3

Presentation of Data

INTRODUCTION

After the data have been collected, the next step is to present them in some suitable form. The need for proper presentation arises because of the fact that statistical data in their raw form almost defy comprehension. When data are presented in easy to-read form, it can help the reader to acquire knowledge in much shorter period of time and also facilitate statistical analysis, Presentation can take two basic forms : (*i*) Statistical Table, and (*ii*) Statistical Chart.

A statistical table is presentation of numbers in a logical arrangement, with some brief explanation to show what they are. However, before, tabulating data it is often necessary to first classify them. A statistical chart or a graph is a pictorial device for presenting data. The present chapter has been divided into three main parts to enable greater clarity : (*A*) Classification of data, (*B*) Tabulation of data, (*C*) Charting data.

(A) CLASSIFICATION OF DATA

After collection and editing of data an important step towards processing the data is classification. Classification is the grouping of related facts into different classes. Facts in one class differ from those of another class with respect to some characteristics called a basis of classification. Sorting facts on one basis of classification and then on another basis is called cross-classification. This process can be repeated as many times as there are possible basis of classification. Classification of data is a function very similar to that of sorting letters in a post office. It is well known that the letters collected in a post office are sorted into different lots on a geographical basis, *i.e.*, in accordance with their destinations as Mumbai, Kolkata, Kanpur, Jaipur, etc. They are then put in separate bags, each containing letters with a common characteristic, *viz.*, having the same destination. Classification of statistical data is comparable to the sorting operation. The process of classification gives prominence to important information gathered while dropping unnecessary details facilitates comparison and enables a statistical treatment of the material collected.

Types of Classification

Broadly, the data can be classified on the following four basis :

(*i*) Geographical, *i.e.,* area-wise, *e.g.,* cities, districts, etc.

(*ii*) Chronological, *i.e.,* on the basis of time.

(*iii*) Qualitative, *i.e.,* according to some attributes.

(*iv*) Quantitative, *i.e.,* in terms of magnitudes.

(*i*) **Geographical classification.** In geographical classification data are classified on the basis of geographical or locational differences between the various items. For example, when we present the production of sugar cane, wheat, rice, etc., for various States, this would be called geographical classification.

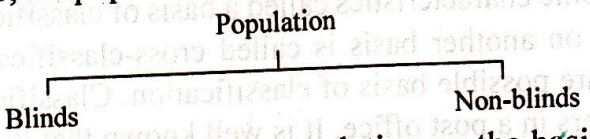
Geographical classifications are usually listed in alphabetical order for easy reference. Items may also be listed by size to emphasise the important areas as in ranking the States by population. Normally in reference tables the first approach is followed and in summary tables the second approach is followed.

(*ii*)**Chronological classification.** When data are observed over a period of time, the type of classification is known as chronological classification. For example, the sales figures of a company are given below :

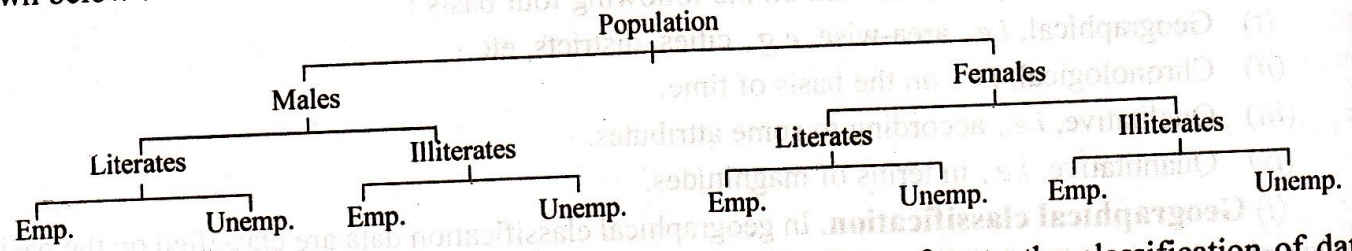
Year	*Sales (Rs. lakh)*	*Year*	*Sales (Rs. lakh)*
2000-01	18810	2005-06	46725
2001-02	23601	2006-07	45724
2002-03	23816	2007-08	50117
2003-04	32435	2008-09	53900
2004-05	39343	2009-10	61795

Time series are usually listed in chronological order normally starting with the earliest time period. When the major emphasis falls on the most recent events, a reverse time order may be used.

(*iii*) **Qualitative classification.** In qualitative classification, data are classified on the basis of some attribute or quality such as sex, colour of hair, literacy, religion, etc. The point to note in this type of classification is that the attribute under study cannot be measured: one can only find out whether it is present or absent in the units of the population under study. For example, if the attribute under study is blindness, we may find out how many persons are blind in a given population. It is not possible to measure the degree of blindness in each case. Thus when only one attribute is studied, two classes are formed; one possessing the attribute and the other not possessing it. This type of classification is known as simple classification. For example, the population under study may be divided into two categories as follows :



In a similar manner, we may classify population on the basis of sex, *i.e.*, males and females, or literacy, *i.e.*, literates and illiterates, and so on. This type of classification where only two classes are formed is also called twofold or dichotomous classification. If instead of forming only two classes we further divide the data on the basis of some attribute or attributes so as to form several classes, the classification is known as manifold classification. For example, we may first divide the population into males and females on the basis of the attribute, 'sex'; each of these classes may be further subdivided into 'literates and illiterates' on the basis of the attribute 'literacy'. Further classification can be made based on some other attribute, say, employment. The type of manifold classification described here is shown below :



(*iv*) **Quantitative classification.** Quantitative classification refers to the classification of data according to some characteristics that can be measured, such as height, weight, income, sales, etc. For example, the workers of a factory may be classified according to wages as follows :

Monthly wages (Rs.)	*No. of workers*	*Monthly wages (Rs.)*	*No. of workers*
4000-4500	50	5500-6000	360
4500-5000	200	6000-6500	90
5000-5500	260	6500-7000	40
		Total	1,000

In this type of classification, there are two elements, namely (*i*) the *variable*, *i.e.*, the monthly wage in the above example, and (*ii*) *frequency*, *i.e.*, the number of workers in each class. There were 50 workers having income between Rs. 4000 and Rs. 4500, 200 workers having income between Rs. 4500 and Rs. 5000, and so on. The quantitative classification gives birth to a frequency distribution.*

Variable. A frequency distribution refers to data classified on the basis of some variable that can be measured such as prices, wages, age, number of units produced or consumed. The term 'variable' refers to the characteristic that varies in amount or magnitude in a frequency distribution. A variable may be either continuous or discrete (also called discontinuous). A continuous variable is capable of manifesting every conceivable fractional value within the range of possibilities, such as the height or weight of persons or the weight of a product. Thus, as a student grows, say, from 90 cm to 150 cm, his height passes through all values between these lines. On the other hand, a discrete variable is that which can try only be finite "jumps" and cannot manifest every conceivable fractional value. For instance, the number of rooms in a house can only take certain values as 1, 2, 3, etc. Similarly, the number of employees and number of machines in an establishment are discrete variables. Generally speaking, continuous data are obtained through measurements, while discontinuous data are derived by counting. Series which can be described by a continuous variable are called continuous series. Series represented by a discrete variable are called discrete series. The following are two examples of discrete and continuous frequency distributions :

No. of Children	*No. of Families*
0	10
1	400
2	800
3	700
4	250
5	150
6	50
	Total 2,360

(*a*) Discrete Frequency Distribution

Age (years)	*No. of Employees*
20—25	10
25—30	15
30—35	40
35—40	45
40—45	26
45—50	4
	Total 140

(*b*) Continuous Frequency Distribution

Although the theoretical distinction between continuous and discrete variables is clear and precise, in practical statistical work it is only an approximation. The reason is that even the most precise instruments of measurement can be used only to finite number of places. Thus from a practical viewpoint continuous series can never be expected to flow continuously with one measurement touching another without any break in actual observations.

Formation of a Frequency Distribution

The process of preparing this type of distribution is very simple. We have just to count the number of times a particular value is repeated which is called the frequency of that class. In order to facilitate counting, prepare a column of 'tally'. In another column, place all possible values of the variable from

* The word 'distribution' refers to the way in which the observations are distributed in different classes.

the lowest to the highest. Then, put a bar (vertical line) opposite the particular value to which it relates. To facilitate counting, blocks of five bars are prepared and some space is left in between each block. We finally count the number of blocks and bars corresponding to each value of the variable and place it in the column of frequency. The process shall be clear from the following example of the number of refrigerators sold on 22 working days by a leading agency house :

23, 30, 20, 26, 30 30, 20, 23, 40., 40, 26, 20, 23, 40, 28, 26, 23, 30, 40, 28, 28, 30.

FREQUENCY DISTRIBUTION OF THE NUMBER OF REFRIGERATORS SOLD

No. of Refrigerators	*Tally Bars*	*Frequency No. of days*
20	|||	3
23	||||	4
26	|||	3
28	|||	3
30	𝍸	5
40	||||	4
	Total	22

The table clearly shows that on 3 days 20 refrigerators were sold each day, on 4 days 23 refrigerators were sold each day, etc.

This method of classifying helps in condensing the data only where values are largely repeated, otherwise there will be hardly any condensation. In order to make the series more compact so that its characteristics can be easily studied, data may be classified according to class-intervals.

Classification according to Class-Intervals

This type of classification is most popular in practice. The following technical terms are important when data are classified according to class-intervals :

(*i*) **Class limits.** The class limits are the lowest and the highest values that can be included in the class. For example, take the class 20—40. The lowest value of this class is 20 and the highest 40. The two boundaries of a class are known as the lower limit and upper limit of the class. The lower limit of a class is the value below which there can be no value in that class. The upper limit of a class is the value above which no value can belong to that class. Of the class 70—89, 70 is the lower limit and 80 is the upper limit, *i.e.*, in this class, there can be no value which is less than 70 or more than 89. Similarly, if we take the class 90—109, there can be no value in that class which is less than 90 or more than 109.

(*ii*) **Class-intervals.** The span of a class, that is, the difference between the upper limit and the lower limit, is known as class-interval. For example, in the class 20—40, the class interval is 20 (*i.e.*, 40 minus 20). The size of the class-interval is determined by the number of classes and the total range in the data.

(*iii*) **Class frequency.** The number of observations corresponding to the particular class is known as the frequency of that class or the class frequency. In the illustration given on page 41, the frequency of the class 5000-6000 is 50 which implies that there are 50 employees having income between Rs. 5000 and Rs. 6000. If we add together the frequencies of all individual classes, we obtain the total frequency. Thus, in the same problem, the total frequency of the six classes is 550 which means that in all there are 550 employees whose income has been studied.

(*iv*) **Class mid-point.** It is the value lying half-way between the lower and upper class limits of a class-interval. Mid-point of a class is ascertained as follows :

$$\text{Mid-point of a class} = \frac{\text{Upper limit of the class} + \text{Lower limit of the class}}{2}$$

For the purpose of further calculations in statistical work the mid-point of each class is taken to represent that class.

There are two methods of classifying the data according to class-intervals, namely (*a*) 'exclusive' method, and (*b*) 'inclusive' method.

(*a*) **'Exclusive' method.** When the class-intervals are so fixed that the upper limit of one class is the lower limit of the next class, it is known as the 'exclusive' method of classification. The following data are classified on this basis:

Income (Rs.)	*No. of Employees*	*Income (Rs.)*	*No. of Employees*
5000-6000	50	8000-9000	150
6000-7000	100	9000-10000	40
7000-8000	200	10000 and above	10
			Total 550

It is clear that the 'exclusive' method ensures continuity of data inasmuch as the upper limit of one class is the lower limit of the next class. Thus, in the above example, there are 50 employees whose income is between Rs. 5000 and Rs. 5999.99. An employee who is getting exactly Rs. 6000 would be included in the class 6000-7000. This method is widely followed in practice. However, it is confusing to a layman who has no knowledge of statistics. For example, if a questionnaire includes an observation asking the respondent the number of times he visits the Super Bazar in a month and he is required to tick one of the categories: 5-10 and 10-15, a person who visits the Super Bazar 10 times may find it difficult to decide whether to put the tick in the space against the class 5-10 and 10-15. In the absence of any specific instructions, some people may tick the class 5-10 while others 10-15. Hence, whenever this method is used it is necessary to give clear instructions in the questionnaire. However, the reader should note that if class-intervals are given like 0-10, 10-20, etc., it is always presumed that upper limit is exclusive, *i.e.*, an observation exactly equal to the upper limit is not included in that class.

(*b*) **'Inclusive' method.** Under the 'inclusive' method of classification, the upper limit of one class is included in that class itself. The example given on the next page illustrates this method:

Income (Rs.)	*No. of Employees*	*Income (Rs.)*	*No. of Employees*
5000-5999	50	8000-8999	150
6000-6999	100	9000-9999	40
7000-7999	200	10000-10999	10
			Total 550

In the class 5000-5999, we include employees whose income is between Rs. 5000 and Rs. 5999. If the income of an employee is exactly Rs. 6000 he is included in the next class. The above example makes it clear that there is no confusion here of the type we find under the 'exclusive' method. We may have classes like 5000-5999.5 or 5000-5999.9, and so on.

It should be noted that both the inclusive and exclusive methods give us the same class frequencies, although the class-intervals are apparently different in the two cases. In the above example, in case of exclusive method the class interval is 100 whereas in case of inclusive method the class-interval is 99. However, 99 is not the correct class-interval. The correct class-interval is 100. It is because whenever 'inclusive' method is used for equal class-intervals, the class-interval is obtained by taking the difference between the two upper limits.

Principles of Classification

It is difficult to lay down any hard and fast rules for classifying the data as the type of classification. However, the following general considerations may be borne in mind for ensuring meaningful classification of data:

(1) The number of classes should preferably be between 5 and 15. However, there is no rigidity about it. The classes can be more than 15 depending upon the total number of observations in the data and the details required, but they should not be less than five because in that case the classification may not reveal the essential characteristics.

Struges suggested the following formula for determining the approximate number of classes:

$k = 1 + 3.322 \log N.$

k = The approximate number of classes.

N = Total number of observations.

log = The ordinary logarithm to the base of 10.

However, the precise number of classes to be used for a given variable depends upon personal judgment and other considerations such as the details required, the ease of calculation of further statistical work, etc.

(2) As far as possible one should avoid odd values of class-intervals, *e.g.*, 3, 7, 11, 26, 39, etc. Preferably, one should have class-intervals of either five or multiples of five like 10, 20, 25, 100, etc. The reason is that the human mind is accustomed more to think in terms of certain multiples of 5, 10 and the like. However, where the data necessitate a class-interval of less than 5 it can be any value between 1 and 4.

(3) The starting point, *i.e.*, the lower limit of the first class, should either be zero or 5 or multiple of 5. For example, if the lowest value of the data is 63 and we have taken a class-interval of 10, then the first class should be 60-70, instead of 63-73. Similarly, if the lowest value of the series is 76 and the class-interval is 5 then the first class should be 75 to 80 rather than 76 to 81.

(4) To ensure continuity and to get correct class-interval we should adopt 'exclusive' method of classification. However, where 'inclusive' method has been adopted it is necessary to make an adjustment to determine the correct class-interval and to have continuity. The adjustment consists of finding the difference between the lower limit of the second class and the upper limit of the first class, dividing the difference by two, subtracting the value so obtained from all lower limits and adding the value to all upper limits. This can be expressed in the formula as follows :

$$\text{Correction factor} = \frac{\text{Lower limit of the 2nd class} - \text{Upper limit of the lst class}}{2}$$

How the adjustment is made when data are given by inclusive method can be seen from the following example :

Monthly Wages (Rs.)	*No. of Workers*	*Monthly Wages (Rs.)*	*No. of Workers*
5000-5999	5	8000-8999	18
6000-6999	10	9000-9999	12
7000-7999	15	10000-10999	4

To adjust the class limits, we take the difference between 6000 and 5999 which is one. By dividing it by two we get 1/2 or 0.5. This (0.5) is called the correction factor. Deduct 0.5 from the lower limits of all classes and add 0.5 to upper limits. The adjusted classes would then be as follows :

Monthly Wages (Rs.)	*No. of Workers*	*Monthly Wages (Rs.)*	*No. of Workers*
4999.5-5999.5	5	7999.5-8999.5	18
5999.5-6999.5	10	8999.5-9999.5	12
6999.5-7999.5	15	9999.5-10999.5	4

(5) Whenever possible all classes should be of the same size. If intervals are not of uniform width, it is difficult to make meaningful comparison between classes. At times, however, extreme observations may require the inclusion of so many class-intervals that the frequency distribution will become unwieldy. The observations are then classified as follows: below 200, 200-400, 400-600, 600-800, 800 and above. These classes are called *open-end classes* and distribution is known as an *open-end frequency distribution*. When frequency distribution is being employed as a technique of presentation only, open-end classes do not seriously reduce its usefulness as long as only a few observations fall in these classes. However, use of the frequency distribution for purposes of further mathematical computation is not helpful because a mid-point value, which can be used to represent the class, cannot be determined for an open-end class.

