find out the mean and the median of the data given below :

140, 150, 280, 185, 300, 156, 230
209, 105, 80, 100, 95

Find out the Mode of the following frequencies

27, 20, 25, 26, 31, 27, 7, 22, 27
27, 20, 18, 20, 20, 27, 20, 18, 17