It may be noted that the frequency table, like other types of data presentation, is always constructed to serve some specific purpose. The technical requirements outlined above must be supplemented by sound subjective judgement if proper and useful frequency distributions are to be formed.

Illustration 1. The profits (in lakhs of rupees) of 30 companies for the year 1999-2000 are given below :

20, 22, 35, 42, 37, 42, 48, 53, 49, 65, 39, 48, 67, 18, 16, 23, 37, 35, 49, 63, 65, 55, 45, 58, 57, 69, 25, 29, 58, 65.

Classify the above data taking a suitable class-interval.

Solution: Let us determine the suitable class-interval with the help of the following formula:

$$i = \frac{\text{Range}}{1 + 3.322 \log N}$$

Range = (69–16) = 53, $N = 30$

$$i = \frac{53}{(1 + 3.322 \times 1.4771)} = \frac{\cdot 53}{1 + 4.91} = \frac{53}{5.91} = 8.97 \text{ or } 9.$$

Since values like 3, 7, 9, etc., should be avoided and therefore, we will take 10 as the class-interval and the first class as 15–25.

FREQUENCY DISTRIBUTION OF THE PROFITS

Profits (Rs. lakhs)	*Tally Bars*	*No. of Companies*
15–25	𝍸	5
25–35	𝍸	2
35–45	𝍸 ‖	7
45–55	𝍸 |	6
55–65	𝍸	5
65–75	𝍸	5
		Total 30

Illustration 2. Present the following data of the marks of 60 applicants who were given a certain test for the purpose of selection to a post :

41	17	83	63	55	92	60	58	70	06
67	82	33	44	57	49	34	73	54	63
36	52	32	75	60	33	09	79	28	30
42	93	43	80	03	32	57	67	84	64
63	11	35	28	10	23	08	41	60	32
72	53	92	88	62	55	60	33	40	57

Take first class as 0–9.

Solution : **FREQUENCY DISTRIBUTION OF THE MARKS OF 60 APPLICANTS**

Marks	*Tally Bars*	*Frequency*
0–9	‖‖	4
10–19	‖‖	3
20–29	‖‖	3
30–39	𝍸 𝍸	10
40–49	𝍸 ‖	7
50–59	𝍸 ‖‖	9
60–69	𝍸 𝍸 ‖	11
70–79	𝍸	5
80–89	𝍸	5
90–99	‖‖	3
		Total 60

Illustration 3. The data given below relate to the sales and advertisement expenditure of 20 companies. You are required to form a bivariate frequency distribution with class interval 62 to 64, 64 to 66, and so on and 115 to 125, 125 to 135 and so on.

Company	*Sales (Rs. Lakhs)*	*Adv. Exp. (Rs. Lakhs)*	*Company*	*Sales (Rs. Lakhs)*	*Adv. Exp. (Rs. Lakhs)*
1	170	70	11	163	70
2	135	65	12	139	67
3	136	65	13	122	63
4	137	64	14	134	68
5	148	69	15	140	67
6	124	62	16	132	69
7	117	65	17	120	66
8	128	70	18	148	68
9	143	71	19	120	67
10	129	62	20	152	67

Solution : As per the requirements of the question, the data are to be divided into five classes according to the advertisement expenditure and six classes according to the sales.

For tabulating the information in appropriate cells, first, the row to which the advertisement expenditure (say, X) should belong is determined. Afterwards on a consideration of the sales (say, Y) the column in which it should be included is determined. The tabulation is recorded by tally bars. Thus the two-way table shall be prepared as follows :*

TWO-WAY FREQUENCY TABLE SHOWING SALES AND ADVERTISEMENT EXPENDITURE OF 20 COMPANIES

Adv. Exp. (X) \ *Sales (Y)*	115–125	125–135	135–145	145–155	155–165	165–175	Total
62–64	II (2)	I (1)	–	–	–	–	3
64–66	I (1)	–	III (3)	–	–	–	4
66–68	II (2)	–	II (2)	I (1)	–	–	5
68–70	–	II (2)	–	II (2)	–	–	4
70–72	–	I (1)	I (1)	–	I (1)	I (1)	4
Total	5	4	6	3	1	1	20

*The figure in brackets denote the frequency corresponding to each cell, advertisement expenditure incurred and six classes according to the sales. There will be thus $5 \times 6 = 30$ cells.

(B) TABULATION OF DATA

One of the simplest and most revealing devices for summarizing data and presenting them in meaningful fashion is the statistical table.* A table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangement, whereas columns are vertical ones. The purpose of a table is to simplify the presentation and to facilitate comparisons. The simplification results from the clear-cut and systematic arrangement, which enables the reader to quickly locate desired information. Comparison is facilitated by bringing related items of information close together.

Parts of a Table

The various parts of a table may vary from case to case depending upon the given data. But a good table must contain at least the following parts :

1. Table number
2. Title of the table
3. Caption
4. Stub
5. Body of the table
6. Headnote
7. Footnote

1. **Table number.** Each table should be numbered. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or in the side of the table at the top or at the bottom of the table on the left-hand side. However, if space permits the table number should be given in the centre. Where there are many columns, it is also desirable to number each column so that easy reference to it is possible.

2. **Title of the table.** Every table must have a suitable title. The title is a description of the contents of the table. A complete title has to answer the questions *what, where* and *when* in that sequence. In other words,

(*a*) What precisely are the data in the table (*i.e.*, what categories of statistical data are shown)?

(*b*) Where the data occurred (*i.e.*, the precise geographical, political or physical area covered)?

(*c*) When the data occurred (*i.e.*, the specific time or period covered by the statistical material on the table)?

The title should be clear, brief and self-explanatory. However, clarity should not be sacrificed for the sake of brevity. Long title cannot be read as promptly as short title, but at times they may have to be used for the sake of clarity. The title should be so worded that it permits one and only one interpretation. It should be in the form of a series of phrases rather than complete sentences. *Its lettering should be the most prominent of any lettering in the table.*

3. **Caption.** Caption refers to the column headings. It explains what the column represents. It may consist of one or more column headings. Under a column-heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column. If the different columns are expressed in different units, the unit should be specified along with the captions. As compared with the main part of the table the caption should be shown in smaller letters. This helps in saving space.

4. **Stub.** As distinguished from caption, stubs are the designation of the rows or row headings. They are at the extreme left and perform the same function for the horizontal rows or numbers in the table as the column headings do for the vertical columns or numbers. The stubs are usually wider than column headings but should be kept as narrow as possible without sacrificing precision and clarity of statements.

---

* A statistical table is the logical listing of related quantitative data in vertical columns and horizontal rows of numbers with sufficient explanatory and qualifying words, phrase and statements to the form of titles, headings and notes to make clear the full meaning of data and their origin.

(B) TABULATION OF DATA

One of the simplest and most revealing devices for summarizing data and presenting them in meaningful fashion is the statistical table.* A table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangement, whereas columns are vertical ones. The purpose of a table is to simplify the presentation and to facilitate comparisons. The simplification results from the clear-cut and systematic arrangement, which enables the reader to quickly locate desired information. Comparison is facilitated by bringing related items of information close together.

Parts of a Table

The various parts of a table may vary from case to case depending upon the given data. But a good table must contain at least the following parts :

1. Table number
2. Title of the table
3. Caption
4. Stub
5. Body of the table
6. Headnote
7. Footnote

1. **Table number.** Each table should be numbered. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or in the side of the table at the top or at the bottom of the table on the left-hand side. However, if space permits the table number should be given in the centre. Where there are many columns, it is also desirable to number each column so that easy reference to it is possible.

2. **Title of the table.** Every table must have a suitable title. The title is a description of the contents of the table. A complete title has to answer the questions *what, where* and *when* in that sequence. In other words,

(*a*) What precisely are the data in the table (*i.e.*, what categories of statistical data are shown)?

(*b*) Where the data occurred (*i.e.*, the precise geographical, political or physical area covered)?

(*c*) When the data occurred (*i.e.*, the specific time or period covered by the statistical material on the table)?

The title should be clear, brief and self-explanatory. However, clarity should not be sacrificed for the sake of brevity. Long title cannot be read as promptly as short title, but at times they may have to be used for the sake of clarity. The title should be so worded that it permits one and only one interpretation. It should be in the form of a series of phrases rather than complete sentences. *Its lettering should be the most prominent of any lettering in the table.*

3. **Caption.** Caption refers to the column headings. It explains what the column represents. It may consist of one or more column headings. Under a column-heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column. If the different columns are expressed in different units, the unit should be specified along with the captions. As compared with the main part of the table the caption should be shown in smaller letters. This helps in saving space.

4. **Stub.** As distinguished from caption, stubs are the designation of the rows or row headings. They are at the extreme left and perform the same function for the horizontal rows or numbers in the table as the column headings do for the vertical columns or numbers. The stubs are usually wider than column headings but should be kept as narrow as possible without sacrificing precision and clarity of statements.

---

* A statistical table is the logical listing of related quantitative data in vertical columns and horizontal rows of numbers with sufficient explanatory and qualifying words, phrase and statements to the form of titles, headings and notes to make clear the full meaning of data and their origin.

5. **Body of the table.** The body of the table contains the numerical information. This is the most vital part of the table. Data presented in the body arranged according to descriptions are classifications of the captions and stubs.

6. **Headnote.** It is a brief explanatory statement applying to all or a major part of the material in the table, and is placed below the title entered and enclosed in brackets. It is used to explain certain points relating to the whole table that have not been included in the title nor in the captions or stubs. For example, the unit of measurement is frequently written as the headnote, such as "in thousands" or "in million tonnes", or "in crores", etc.

7. **Footnote.** Anything in a table which the reader may find difficult to understand from the title, captions and stubs should be explained in footnotes. If footnotes are needed, they are placed directly below the body of the table. Footnotes are used for four main purposes :

(*a*) To point out any exceptions as to the basis of arriving at the data, for example, sales recorded at 'delivered price' for others. Any heterogeneity in the data recorded must be disclosed to avoid wrong conclusions.

(*b*) Any special circumstances affecting the data, for example, strike, lock-out, fire, etc.

(*c*) To clarify anything in the table.

(*d*) To give the source in case of secondary data. The reference to the source should be complete in itself, for example, if the data are obtained from some periodical, its name, date of publication, page number, table number, etc., should be mentioned so that if the user wishes to check the data from the original source, he will know where to look for the information.

There are various systems of identifying the footnotes. One is numbering them consecutively with small number ^{1}, ^{2}, ^{3}, or letters *a, b, c, d.* Another system identifies the first footnote with one star (*), second footnote with two stars (**), third footnote with three stars (***), and so on. Sometimes instead of star another sign (dagger) is used. However, where several footnotes are required, it is more convenient to use small numbers like ^{1}, ^{2}, ^{3}, ... etc.

Review of the Table

Before a table is released it should be reviewed for form, content, validity and clerical accuracy. It is difficult for the person preparing the table to make a thoroughly satisfactory checks on all the four aspects. The person, who has prepared the table, might have done his best, but he can hardly review it objectively. He should, if possible, get his work reviewed by some experienced person.

In case of a summary table, the reviewer should ask himself the following questions to determine whether or not table is satisfactory :

(1) Does the title clearly state what is in the table ?

(2) Are all the entries pertinent ?

(3) Is there unity of subject-matter ?

(4) Are the classifications arranged so as to focus attention on the main comparisons ?

(5) Are the data arranged so as to emphasise important points ?

(6) Does the table include adequate interpretative figures such as totals, percentages and averages ?

(7) Are there notations about peculiarities of the data ?

(8) Is the source properly stated ?

(9) Is the table in proper form, so that it presents an attractive appearance ?

Types of Tables

Tables may broadly be classified into two categories :

1. Simple and complex tables ; and
2. General purpose and special purpose (or summary) tables.

1. **Simple and Complex Tables.** The distinction between simple and complex tables is based upon the number of characteristics studied.

In a simple table only one characteristic is shown. Hence, this type of table is also known as one-way table. In a complex table, on the other hand, two or more characteristics are shown. Such tables are more popular in practice because they enable full information to be incorporated and facilitate a proper consideration of all related facts. When two characteristics are shown, a table is known as two-way table or double tabulation. When three characteristics are shown in a table, this type of tabulation is known as treble tabulation. When four or more characteristics are simultaneously shown, it is a case of manifold tabulation. The following examples will illustrate the distinction between simple and complex tables :

(*i*) *Simple table or one-way table.* In this type of table only one characteristic is shown. This is the simplest of tables. The following is the illustration of such a table :

NUMBER OF EMPLOYEES IN AN ORGANISATION ACCORDING TO AGE GROUP

Age (in years)	*No. of Employees*
Below 25	50
25—35	67
35—45	43
45—55	15
55 and above	5
	Total 180

(*ii*) *Two-way table.* Such a table shows two characteristics and is formed when either the stub or the caption is divided into two coordinate parts. The example given on page 47 illustrates the nature of such a table :

NUMBER OF EMPLOYEES IN AN ORGANISATION ACCORDING TO AGE AND SEX

Age (in years)	*Employees*		*Total*
	Males	*Females*	
Below 35	32	18	50
25—35	40	27	67
35—45	25	18	43
45—55	10	5	15
55 & above	5	—	5
Total	112	68	180

(*iii*) *Higher order table.* When three or more characteristics are represented in the same table, such a table is called higher order table. The need for such a table arises when we are interested in presenting a number of characteristics simultaneously. While constructing such a table it is necessary to first establish an order of precedence among the attributes or characteristics sought to be classified having regard to their relative importance.

It should be remembered that as the number of characteristics represented increases, the table becomes more and more confusing and as such normally not more than four characteristics should be represented in the same table. Where more than four characteristics are to be represented we can have more than one table depicting relationship between different attributes.

2. **General Purpose and Special Purpose Tables.** General purpose tables, also known as the reference tables or repository tables, provide information for general use or reference. They usually contain detailed information and are not constructed for specific discussion. In other words, these tables serve as a repository of information and are arranged for easy reference. Tables published by governmental agencies are mostly of this kind, such as the tables contained in the *Statistical Abstract of the Indian Union,* detailed tables contained in the census reports, etc. Such tables tell facts which are not for particular discussion. When such table are used by a researcher, they are usually placed in the appendix of the report for easy reference.

Special purpose tables, also known as summary or analytical tables, provide information for prticular discussion. They show relationship between different groups of figures. When attached to a report they are found in the body of the text. These tables are also called derivative tables since they are often derived from general tables. Thus the large detailed tables in the census records of the Government of India are general purpose tables. When such data are used, they are ordinarily taken from the general purpose tables and presented as special purpose tables, which emphasise the relation the user wishes to stress. A special purpose table should be designed in such a way that a reader may easily refer to the table for comparison, analysis or emphasis concerning the particular discussion.

Illustration 4. In a sample study about the coffee habits in two towns, following data were observed:

Town *X*
52% persons were males,
25% were coffee drinkers, and
16% were male coffee drinkers

Town *Y*
55% persons were males,
28% were coffee drinkers, and
18% were male coffee drinkers.

Tabulate the above observations. *(MBA, HPU, 2006)*

Solution.

TABLE SHOWING PERCENTAGE OF COFFEE DRINKERS

Attributes	*Town X*			*Town Y*		
	Males	*Females*	*Total*	*Males*	*Females*	*Total*
Coffee drinkers	16	9	25	18	10	28
Non-coffee drinkers	36	39	75	37	35	72
Total	52	48	100	55	45	100

Illustration 5. Present in a tabular form with suitable captions, etc., for the information contained in the following:

"In 2000, out of a total of 2,000 workers of a factory 1,500 workers were members of a trade union. The number of women employed was 150 of which 128 did not belong to a trade union. In 2005, the number of trade union workers increased to 1,620 of which 1,582 were men. On the other hand, the number of non-union workers fell down to 448 of which 318 were men. In 2010, there were on the payrolls of the factory 2,200 workers of whom 2,000 belonged to a trade union. Of all the employees in 2010, 200 were women of whom only 25 did belong to a trade union."

Solution.

TABLE SHOWING TRADE UNION MEMBERSHIP

	2000			2005			2010		
Category	*Trade Union members*	*Non-members of T.U.*	*Total*	*Trade Union members*	*Non-members of T.U.*	*Total*	*Trade Union members*	*Non-members of T.U.*	*Total*
Men	1,478	372	1,850	1,582	318	1,900	1,825	175	2,000
Women	22	128	150	38	130	168	175	25	200
Total	1,500	500	2,000	1,620	448	2,068	2,000	200	2,200

(C) CHARTING DATA

One of the most convincing and appealing ways in which data may be presented is through charts. Evidence of this can be found in the financial pages of newspapers, journals, advertisements, etc. Pictorial presentation helps in quick understanding of the data. As the number and magnitude of figures increases, they become more confusing and their analysis tends to be more strenuous. A picture is said to be worth 10,000 words, *i.e.*, through pictorial presentation data can be presented in an interesting form. Not only this, charts have greater memorizing effect as the impressions created by them last much longer than those created by the figures.

A chart can take the shape of either a diagram or a graph. For the sake of clarity we will discuss them under two separate heads :

(*i*) Diagram, and (*ii*) Graphs.

(*i*) Diagrams

For representing data diagrams are more commonly used than graphs. However, before discussing different types of diagrams it would be worthwhile to consider some general rules for constructing diagrams.

General Rules for Constructing Diagrams

The following general rules should be observed while constructing diagrams :

1. *Title.* Every diagram must be given a suitable title. The title should convey in as few a words as possible the main idea that the diagram is intended to portray. However, the brevity should not be secured at the cost of clarity or omission of essential details. The title may be given either at the top of the diagram or below it.

2. *Proportion between width and height.* A proper proportion between the height and width of the diagram should be maintained. If either the height or width is too short or too long in proportion, the diagram would give an ugly look. While there are no fixed rules about the dimensions, convenient standard as suggested by Lutz in the book entitled *Graphic Presentation,* may be adopted for general use. It is known as "Root-two", that is ratio of 1 (short side) to 1.414 (long side). Modifications wherever necessary may be made to accommodate a diagram in the space available.

3. *Selection of appropriate scale.* The scale showing the values should be in even numbers or in multiples of five or ten, *e.g.*, 25, 50, 75 or 20, 40, 60. Odd values like 1, 3, 5, 7 should be

avoided. No rigid rules can be laid down about the selection of appropriate scale. The given data and the required size of diagram are the guiding factors. The scale should specify the size of the unit and what it represents, for example, "millions of tonnes", "number of persons in thousands", "units produced in lakhs", etc. *All lettering should be easily readable without turning the chart sidewise.*

4. *Footnotes.* In order to clarify certain points about the diagram, footnotes may be given at the bottom of the diagram.

5. *Index.* An index illustrating different types of lines or different shades, colours, should be given so that the reader can easily make out the meaning of the diagram.

6. *Neatness and cleanliness.* Diagrams should be absolutely neat and clean.

7. *Simplicity.* Diagrams should be as simple as possible so that the reader can understand their meaning clearly. For the sake of simplicity, it is important that too much material should not be loaded in a single diagram otherwise it may become too confusing and prove useless. Several simple charts are much better and more effective than one or two complex ones which present the same material in a confusing way.

Types of Diagrams

In practice, a very large variety of diagrams are in use and new ones are constantly being added. It would be outside the scope of this text to deal exhaustively with the subject and as such only more frequently used diagrams are discussed. For the sake of convenience and simplicity different types of diagrams are divided under the following heads :

I. One-dimensional diagrams, *e.g.*, bar diagrams.
II. Two-dimensional diagrams, *e.g.*, rectangles, squares and circles.
III. Pictograms and cartograms.

Each of these types is discussed below in detail :

I. One-dimensional or Bar Diagrams

Bar diagrams are the most common type of diagrams used in practice. A bar is a thick line whose width is shown merely for attention. They are called one-dimensional because it is only the length of the bar that matters and not the width. When the number of observations is large, lines may be drawn instead of bars to economise space. Special merits of bar diagram are the following :

(*i*) They are readily understood even by those unaccustomed to reading charts or those who are not chart-minded.

(*ii*) They possess the outstanding advantage that they are the simplest and the easiest to make.

(*iii*) When a large number of observations are to be compared, they are the only form that can be used effectively.

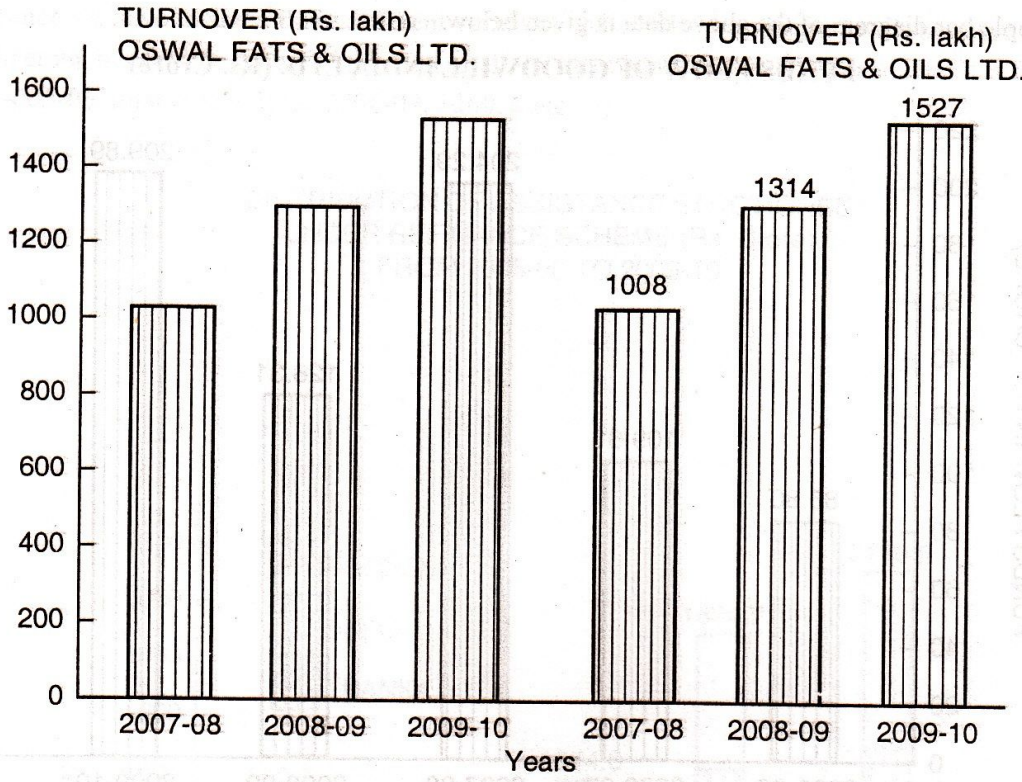
Points to be kept in mind while constructing Bar Diagrams

(*i*) The width of the bars should be uniform throughout the diagram.

(*ii*) The gap between one bar and another should be uniform throughout.

(*iii*) Bars may be either horizontal or vertical. The vertical bars should be preferred because they give a better look and also facilitate comparison.

(*iv*) While constructing the bar diagrams, it is desirable to write the respective figure at the end of each bar so that the reader can know the precise value without looking at the scale. This is particularly important when the scale is too narrow; for example, 1 cm on paper may represent 10 crore people. The following two diagrams would clarify the difference :



It is clear from the above two diagrams that from the left one it is difficult to read precise values whereas the right side diagram makes it clear.

Types of Bar Diagrams

Bar diagrams are of the following types:

(*a*) Simple bar diagrams
(*b*) Subdivided bar diagrams*
(*c*) Multiple bar diagrams
(*d*) Precentage bar diagrams
(*e*) Deviation bars
(*f*) Broken bars

(a) Simple Bar Diagrams

A simple bar diagram is used to represent only one variable. For example, the figures of sales, production, population, etc., for various years may be shown by means of a simple bar diagram. Since the bars are of the same width and only the length varies, it becomes very easy for the reader to study the relationship. Simple bar diagrams are very popular in practice. However, an important limitation of such diagrams is that they can present only one classification or one category of data. For example, while presenting the population for the last five decades, one can only depict the total population in the simple bar diagrams and not its sex-wise distribution.

Illustration 6. The funds flow of Goodwill India Ltd. from 2005-06 to 2009-10 are given below :

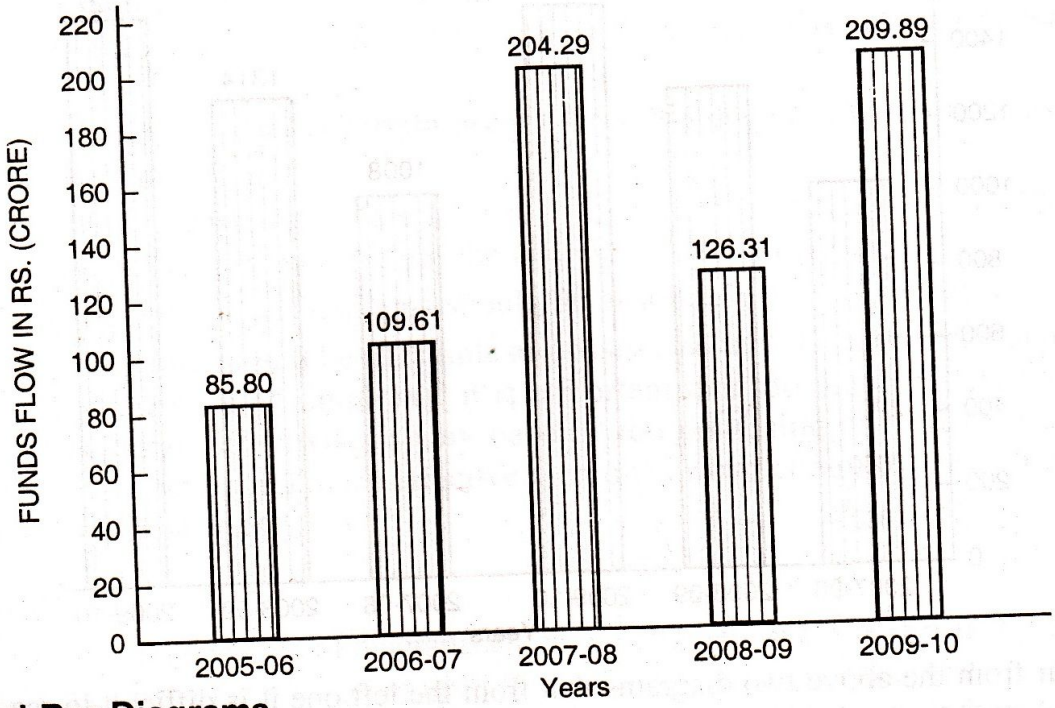
Year	*Funds Flow (Rs. crores)*
2005-06	85.80
2006-07	109.61
2007-08	204.29
2008-09	126.31
2009-10	209.89

Represent this data by a suitable bar diagram.

*Such diagrams are also known as component bar diagrams.

Solution. The simple bar diagram of the above data is given below.

FUNDS FLOW OF GOODWILL INDIA LTD. (Rs. Crore)



(*b*) Subdivided Bar Diagrams

These diagrams are used to represent various parts of the total. For example, the number of employees in various departments of a company may be represented by a subdivided bar diagram. While constructing such a diagram, the various components in each bar should be kept in the same order. A common and helpful arrangement is that of presenting each bar in the order of magnitude from the largest component at the base of the bar to the smallest at the end. To distinguish between the different components, it is useful to use different shades or colours. Index or key should be given explaining these differences. Sub-divided bar diagrams can be vertical as well as horizontal.

Subdivided bar diagrams should not be used where the number of components is more than 10 or 12, for, in that case, the diagram would be overloaded with information which cannot be easily compared and understood.

The component bar diagrams can be used to represent either the absolute data or distribution ratios such as percentage distribution ratios is, in fact, an excellent method for presenting a set of distribution ratios diagrammatically.*

Illustration 7. Represent the following data by subdivided bar diagram.

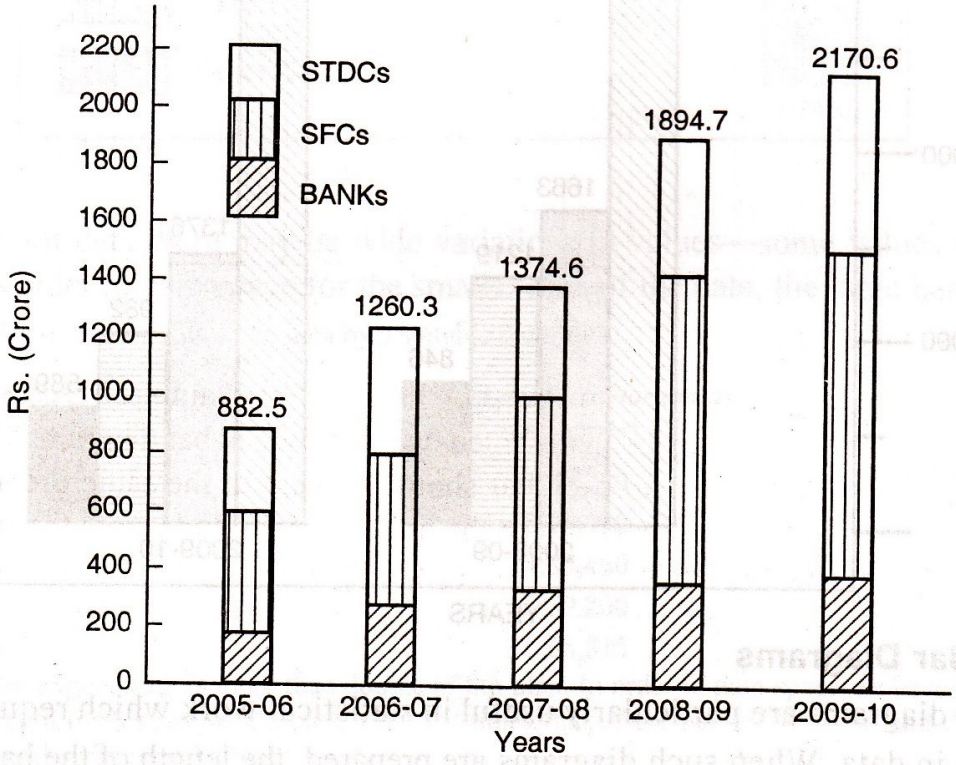
INSTITUTIONWISE ASSISTANCE SANCTIONED UNDER REFINANCE SCHEME (Rs. Crore)

Year	*Banks*	*SFCs*	*STDCs*	*Total*
2005-06	233.8	365.3	283.4	882.5
2006-07	301.8	484.7	473.8	1260.3
2007-08	303.2	668.6	402.8	1374.6
2008-09	365.3	992.8	536.6	1894.7
2009-10	416.4	1067.4	686.8	2170.6

*The other alternatives for this purpose are the relative pie diagram and the relative component line chart. The latter can be used only in cases where the classification is chronological. When the number of time period is not large, the bar chart is undoubtedly superior to these diagrammatic methods.

Solution. Since we have to show three different variables, subdivided bar diagram will be more appropriate. In order to prepare such a diagram, bar is to be drawn of the total of all the three heads for each year and then it is to be subdivided in three heads. Thus for 2005-06, total is 882.5; for 2006-07, 1260.3, etc.

DISTRIBUTION OF ASSISTANCE SANCTIONED
UNDER REFINANCE SCHEME (Rs. Crore)
FROM 2005-06 TO 2009-10



(c) Multiple Bar Diagrams

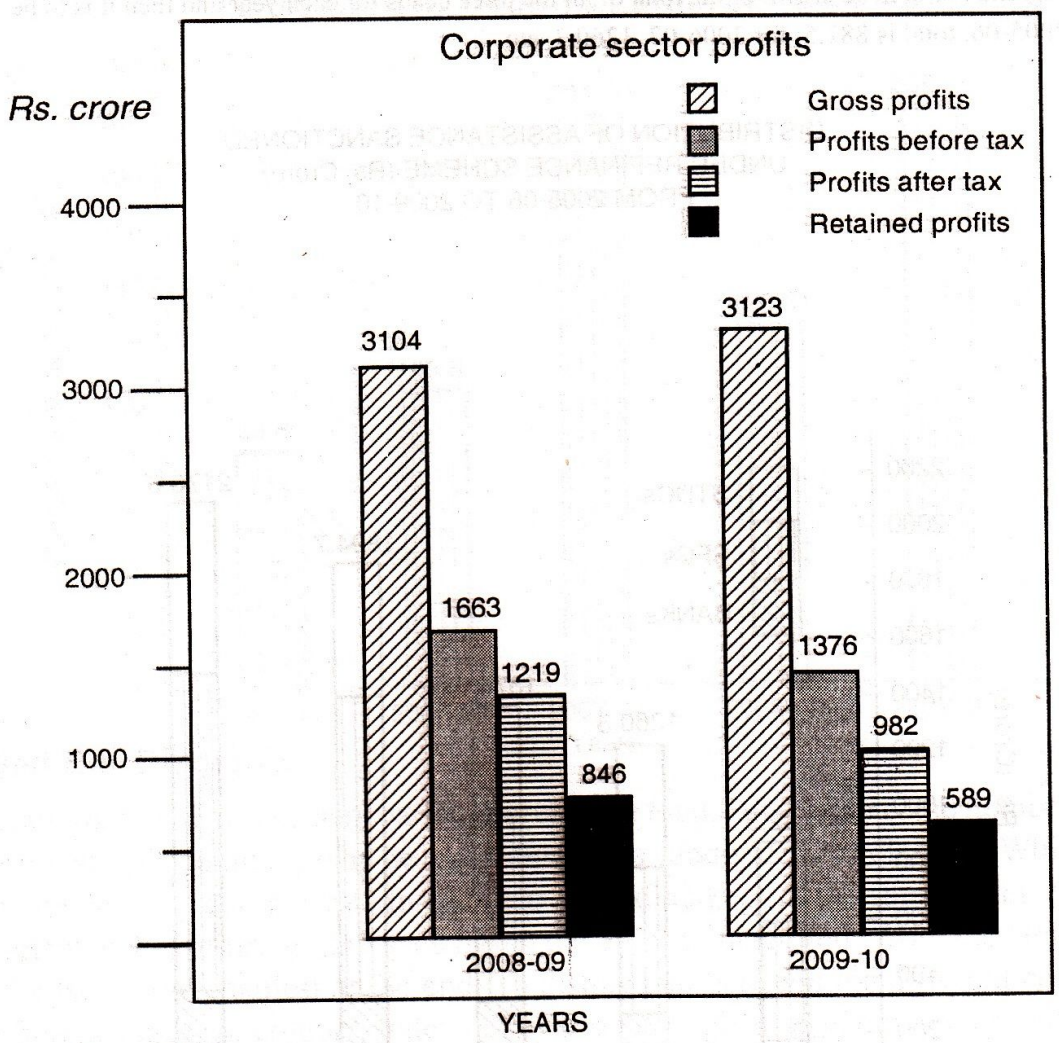
In multiple bar diagram two or more sets of interrelated data are represented. The technique of drawing such a diagram is the same as that of simple bar diagram. The only difference is that since more than one phenomenon is represented, different shades, colours, dots, or crossings are used to distinguish between the bars.

Illustration 8. Represent the following data by multiple bar diagram :

CORPORATE SECTOR PROFITS (Rs. crore)

	2008-09	2009-10
Gross profits	3104	3123
Profits before tax	1663	1376
Profits after tax	1219	982
Retained profits	846	589

Solution. The multiple bar diagram of the above data is given below :



(*d*) Percentage Bar Diagrams

Percentage bar diagrams are particularly useful in statistical work which requires the portrayal of relative changes in data. When such diagrams are prepared, the length of the bars is kept equal to 100 and segments are cut in these bars to represent the components (percentages) of an aggregate.

(*e*) Deviation Bars

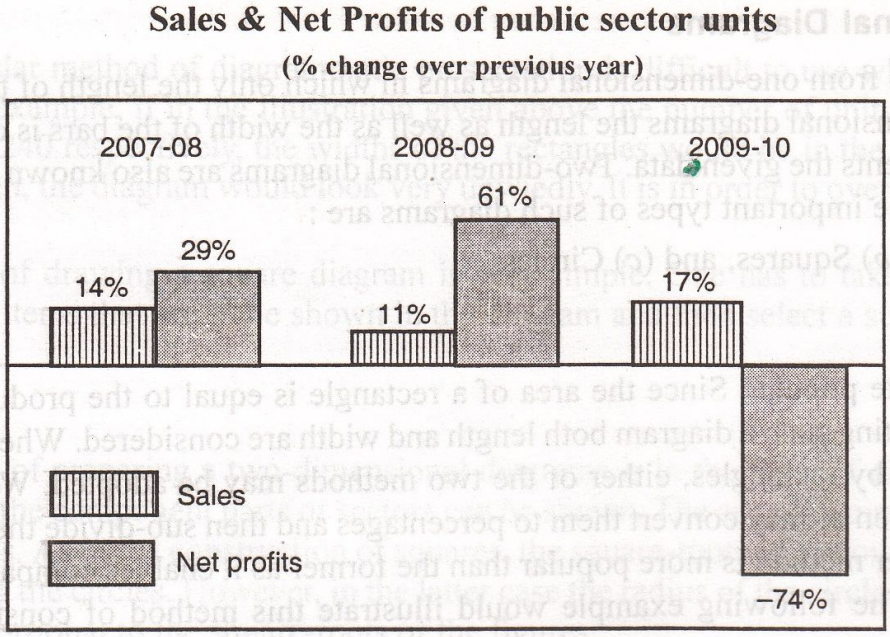
Deviation bars are popularly used for representing net quantities—excess or deficit, *i.e.*, net profit, net loss, net exports or imports, etc. Such bars can have both positive and negative values. Positive values are shown above the base line and negative values below it. The following illustration would explain this type of diagram :

Illustration 9. The following are the figures of sales and net profits of public sector units over the last three years. Represent the data by a suitable diagram.

(*% change over previous year*)

Year	*Sales*	*Net profits*
2007-08	14%	29%
2008-09	11%	61%
2009-10	17%	–74%

Solution. The above data can best be represented by deviation bars.



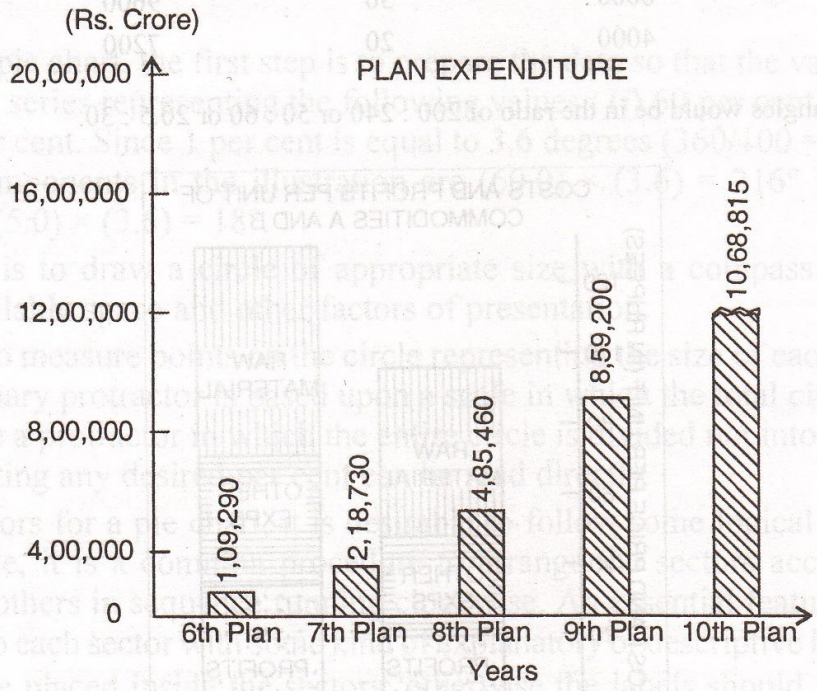
Broken Bars

In certain type of data there may be wide variations in values—some values may be very small, other very large. In order to gain space for the smaller bars of the data, the large bars may be broken.

Illustration 10. Represent the following data by a suitable diagram :

Five-Year Plan	*Actual public sector outlay (Rs. crore)*
6th	1,09,290
7th	2,18,730
8th	4,85,460
9th	8,59,200
10th	10,68,815

Solution. 10th Plan expenditure is more than double of 8th plan. In order to gain space we have broken the bar for 10th Plan. Otherwise, the length of this bar would have almost 2 times that of the bar for 9th Plan and the diagram would have occupied a lot of space and given an ugly look.



* Estimate.

II. Two-dimensional Diagrams

As distinguished from one-dimensional diagrams in which only the length of the bars is taken into account, in two-dimensional diagrams the length as well as the width of the bars is considered. Thus the area of the bar represents the given data. Two-dimensional diagrams are also known as *surface diagrams* or *area diagrams*. The important types of such diagrams are :

(*a*) Rectangles, (*b*) Squares, and (*c*) Circles.

(*a*) Rectangles

This form is quite popular. Since the area of a rectangle is equal to the product of its length and width, while constructing such a diagram both length and width are considered. When two sets of figures are to be represented by rectangles, either of the two methods may be adopted. We may represent the figures as they are given or may convert them to percentages and then sub-divide the length into various components. The latter method is more popular than the former as it enables comparison to be made on a percentage basis. The following example would illustrate this method of constructing rectangular diagrams :

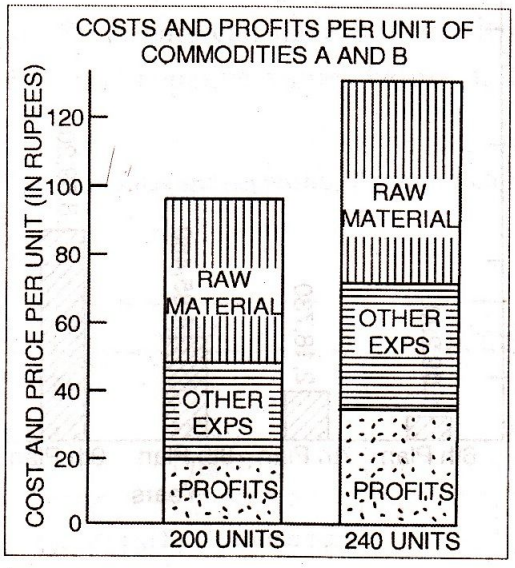
Illustration 11. Present the following data by a rectangular diagram :

	Commodities	
	A	*B*
Price per unit of commodity (Rs.)	100	120
Quantity sold	200	240
Cost of raw materials used (Rs.)	10000	12000
Other costs (Rs.)	6000	9600
Profit (Rs.)	4000	7200

Solution. Let us calculate the cost of raw materials, other expenses and profits per unit.

	Commodity A 200 units		*Commodity B 240 units*	
	Total (Rs.)	*Per unit (Rs.)*	*Total (Rs.)*	*Per unit (Rs.)*
Cost of raw material	10000	50	12000	50
Other expenses	6000	30	9600	40
Profit	4000	20	7200	30

The widths of the rectangles would be in the ratio of 200 : 240 or 50 : 60 or 20.5 : 30.



(b) Squares

The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. For example, if in the illustration given above the number of units sold of commodities *A* and *B* are 20 and 240 respectively, the widths of the rectangles would be in the ratio of 5 : 60 or 1 : 12. If this ratio is taken, the diagram would look very unwiedly. It is in order to overcome this difficulty that squares are used.

The method of drawing a square diagram is very simple. One has to take the square-root of the values of various items that are to be shown in the diagram and then select a suitable scale to draw the squares.

(c) Circles

Another way of preparing a two-dimensional diagrams is in the form of circles. In such diagrams both the total and the component parts or sectors can be shown. The area of a circle is proportional to the square of its radius. As in the construction of squares, the square-roots of various figures are worked out while constructing the circles. However, in the latter case the radius of the circles (rather than the side of squares) are proportional to the square-roots of the figures.

Circles can be used in all those cases in which squares are used. However, in both these types of diagrams it is difficult to judge the relative magnitude with precision.

Circles are difficult to compare and as such they are not very popular in statistical work. When it is necessary to use circles, they should be compared on an area basis rather than on a diameter basis as the diameter basis is very misleading. Compared to rectangles, circles are more difficult to construct and interpret.

Pie Diagram

This type of diagram enables us to show the partitioning of a total into component parts. A very common use of the pie chart is to represent the division of a sum of money into its components. For example, the entire circle, or pie, may represent the budget of a family for a month and the sections may represent portions of the budget alloted to rent, food, clothing, and so on. Similarly, through a pie diagram we can show how a rupee spent by a firm is distributed over various heads such as wages, raw materials, administration expenses, etc.

The pie chart is so called because the entire graph looks like a pie, and the components resemble slices cut from it.

In constructing a pie chart, the first step is to prepare the data so that the various component values can be transposed in a series representing the following values : (*i*) 60 per cent, (*ii*) 25 per cent, (*iii*) 10 per cent, and (*iv*) 5 per cent. Since 1 per cent is equal to 3.6 degrees (360/100 = 3.6), the corresponding values of the four components in the illustration are (60.0) × (3.6) = 216° ; (25.0) × (3.6) = 90° ; (10.0) × (3.6) = 36° ; (5.0) × (3.6) = 18°.

The second step is to draw a circle of appropriate size with a compass. The size of the radius depends upon the available space and other factors of presentation.

The third step is to measure points on the circle representing the size of each sector with the help of a protractor. The ordinary protractor is based upon a scale in which the total circle is 360 degree, but it is possible to purchase a protractor in which the entire circle is divided not into 360 but 100 equal parts so that angle representing any desired per cent can be read directly.

In laying out sectors for a pie chart, it is desirable to follow some logical arrangement, pattern or sequence. For example, it is a common procedure to arrange the sectors according to size, with the largest at the top and others in sequence running clockwise. An essential feature of the pie chart is the careful identification to each sector with some kind of explanatory or descriptive label. If there is sufficient room the labels can be placed inside the sectors; otherwise the labels should be placed in continuous positions outside the circle, usually with an arrow pointing to the appropriate sector.

Illustration 12. The following funds in assistance were sanctioned and disbursed during 2002-03 to 2009-10 by a leadi financial institution.

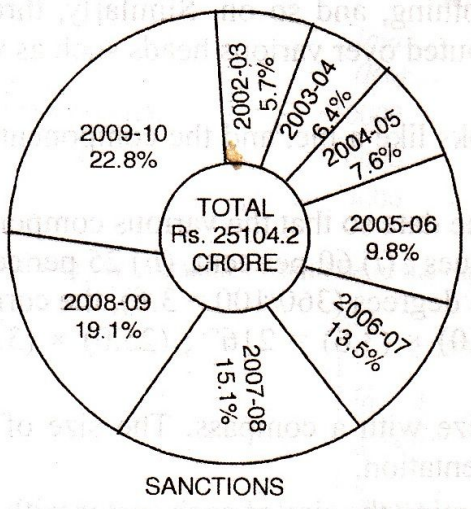
Year	*Amount sanctioned (Rs. crore)*	*Amount disbursed (Rs. crore)*
2002-03	1444.3	1066.1
2003-04	1607.0	1339.5
2004-05	1905.5	1582.4
2005-06	2449.1	1961.5
2006-07	3394.6	2293.4
2007-08	3784.4	2787.0
2008-09	4791.5	3258.0
2009-10	5727.8	4041.6

Represent the data by a pie diagram.

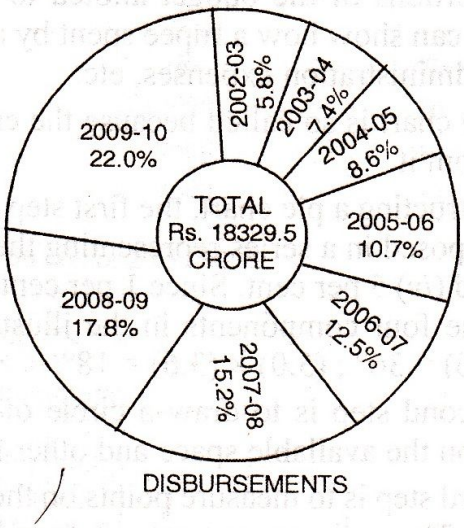
Solution. Convert the given data into percentages and then prepare two pie diagrams and divide them into segments.

Year	*Amount sanctioned (Rs. crore)*	*%*	*Amount disbursed (Rs. crore)*	*%*
2002-03	1444.3	5.7	1066.1	5.8
2003-04	1607.0	6.4	1339.5	7.4
2004-05	1905.5	7.6	1582.4	8.6
2005-06	2449.1	9.8	1961.5	10.7
2006-07	3394.6	13.5	2293.4	12.5
2007-08	3784.4	15.1	2787.0	15.2
2008-09	4791.5	19.1	3258.0	17.8
2009-10	5727.8	22.8	4041.6	22.0
Total	25104.2	100.0	18329.5	100.0

Composition of Trends in Assistance Sanctioned and Disbursed during the year 2002-03 to 2009-10.



SANCTIONS



DISBURSEMENTS

Limitations of Pie Diagrams

Pie diagrams are less effective than bar diagrams for accurate reading and interpretation, particularly when series are divided into a large number of components or the difference among the components is very small. It is generally inadvisable to attempt to portray a series of more than five or six categories by means of a pie chart. If, for example, there are eight, ten or more categories it may be very confusing to differentiate the relative values portrayed especially when the several small sectors are of approximately the same size. This type of diagram, although frequently used, appears upon comparison inferior to simple bar diagram the divided bar diagram or a group of curves.

III. Pictograms and Cartograms

(i) Pictograms

Pictograms, also known as picturegrams, are very popularly used in presenting statistical data. They are not abstract presentations such as lines or bars, but really depict the kind of data we are dealing with. Pictures are attractive and easy to comprehend and as such this method is particularly useful in presenting statistics to the layman. When pictograms are used, data are represented through a pictorial symbol that is carefully selected. The picture symbol should be self-explanatory in nature, *i.e.*, it should represent clearly the phenomena. For example, if the increase in number of buses on road is shown over a period of time the appropriate symbol would be a bus.

Illustration 13. The following table gives the production of tea in India by a leading company :

Years	2006	2007	2008	2009	2010
Production (Million kgs)	421	561	587	645	660

Represent the data by a pictogram.

Solution. For representing the above data by a pictogram we will use the symbol of a cup.

PICTOGRAM

PRODUCTION OF TEA	
YEAR	(million kgs)
2006	421
2007	561
2008	587
2009	645
2010	660

Merits. (1) As compared with other types of diagrams, pictograms have a greater attraction value and, therefore, where the attention of masses is to be drawn such as in exhibitions, fairs, etc., they are very popularly used. They stimulate interest in the information being presented.

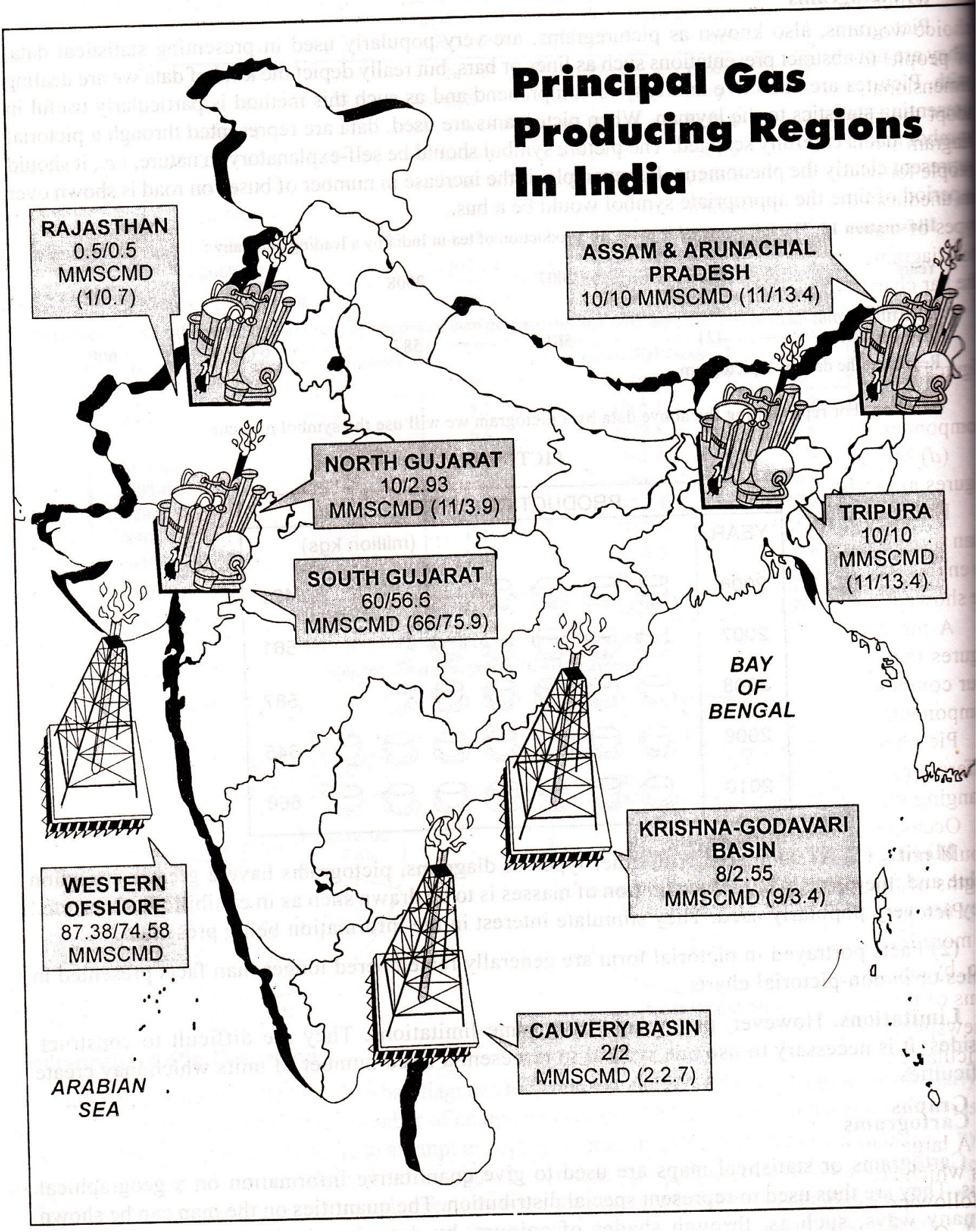
(2) Facts portrayed in pictorial form are generally remembered longer than facts presented in tables or in non-pictorial charts.

Limitations. However, pictograms have some limitations. They are difficult to construct. Besides, it is necessary to use one symbol to represent a fixed number of units which may create difficulties.

(ii) Cartograms

Cartograms or statistical maps are used to give quantitative information on a geographical basis. They are thus used to represent special distribution. The quantities on the map can be shown in many ways, such as, through shades of colours, by dots, by placing pictograms in each geographical unit and by placing the appropriate numerical figure in each geographical unit.

The cartogram shows the principal gas producing regions in india.



CHOICE OF A SUITABLE DIAGRAM

Which diagram out of several ones to select in a given situation is a ticklish problem. The choice would primarily depend upon two factors, namely : (*i*) the nature of the data : and (*ii*) the type of people for whom the diagram is meant. The nature of data would depend whether to use one-dimensional, two-dimensional or three-dimensional diagram, and if it is one-dimensional, whether to adopt the simple bar or sub-divided bar, multiple bar or some other type. As already stated, a cubic diagram would be preferred to a bar if the magnitudes of the figures are very wide apart. The type of people for whom the diagram is intended must also be considered. For example, for drawing attention of an uneducated mass pictograms and cartograms are more effective than cubes, circles, etc. Different types of diagrams such as bars rectangles, cubes, pictograms, pie charts have specific uses. However, bar diagrams are most popular in practice. There are different types of bars and the appropriate type of the bar chart can be decided on the following basis :

(*a*) Simple bar charts should be used where changes in totals are required to be conveyed.

(*b*) Component bar charts are more useful where changes in totals as well as in the size of component figures (absolute ones) are required to be displayed.

(*c*) Percentage sub-divided bar charts are better suited where changes in the relative size of component figures are to be exhibited.

(*d*) Multiple bar charts should be used where changes in the absolute values of the component figures are to be emphasised and the overall total is of no importance.

However, multiple and component bar charts should be used only when there are not more than three or four components, as a large number of components make the bar charts too complex to enable worthwhile visual impression to be gained. When a large number of components have to be shown a pie chart is more suitable.

A pie chart is particularly useful where it is desired to show the relative proportion of the figures that go to make up a single overall total. Unlike bar charts it is not restricted to three or four component figures although its effectiveness tends to dwindle with more than seven or eight components.

Pie charts cannot be used effectively where a series of figures is involved, as a number of different pie charts are not easy to compare. Nor should changes in one overall total be shown by changing the size of the 'pie'.

Occasionally, circles are used to represent size. But it is difficult to compare them and they should not be used when it is possible to use bars. This is because it is easier to compare the lengths of lines or bars than to compare areas or volumes.

Pictograms and cartograms are very elementary forms of visual presentation. However, they are more informative and more effective than other forms for presenting data to the general public who, by and large, neither possess much ability to understand nor take interest in the less attractive forms of presentation. The pictogram is admirably suited to the illustrations of exhibits or articles in newspapers and magazines or for dressing up annual reports. Cartograms or statistical maps are particularly effective in bringing out the geographical pattern that may be concealed in the data.

(iii) Graphs

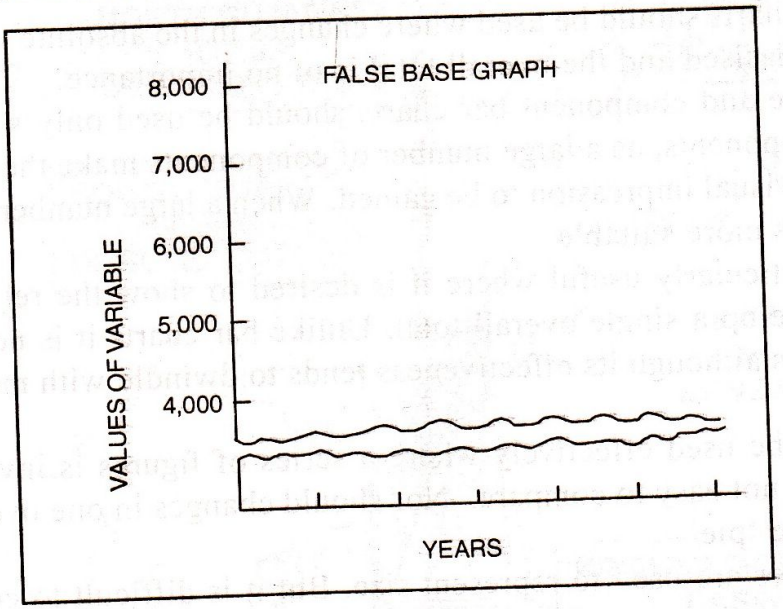
A large variety of graphs are in practical use. However, we shall discuss only some important ones which are more popularly used in practice. Broadly, the various graphs can be divided under the following two heads :

1. Graphs of time series or line graphs.
2. Graphs of frequency distributions.

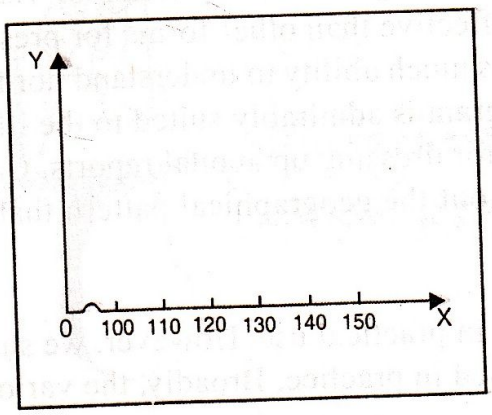
GRAPHS OF TIME SERIES OR LINE GRAPHS

When we observe the values of a variable at different points of time the series so form known as time series. The technique of graphic presentation is extremely helpful in anal changes at different points of time. On the X-axis we generally take the time and on the Y-axi value of the variable and join the various points by straight lines. The graph so formed is know line graph. Such graphs are most widely used in practice. They are simplest to understand, ea to make and most adaptable to many uses. Also several variables can be shown on the same g and a comparison can be made.

One of the fundamental rules while constructing graphs is that the scale on the Y-axis sh begin from zero even if the lowest Y-figures associated with any X-period or value is far a zero. However, if this rule is strictly followed the curve would be very much pulled up towards right, *i.e.*, away from the point of origin. When the gap between zero and smallest value of variable is large, for example, if the variable starts from 50,000, a lot of space would be requ to show the variable. It is in order to solve this difficulty that the use of false base is made. Wh a false base has been used, the space between zero and the smallest value of the variable is omitted To bring out this fact clearly that the false base has been used, two zig-zag horizontal lines a drawn above the base line as shown in the graph below :



Just as we have talked of Y-axis starting from zero, we also talk of X-axis starting from zer To represent false base on the X-axis, we draw a kinked line as follows :



It is clear from above that a considerable saving in space is possible in case the variable sta from a value much away from 0. It may, however, be noted there is a growing feeling that there

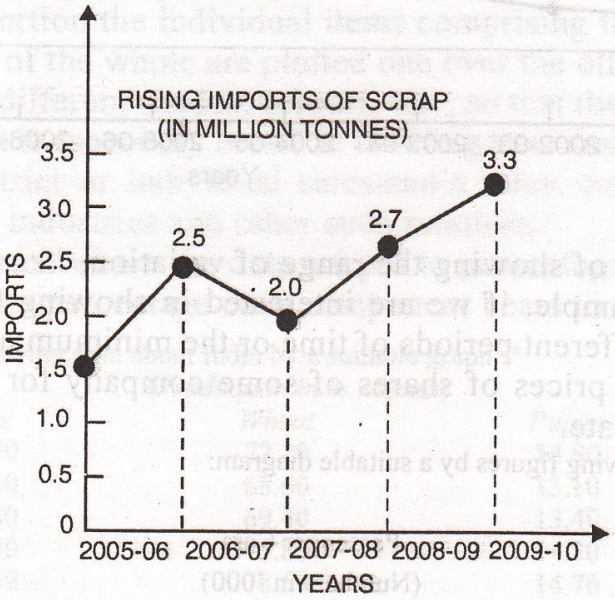
sanctity in X-axis and Y-axis starting from zero, they can well start from the lowest value or near about from the data given. If that is so the false base line and kinked line need not be used.

Illustration 14. The following data relate to the imports of scrap of Jharkhand Sponge Iron Limited :

Year	:	2005-06	2006-07	2007-08	2008-09	2009-10
Imports (in million tonnes)	:	1.5	2.5	2.0	2.7	3.3

Represent the data graphically.

Solution. The given data can be represented by a graph as shown below :



If the unit of measurement is the same, we can represent two or more variables on the same graph. This facilitates comparison. However, when the number of variables is very large (say, exceeding five or six) and they are all shown on the same graph the chart becomes quite confusing because different lines may cut each other and make it difficult to understand the behaviour of the variables. Therefore, for the sake of clarity we should not represent more than 5 or 6 variables on the same graph. When two or more variables are shown on the same graph, it is desirable to use thick, thin, broken, dotted lines, etc., to distinguish between the various variables.

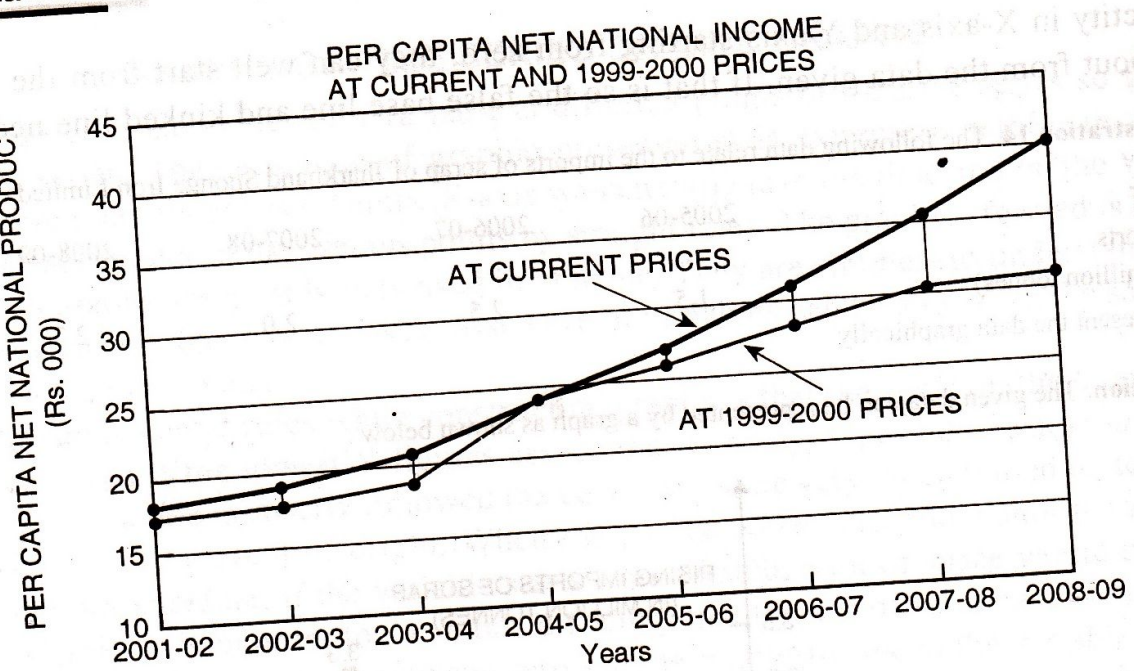
Illustration 15. Represent the following data by a suitable graph.*

PER CAPITA NATIONAL PRODUCT* (Rs.)

Year	*At Current Prices*	*At 1999-2000 Prices*
2001-02	17782	16769
2002-03	18885	17109
2003-04	20871	18301
2004-05	24095	24095
2005-06	27183	25969
2006-07	31080	28074
2007-08	35430	30316
2008-09	40141	31821

*Source : Economic Survey 2009-10, Govt. of India A-3.

Solution.



Range Chart

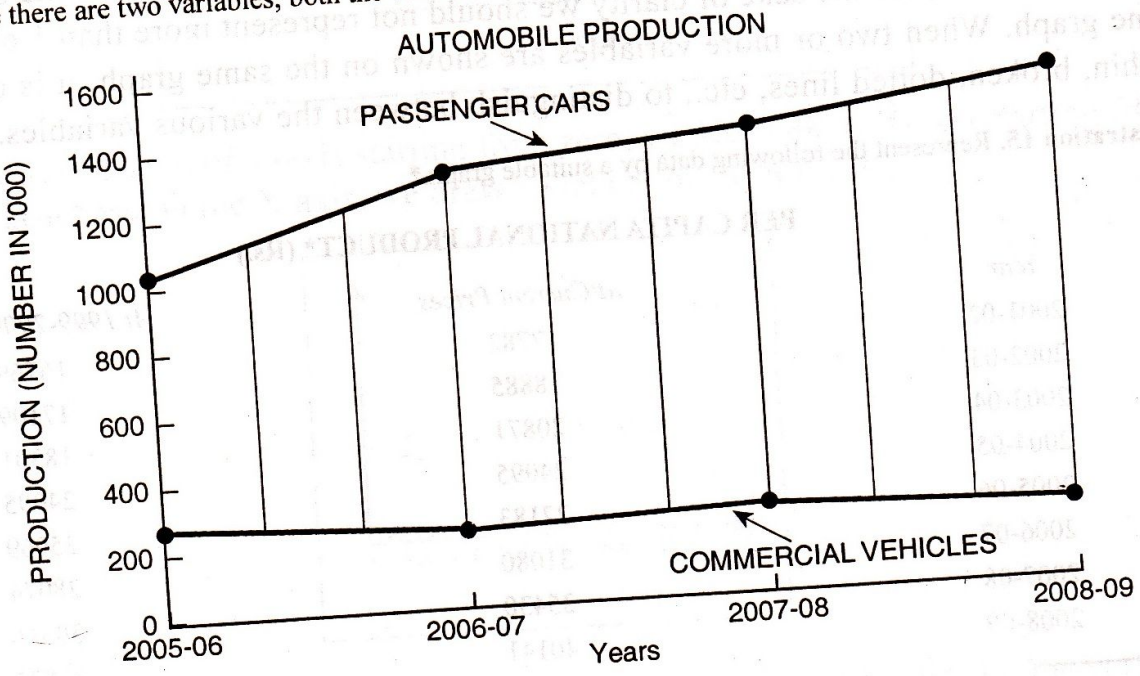
It is a very good method of showing the range of variation, *i.e.*, the minimum and maximum values of a variable. For example, if we are interested in showing the minimum and maximum prices of a commodity for different periods of time or the minimum and maximum temperature or the minimum and maximum prices of shares of some company for different periods, the range chart would be very appropriate.

Illustration 16. Show the following figures by a suitable diagram:

Automobile Production

Year	Passenger Cars (Numbers in '000)	Commercial Vehicles (Numbers in '000)
2005-06	1046	263
2006-07	1323	222
2007-08	1426	246
2008-09	1517	218

Solution. Since there are two variables, both the variables we will show on the same graph.



* *Economic Survey* : Govt. of India, 2008-09 p. 212

The following are the steps in constructing such a chart :

1. Take the time on X-axis and the variable on the Y-axis.
2. Draw two curves by plotting the given data–one curve representing the highest values and the other one lowest values. In the given case, curve A represents lowest prices, whereas curve B highest prices. The gap between curves A and B represents the range of variation.
3. For highlighting the difference between the lowest and highest values the use of some colour or shade, etc., should be made.

Band Graphs

A band graph is a type of line graph which shows the total for successive time periods broken up into sub-totals for each of the component parts of the total. In other words, the band graph shows how and in what proportion the individual items comprising the aggregate are distributed. The various component parts of the whole are plotted one over the other and the gaps between the successive lines are filled by different shades, colours, etc., so that the chart has the appearance of a series of bands. Such a chart is especially useful in dividing total costs into component cost, total sales into department or district or individual salesman's sales, total production by nature of commodity, States, plants, or industries and other such relations.

Band graph can also be used where the data are put to percentage form ; the whole chart will depict 100% and the bands, the percentage each component bears to whole.

Illustration 17. Present the following data about India by a suitable graph :*

(*Production in m. tonnes*)

Year	*Rice*	*Wheat*	*Pulses*	*Coarse cereals*
2003-04	88.50	72.20	14.90	37.60
2004-05	81.10	68.60	13.10	33.50
2005-06	91.80	69.40	13.40	34.10
2006-07	93.40	75.80	14.20	33.90
2007-08	96.69	78.57	14.76	40.76
2008-09	99.15	80.58	14.66	39.48

Solution. The above data can be most suitably presented through a band graph. The procedure of constructing such a graph is as follows :

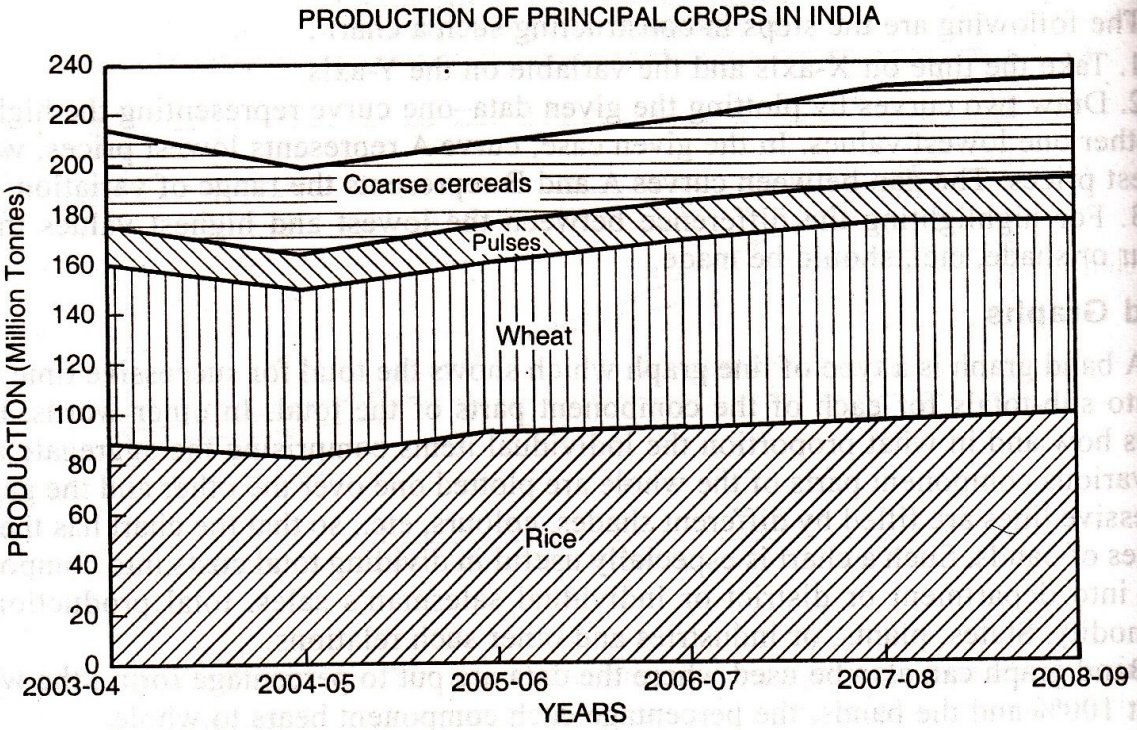
1. Take the years on the X-axis and the variable on the Y-axis.
2. Plot the various points for different years for rice and join them by straight lines. This is represented by curve *A*.
3. Add the figures of rice for various years to the figures of wheat and plot the points and join them by straight lines. This is represented by curve *B*. The difference between the two curves, *i.e.*, *B* and *A*, gives the production of wheat.
4. Add the figures of rice and wheat to pulses and plot the various points. This is represented by curve *C*. The difference between curve *C* and curve *B* represents production of the pulses.
5. Add the figures of rice, wheat and pulses to other cereals, and draw a curve. This is represented by curve *D*. The difference between curve *D* and curve *C* gives the production figures for other cereals.

GRAPHS OF FREQUENCY DISTRIBUTIONS

A frequency distribution can be presented graphically in any of the following ways :

1. Histogram.
2. Frequency polygon.
3. Smoothed frequency curve.
4. Cumulative frequency curves or 'Ogives'.

* *Economic Survey* : Govt. of India, 2009-10, p. 181.



1. Histogram

Out of several methods of presenting a frequency distribution graphically, histogram or the column diagram, as it is sometimes called, is the most popular and widely used in practice. The statistical meaning of histogram is that it is a graph that represents the class frequencies in a frequency distribution by vertical adjacent rectangles. A histogram is a graphical method for presenting data, where the observations are located on a horizontal axis (usually grouped into intervals) and the frequency of those observations is depicted along the vertical axis.

While constructing histogram the variable (class interval) is always taken on the X-axis and the frequencies depending on it on the Y-axis. Each class is then represented by a distance of the scale that is proportional to its class-interval. The distance for each rectangle on the X axis shall remain the same in case the class-intervals are uniform throughout ; if they are different the width of the rectangles shall also vary. The Y-axis represents the frequencies of each class which constitute the height of its rectangle. In this manner, we get a series of rectangles each having a class-interval distance as its width and the frequency distance as its height. The area of the histogram represents the total frequency as distributed throughout the classes.

The histogram should be clearly distinguished from a bar diagram. The distinction lies in the fact that whereas a bar diagram is one-dimensional, *i.e.*, only the length of the bar is material and not the width ; a histogram is two-dimensional, that is, in a histogram both the length as well as the width are important.

The histogram is most widely used for graphical presentation of a frequency distribution. However, we cannot construct a histogram for distributions with open-end classes. Moreover, a histogram can be quite misleading if the distribution has unequal class-intervals and suitable adjustments in frequencies are not made.

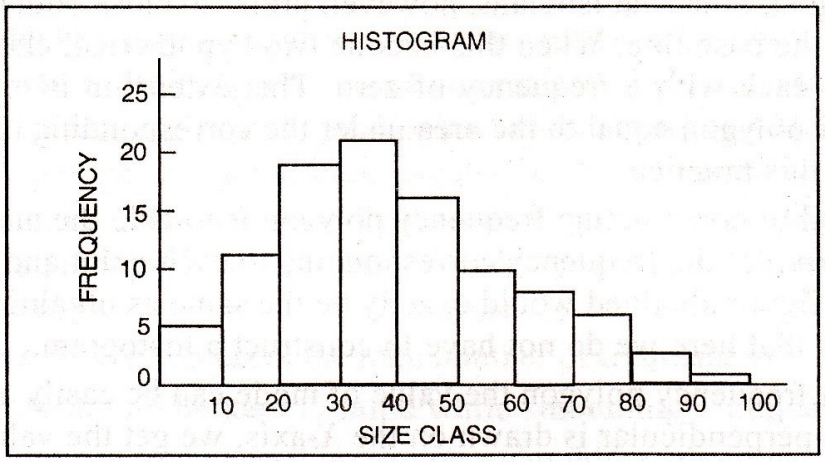
The technique of constructing histogram is now illustrated (*i*) for distributions having equal class-intervals, and (*ii*) for distributions having unequal class-intervals.

Construction of Histogram when Class-Intervals are Equal. When class-intervals are equal, take frequency on the Y-axis, the variable on the X-axis and construct adjacent rectangles. In such a case, the heights of the rectangles will be proportional to the frequencies.

Illustration 18. Represent the following data by a histogram :

Size class	*Frequency*	*Size class*	*Frequency*
0—10	5	50—60	10
10—20	11	60—70	8
20—30	19	70—80	6
30—40	21	80—90	3
40—50	16	90—100	1

Solution. The histogram of the above data is given below.

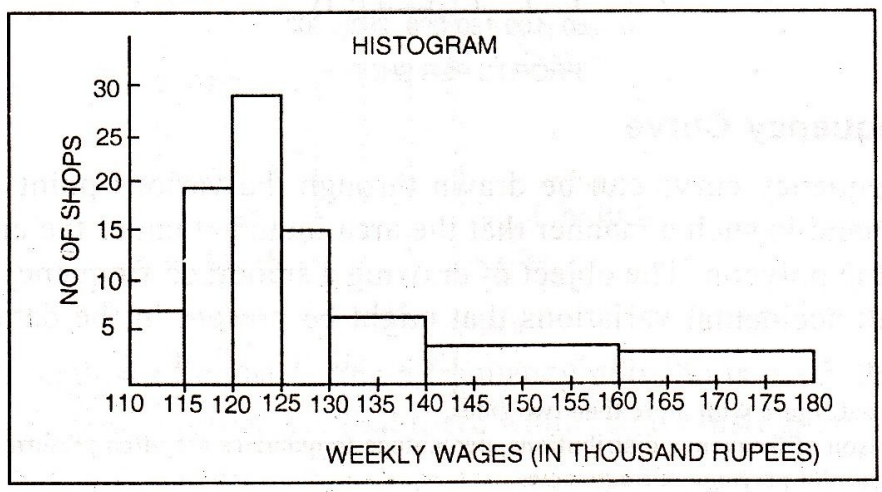


Construction of Histogram when Class-Intervals are Unequal. When class-intervals are unequal, the frequencies must be adjusted before constructing the histogram. For making the adjustment we take that class which has the lowest class-interval and adjust the frequencies of other classes in the following manner. If one class-interval is twice as wide as the one having lowest class-interval, we divide the height of its rectangle by two ; if it is three times more, we divide the height of its rectangles by three, etc., *i.e., the heights will be proportional to the ratios of the frequencies to the width of the classes.* Great care is needed for construction of histograms with unequal class-interval widths. Quite often they are illconstructed and misinterpreted. This would be clear from the following example :

Illustration 19. Represent the following data by means of a histogram :

Weekly profits (in 000's Rs.)	*No. of shops*	*Weekly profits (in 000's Rs.)*	*No. of shops*
110—115	7	130—140	12
115—120	19	140—160	12
120—125	27	160—180	8
125—130	15		

Solution. Since the class-intervals are unequal, frequencies must be adjusted otherwise the histogram would give a misleading picture. The adjustment is done as follows : The lowest class-interval is 5. The frequency of the class 130–140 shall be divided by two since the class-interval is double, that of 140–160 by 4, and so on.



2. Frequency Polygon*

A frequency polygon is a graph of frequency distribution. It has more than four sides. It particularly effective in comparing two or more frequency distributions. There are two ways which a frequency polygon may be constructed :

1. We may draw a histogram of the given data and then join by straight lines the mid-points the upper horizontal side of each rectangle with the adjacent rectangle. The figure so formed called frequency polygon. Some statisticians, however, prefer to close both the ends of the polygon by extending them to the base line. When this is done two hypothetical classes at each end would have to be included—each with a frequency of zero. This extension is made with the object making the area under polygon equal to the area under the corresponding histogram. The students are advised to follow this practice.

2. Another method of constructing frequency polygon is to take the mid-points of the various class-intervals and then plot the frequency corresponding to each point and to join all these points by straight lines. The figure obtained would exactly be the same as obtained by the other method. The only difference is that here we do not have to construct a histogram.

By constructing a frequency polygon the value of mode can be easily ascertained. If from the apex of the polygon a perpendicular is drawn on the X-axis, we get the value of mode. Moreover, frequency polygons facilitate comparison of two or more frequency distributions on the same graph.

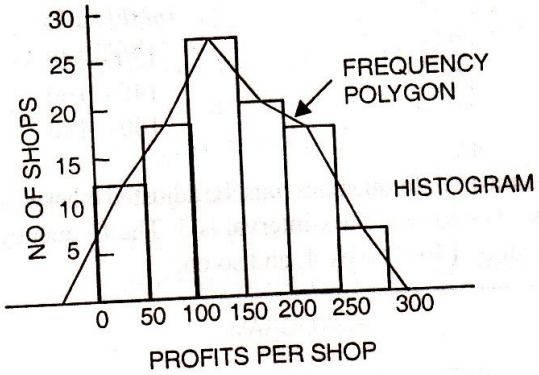
Frequency polygon has a special advantage over the histogram. The frequency polygons of several distributions may be plotted on the same axis, thereby making certain comparisons possible whereas histograms cannot be usefully employed in the same way. To compare histograms, we must have a separate graph for each. Because of this limitation for purposes in making a graphic comparison of frequency distribution, frequency polygons are preferred.**

Illustration 20. The daily profits (in thousand rupees) of 100 shops are distributed as follows :

Daily Profits :	0—50	50—100	100—150	150—200	200—250	250—300
No. of shops :	12	18	27	20	17	6

Prepare a histogram and frequency polygon of the above data.

Solution. The histogram and frequency polygon of the above data are shown in the diagram below :



3. Smoothed Frequency Curve

A smoothed frequency curve can be drawn through the various points of the polygon. The curve is drawn freehand in such a manner that the area included under the curve is approximately the same as that of the polygon. The object of drawing a smoothed frequency curve is to eliminate as far as possible all accidental variations that might be present in the data. While smoothing a

*"Polygon" is a closed figure with more than two sides.

** To make comparison of frequency distributions, percentage frequencies are often preferred. Accordingly, to compare frequency polygons we can plot percentage frequencies.

frequency polygon the fact that it is really derived from the histogram should always be kept in mind. This would imply that the top of the curve would overtop the highest point of the polygon particularly when the magnitude of class-interval is large. The curve should look as regular as possible and all sudden turns should be avoided. The extent of smoothing would, however, depend upon the nature of the data. If it is a natural phenomenon like tossing of coins, smoothing may be freely resorted to as such phenomenon normally has symmetrical curves, but if the phenomenon is social or economic the curve is generally skewed and in such a case smoothing cannot be carried too far.

For drawing smoothed frequency curve it is necessary to first draw the polygon and then smooth it out. As discussed earlier, the polygon can be constructed even without first constructing a histogram by plotting the frequencies at the mid-points of class-intervals. This may save some time but the smoothing of the polygon cannot be done properly without a histogram. Hence, it is desirable to proceed in a sequence, *i.e.*, first to draw a histogram, then a polygon and lastly to smooth it to obtain the smoothed frequency curve. This curve should begin and end at the base line and, as a general rule, it may be extended to the mid-points of the class-intervals just outside the histogram. The area under the curve should represent the total number of frequencies in the entire distribution.

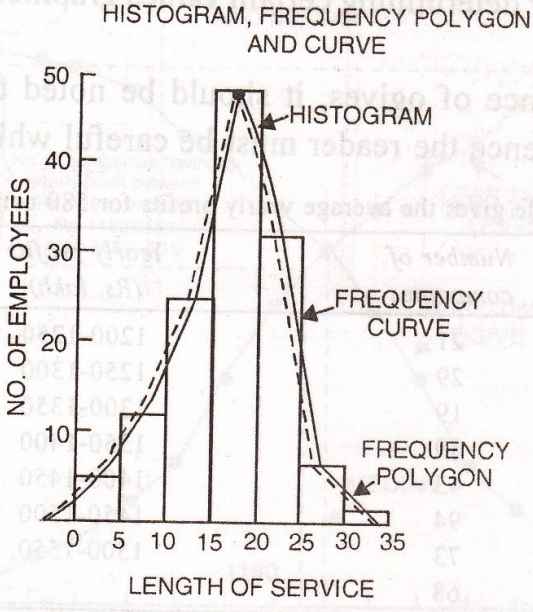
The following points should be kept in mind while smoothing a frequency graph :

1. Only frequency distributions based on samples should be smoothed.
2. Only continuous series should be smoothed.
3. The total area under the curve should be equal to the area under the original histogram or polygon.

Illustration 21. Draw a histogram, frequency polygon and frequency curve representing the following figures :

Length of Service (in years)	*No. of Employees*	*Length of Service (in years)*	*No. of Employees*
Less than 5	5	20—25	32
5—10	12	25—30	6
10—15	25	30—35	1
15—20	48		

Solution. The histogram, frequency polygon and frequency curve of the above data are shown in the following diagram :



When the second method of constructing a frequency polygon is used, the graph would take the same shape as above with a difference that there would be no histogram.

In the construction of frequency polygon, the same difficulties are faced as with histogram, they cannot be used for distributions having open-end classes and suitable adjustment as in case histogram is necessary when there are unequal class-intervals. To get the frequency curve join all points of polygon and smoothen it to obtain the frequency curve as shown in the diagram.

4. Cumulative frequency curves or 'Ogives'

Sometimes one needs to know the answers to questions like 'how many workers of a fact earn more than Rs. 1,500 per month' or 'how many workers earn less than Rs. 1,200 per mont To answer these questions it is necessary to add the frequencies. When frequencies are adde they are called cumulative frequencies. These frequencies are then listed in a table called a cumulat frequency table. The graph of such a distribution is called a cumulative frequency curve or *Ogive* (pronounced Ogive).

There are two methods of constructing ogive, namely:

(*a*) The 'less than' method, and

(*b*) The 'more than' method.

(*a*) *'Less than' method.* In the 'less than' method, we start with the upper limits of the class and go on adding the frequencies. When these frequencies are plotted, we get a rising curve.

(*b*) *'More than' method.* In the 'more than' method, we start with the lower limits of th classes and from the total frequencies we subtract the frequency of each class. When thes frequencies are plotted, we get a declining curve.

Utility of Ogives. From the standpoint of graphic presentation, the ogive is especially used for the following purposes:

1. To determine as well as to portray the number or proportion of cases above or below a given value.

2. To compare two or more frequency distributions. Generally there is less overlapping when comparing several ogives on the same grid than when comparing several simple frequency curves in this manner.

3. Ogives are also drawn for determining certain values graphically, such as median, quartiles deciles, etc.

Despite the great significance of ogives, it should be noted that they are not as simple to interpret as one may feel and hence the reader must be careful while using them.

Illustration 22. The following table gives the average yearly profits for 580 companies :

Yearly profits (Rs. lakh)	*Number of companies*	*Yearly profits (Rs. lakh)*	*Number of companies*
800-850	21	1200-1250	36
850-900	29	1250-1300	45
900-950	19	1300-1350	27
950-1000	39	1350-1400	48
1000-1050	43	1400-1450	21
1050-1100	94	1450-1500	12
1100-1150	73	1500-1550	5
1150-1200	68		

(*a*) Draw ogives by 'less than' and 'more than' method for the data given above.

(*b*) Compute Q_1, D_4 and P_{65} and verify your result with the help of the drawn ogive

(c) Find the number of companies whose yearly profits (in Rs. lakh) lie between Rs. 1180 and Rs. 1480

(MBA, GGSIP Univ., 2000, MBA, DU, 2002, 2007)

Solution. Let us first arrange the frequency distribution for 'less than' and 'more than' methods as given below :

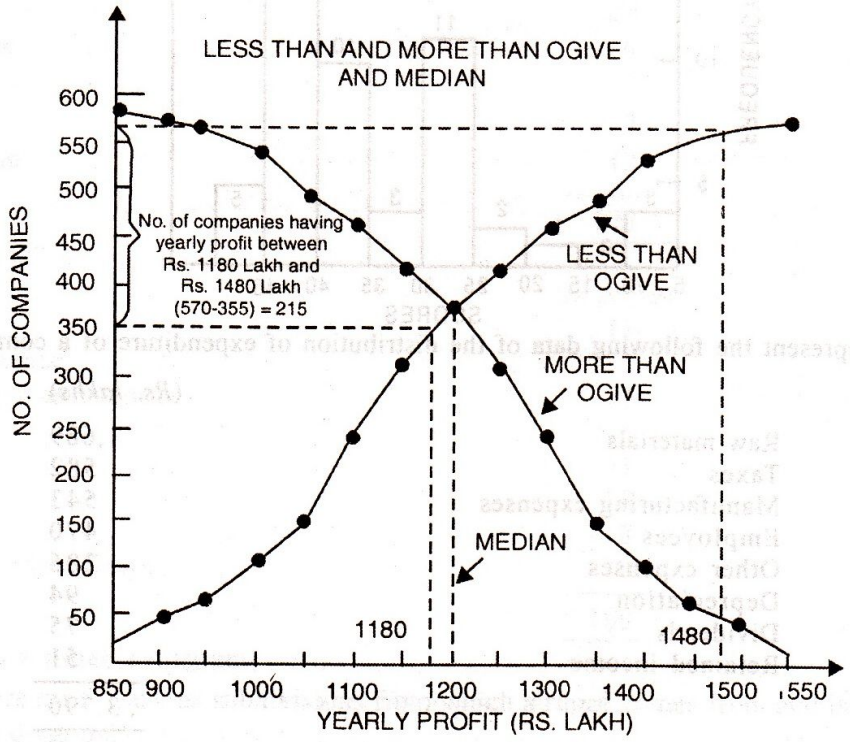
OGIVE BY LESS THAN METHOD

Yearly profits (Rs. lakh)	*Number of companies*	*Yearly profits (Rs. lakh)*	*Number of companies*
Less than 850	21	Less than 1250	422
" " 900	50	" " 1300	467
" " 950	69	" " 1350	494
" " 1000	108	" " 1400	542
" " 1050	151	" " 1450	563
" " 1100	245	" " 1500	575
" " 1150	318	" " 1550	580
" " 1200	386		

OGIVE BY MORE THAN METHOD

Yearly profits (Rs. lakh)	*Number of companies*	*Yearly profits (Rs. lakh)*	*Number of companies*
More than 800	580	More than 1200	194
" " 850	559	" " 1250	158
" " 900	530	" " 1300	113
" " 950	511	" " 1350	86
" " 1000	472	" " 1400	38
" " 1050	429	" " 1450	17
" " 1100	335	" " 1500	5
" " 1150	262		

With the help of these frequency distribution tables, we can draw ogives by less than and more than method as shown below :



It is clear from the above graph that there are 570 companies who are earning profits less than Rs. 1480 lakhs and there are 355 companies who are getting wages less than Rs. 1180 lakhs. Thus the number of companies earning profits between Rs. 1180 lakhs and Rs. 1480 lakhs is (570–355) = 215.

MISCELLANEOUS ILLUSTRATIONS

Illustration 23. You have conducted a market survey with a sample of size 50 regarding the acceptability of a new product which your company wants to launch. The scores of the respondents on the appropriate scale are as follows:

40	45	41	45	45	30	39	8	48	25
26	9	23	24	26	29	8	40	41	42
39	35	18	25	35	40	42	43	44	36
27	32	28	27	25	26	38	37	36	35
32	28	40	41	43	44	45	40	39	41

Prepare a frequency table and present the same information as a histogram. *(MBA, HPU, 2007)*

Solution. Frequency table of the given data is made as follows :

FREQUENCY DISTRIBUTION OF SCORES

Scores	*Tally*	*Frequency*
5-10	III	3
10-15	—	—
15-20	I	1
20-25	II	2
25-30	𝍸 𝍸 I	11
30-35	III	3
35-40	𝍸 𝍸	10
40-45	𝍸 𝍸 𝍸	15
45-50	𝍸	5
		Total 50

HISTOGRAM OF THE GIVEN DATA

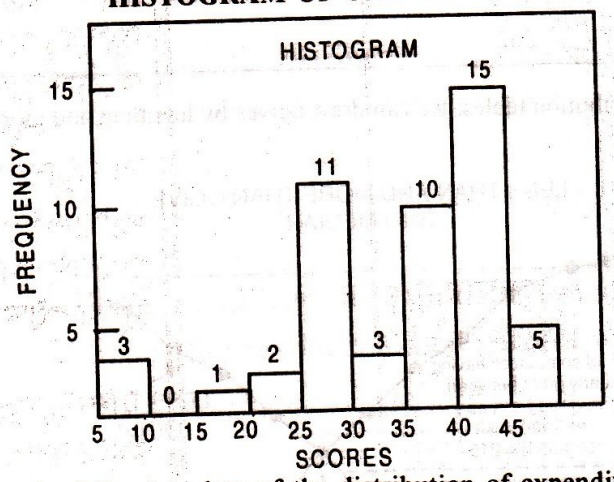


Illustration 24. Represent the following data of the distribution of expenditure of a company by suitable diagram:

	(Rs. lakhs)
Raw materials	1,689
Taxes	582
Manufacturing expenses	543
Employees	470
Other expenses	286
Depreciation	94
Dividends	75
Retained income	51
	3,790

(MBA, KU., 2005)

Solution. The above data can be represented by a pie diagram. Convert the given data in terms of percentage as follows:

Distribution	*Rs. lakhs*	*Percentage*
Raw materials	1,689	44.6
Taxes	582	15.4
Manufacturing expenses	543	14.3
Employees	470	12.4
Other expenses	286	7.5
Depreciation	94	2.5
Dividends	75	2.0
Retained income	51	1.3
Total	3,790	100.0

PIE DIAGRAM SHOWING THE DISTRIBUTION OF EXPENDITURE

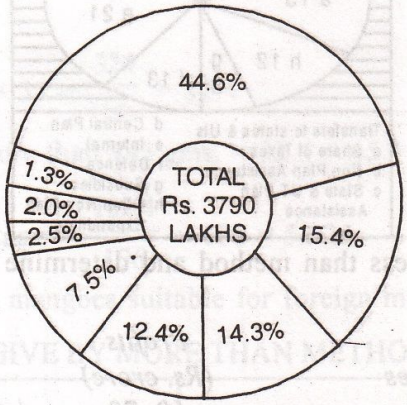


Illustration 25. The following data relate to the Central Budget 2007-08 How a Rupee comes from :

		(in paise)
(a)	Internal borrowings	20
(b)	External assistance	3
(c)	Other capital receipts	6
(d)	Borrowings from RBI	6
(e)	Customs	20
(f)	Excise	21
(g)	Corporation tax	5
(h)	Income tax	5
(i)	Other taxes	2
(j)	Non-tax revenue	12
		100

Where a Rupee goes to?

		(in paise)
(a)	Share of taxes	13
(b)	Non-plan assistance	7
(c)	State & U.T. plan assistance	11
(d)	Central Plan	15
(e)	Interest	21
(f)	Defence	13
(g)	Subsidies	8
(h)	Other non-plan expenditure	12
		100

Represent the data by a suitable diagram.

Solution. Since we are given various subdivisions from which a rupee comes from and is spent, suitable diagram would be a subdivided pie diagram.

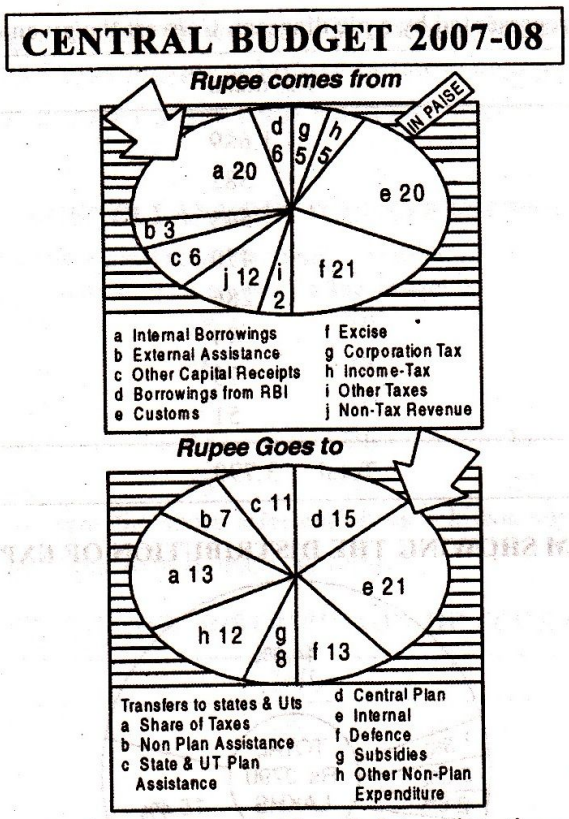


Illustration 26. Draw an ogive by less than method and determine the number of companies getting profits between Rs. 45 crores and Rs. 75 crores:

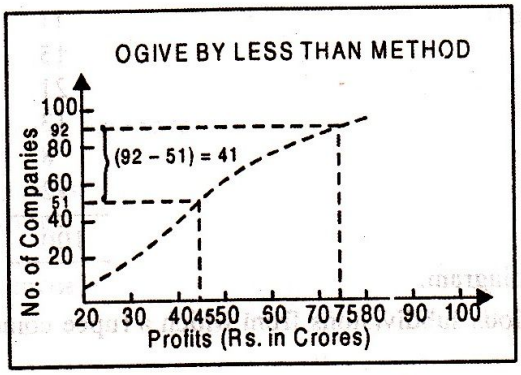
Profits (Rs. crore)	*No. of Companies*	*Profits (Rs. crore)*	*No. of Companies*
10–20	8	60–70	10
20–30	12	70–80	7
30–40	20	80–90	3
40–50	24	90–100	1
50–60	15		

(MBA, DU, 1999)

Solution.

OGIVE BY LESS THAN METHOD

Profits (Rs. crore)	*No. of Companies*
Less than 20	8
" " 30	20
" " 40	40
" " 50	64
" " 60	79
" " 70	89
" " 80	96
" " 90	99
" " 100	100



It is clear from the graph that the number of companies getting profits less than Rs. 75 crores is 92 and the number of companies getting profits less than Rs. 45 crores is 51. Hence the number of companies getting profits between Rs. 45 crores and Rs. 75 crores is 92–51=41.

Illustration 27. The following distribution is with regard to weight in grams of mangoes of a given variety. If mangoes of weight less than 443 grams be considered unsuitable for foreign market, what is the percentage of total yield suitable for it? Assume the given frequency distribution to be typical of the variety:

Weight in gms	*No. of mangoes*	*Weight in gms*	*No. of mangoes*
410–419	10	450–459	45
420–429	20	460–469	18
430–439	42	470–479	7
440–449	54		

Draw an ogive of 'more than' type of the above data and deduce how many mangoes will be more than 443 grams.

(M.B.A., Delhi Univ., 2006)

Solution. Mangoes weighing more than 443 gms. are suitable for foreign market. Number of mangoes weighing more than 443 gms. lies in the last four classes. Number of mangoes weighing between 444 grams and 449 grams would be:

$$\frac{6}{10} \times 54 = \frac{324}{10} = 32.4$$

Total number of mangoes weighing more than 443 gms. = 32.4 + 45 + 18 + 7 = 102.4

$$\text{Precentage} = \frac{102.4}{196} \times 100 = 52.25$$

Therefore, the percentage of the total mangoes suitable for foreign market is 52.25.

OGIVE BY MORE THAN METHOD

Weight more than (gms.)	*No. of mangoes*
410	196
420	186
430	166
440	124
450	70
460	25
470	7

From the graph, it can be seen that there are 103 mangoes whose weight will be more than 443 gms. and are suitable for foreign market.

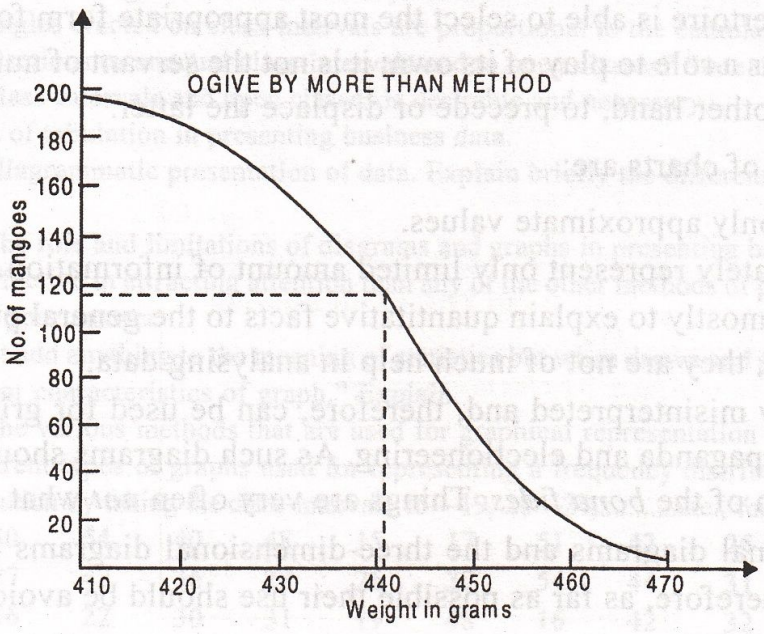
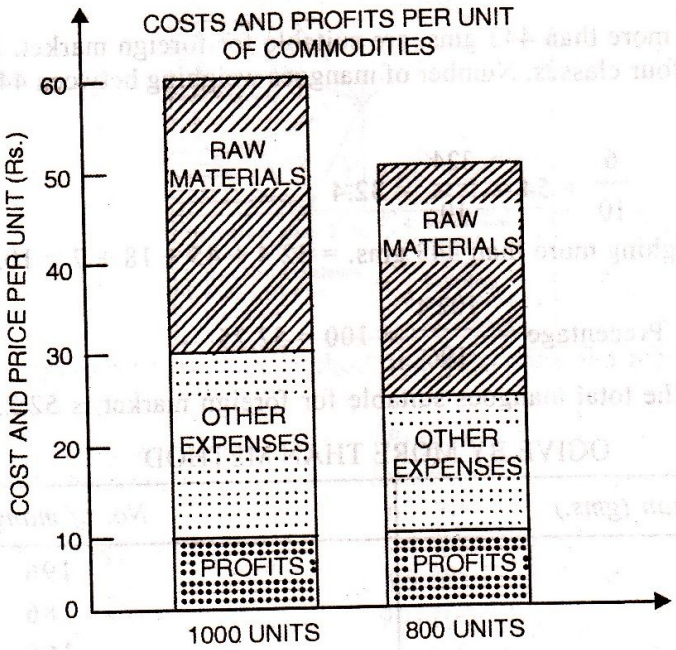


Illustration 28. Draw a rectangular diagram to represent the following information:

	Factory A (Rs.)	*Factory B* (Rs.)
Price per unit	60	50
Units produced	1,000	800
Value of Raw Materials used	3,000	2,400
Other expenses	2,000	1,400
Profit	1,000	1,000

(MBA, HPU, 2006)

Solution.



Limitations of Charts

Although charts are powerful and effective method for presenting statistical data, they are not under all circumstances and for all purposes complete substitutes for tabular and other forms of presentation. The well-trained specialist in the field is one who recognises not only the advantages but also the limitations of these techniques. He knows when to use and when not to use these methods and from his repertoire is able to select the most appropriate form for every purpose. It is said "Graphic statistics has a role to play of its own; it is not the servant of numerical statistics, but it cannot pretend, on the other hand, to precede or displace the latter."

The main limitations of charts are:

1. They can present only approximate values.
2. They can appropriately represent only limited amount of information.
3. They are intended mostly to explain quantitative facts to the general public. From the point of view of the statistician, they are not of much help in analysing data.
4. They can be easily misinterpreted and, therefore, can be used for grinding one's own axe during advertisement, propaganda and electioneering. As such diagrams should never be accepted without a close inspection of the *bona fides*. Things are very often *not* what they appear to be.
5. The two-dimensional diagrams and the three-dimensional diagrams cannot be accurately appraised visually and, therefore, as far as possible their use should be avoided.

PROBLEMS

Answer the following questions, each question carries **one** mark.

(*i*) State the use of pie diagram.

(*ii*) What is the use of histogram ? (*MBA, Madurai Kamraj, Nov. 2001*)

(*iii*) What do you mean by a pictogram ? (*MBA, Madurai Kamraj, Nov. 2003*)

(*iv*) Explain how the ogives are drawn for any frequency distribution.

(*v*) What is tabulation ?

(*vi*) What is a percentile ?

(*vii*) What is cumulative frequency curve ?

(*viii*) What is open end distribution ?

(*ix*) What is histogram ?

(*x*) What is frequency polygon ?

Answer the following questions, each question carries **four** marks. (*MBA, UP Tech. Univ, 2004*)

(*i*) Distinguish between classification and tabulation of data.

(*ii*) Give at least two uses each of classification and tabulation.

(*iii*) Distinguish between discrete and continuous variables with suitable examples.

(*iv*) Distinguish between one dimensional and two dimensional diagrams.

(*v*) What are the important steps in forming a frequency distribution ?

(*a*) Explain the role of tabulation in presenting business data, and discuss briefly the different methods of presentation.

(*b*) What are the different types of graphs and charts known to you? What are their uses?

(*a*) Explain the term 'classification' and 'tabulation'. Point out their importance in a statistical investigation. What precautions would you take in tabulating statistical data?

(*b*) "In classification and tabulation, commonsense is the chief requisite and experience the chief teacher." Comment.

(*c*) Explain pie diagram and histogram as methods of diagrammatic and graphic presentation of data with suitable examples.

(*a*) What are the requisites of a good table? State the rules that serve as a guide in tabulating statistical material.

(*b*) Explain briefly some of the uses of graphs and charts in presenting business data etc.

(*MBA, Osmania Univ., 2004*)

(*a*) What are the chief functions of tabulation? What precautions would you take in tabulating statistical data?

(*b*) What are the characteristics of a good table?

(*c*) Explain Struge's rule in forming frequency distribution.

(*a*) Distinguish clearly between a continuous variable and a discrete variable. Give two examples of continuous variables and two examples of discontinuous or discrete variables that might be used by a business statistician.

(*b*) Explain how tables, graphs and charts help in the effective presentation of data.

(*c*) State whether the following statement is *true* or *false* :

The heights of rectangles erected on class-intervals are proportional to the cumulative frequency of the classes.

(*a*) What are the objections to unequal class-intervals and to open classes? State the conditions under which the use of unequal class-intervals and open classes is desirable and necessary.

(*b*) Mention the role of tabulation in presenting business data.

Point out the role of diagrammatic presentation of data. Explain briefly the different types of bar diagrams known to you.

(*a*) Explain clearly the role and limitations of diagrams and graphs in presenting business data.

(*b*) Charts are more effective in attracting attention than any of the other methods of presenting data. Do you agree? Give reasons for your answer.

(*a*) "Diagrams do not add anything to the meaning of statistics but when drawn and studied intelligently they bring to view the salient characteristics of graph." Explain.

(*b*) Explain briefly the various methods that are used for graphical representation of frequency distributions.

(*c*) Explain the different types of graphs used for representing a frequency distribution. (*MBA, KU., 2006*)

For a frequency distribution by taking the class-intervals 15—19, 20—24,.....etc., for the following data:

30	42	30	54	40	48	15	17	51	42	25
41	30	27	42	36	28	26	37	54	44	31
36	40	36	22	30	31	19	48	16	42	32
21	22	40	33	41	21					

12. Form a frequency distribution taking a suitable class-interval for the following data giving the age of 52 employees in a government agency.

67	34	36	48	49	31	61	34	43	45	38
32	27	61	29	47	36	50	46	30	46	32
30	33	45	49	48	41	53	36	37	37	47
30	46	50	28	35	35	38	36	46	43	34
62	69	50	28	44	43	60	39			

13. Draft a blank table to show :

(*a*) Sex, (*b*) three ranks—supervisors, assistants and clerks, (*c*) years 2000 and 2010, and (*d*) Age groups—18 years and under, over 18 years but less than 55 years, over 55 years.

14. Prepare a table with a proper title, division and sub-divisions to represent the following heads of information:

(*i*) Export of Cotton piece-goods from India.
(*ii*) To Burma, China, Indonesia, Iran, Iraq.
(*iii*) Amount of piece-goods to each country.
(*iv*) Value of piece-goods to each country.
(*v*) From 2007-08 to 2009-10, year by year.
(*vi*) Total amount of exports each year.
(*vii*) Total value of exports each year.

15. Represent the following data by a suitable diagram :

Year	2005	2006	2007	2008	2009	2010
Sale of steel (in thousand tonnes)	8	8.8	9.2	10.2	7.6	12.5

16. Represent the following data by a suitable diagram:

Year	2004-2005	2005-06	2006-07	2007-08	2008-09	2009-10
Profit before taxes (Rs. lakh)	28	29.4	30.2	27	32.5	40.6

17. Represent the following data by a sub-division bar diagram :

(in lakh for Rs.)

Year	2007-08	2008-09	2009-10
Gross Income	460	482	552
Gross expenditure	400	450	500
Net Income	60	32	22

18. Represent the following information diagrammatically :

Factory	*Wages* (*Rs.*)	*Material* (*Rs.*)	*Other costs* (*Rs.*)	*Profits* (*Rs.*)	*No. of Units*
A	3,000	5,000	1,000	1,000	1,000
B	2,000	3,000	800	500	700

19. Represent the following data by a suitable diagram :

UTILIZATION OF 100 PAISE OF INCOME BY XYZ LTD. IN THE YEAR 2009-10

1.	Raw Material, Manufacturing and other Expenses	40 Paise
2.	Wages, Salaries, Bonus and other Benefits to employees	12 Paise
3.	Selling and Distribution Expenses	4 Paise
4.	Interest–Financing Charges	4 Paise
5.	Depreciation and Development Rebate	3 Paise
6.	Excise Duty of Sales	15 Paise
7.	Taxation	13 Paise
8.	Dividends	6 Paise
9.	Surplus retained in Business	3 Paise

20. Represent the following data by a "Pie Diagram" :
CHEQUES CLEARED IN INDIA IN CLEARING HOUSES IN THE YEARS 2009 AND 2010

Centres	*Amount in Crore of rupees*	
	2009	*2010*
Mumbai	829	2,670
Kolkata	1,070	2,443
Chennai	108	274
Other centres	313	615
Total	2,320	6,002

21. Draw a suitable diagram for the following :

Expenditure	*Family A (Income Rs. 10000)*	*Family B (Income Rs. 12000)*
Food	3000	4000
Clothing	2500	2000
Education	500	3600
Others	3800	3000
Saving or deficit	+200	– 600

22. Draw the histogram, frequency curve and the ogive curve for the following data pertaining to income distribution for 1500 employees working in a company.

Monthly income (in thousand Rs.)	*No. of employees*	*Monthly income (in thousand Rs.)*	*No. of Employees*
18–20	10	28–30	320
20–22	35	30–32	200
22–24	140	32–34	75
24–26	300	34–36	35
26–28	370	36–38	15

23. What is meant by a histogram? State briefly how it is constructed? Indicate clearly how the histogram in respect of the following data can be drawn (only a rough sketch is required). State also how you can draw histogram in respect of unequal class-intervals.

Mid-Value	*Frequency*	*Mid-value*	*Frequency*
115	6	165	60
125	25	175	38
135	48	185	22
145	72	195	3
155	116		

24. The following table gives the total units produced at the beginning of different years. Represent the data graphically and estimate the mid-year value for 2002 and 2010.

Years	*Units Produced*	*Years*	*Units Produced*
2002	20	2007	811
2003	62	2008	1,104
2004	147	2009	1,425
2005	300	2010	1,755
2006	536		

25. Represent the data showing the number of companies in various ranges of subscribed capital by means of a histogram:

Subscribed capital (Rs. crore)	*No. of companies*	*Subscribed capital (Rs. crore)*	*No. of companies*
Up to 10	10	50–80	7
10–20	12	80–100	8
20–30	10	Above 100	5
30–50	14		

26. The data below give the yearly profits (in thousand rupees) of two companies *A* and *B*:

	Profits (In '000 Rs.)	
Year	*Company A*	*Company B*
2005–06	120	90
2006–07	135	95
2007–08	140	108
2008–09	160	120
2009–10	175	130

Represent the data by means of a suitable diagram.

27. Below is given the frequency distribution of weekly wages of 100 workers in a factory:

Monthly wages (Rs.)	*No. of workers*	*Monthly wages (Rs.)*	*No. of workers*
3000–3500	3	5500–6000	10
3500–4000	5	6000–6500	8
4000–4500	12	6500–7000	5
4500–5000	23	7000–7500	3
5000–5500	31		

Draw the ogive for the distribution and use it to determine the median wage of a worker.

28. Present the following information in a suitable tabular form supplying the figures not directly given:

"In 2009, out of a total 2,000 workers in a factory 1,550 were members of a trade union. The number of women workers employed was 250, out of which 200 did not belong to any trade union."

"In 2010, the number of union workers was 1,725 of which 1,600 were men. The number of non-union workers was 380 among which 155 were women." *(MBA, GGSIP Univ.)*

29. Draw an Ogive for the following distribution. Read the median from the graph and verify the result by formula. How many workers earned monthly wages between Rs. 5,400 and Rs. 5470 ?

Monthly wages (in Rs.)	*No. of workers*	*Monthly wages (in Rs.)*	*No. of workers*
5000–5200	6	5800–6000	16
5200–5400	10	6000–6200	12
5400–5600	22	6200–6400	15
5600–5800	30		

30. The proprietor of Goodwill Tyres Co. kept a record of the number of car tyres of each brand that were sold during 2009–10. He arranged the data as follows :

Brand	*No. of Tyres Sold*
Dunlop	280
Modi	270
Firestone	200
Ceat	190
Goodyear	160
J.K.	100

(*a*) What kind of a distribution is this?

(*b*) What are the class boundaries of each class?

(*c*) Present the data by a suitable diagram/graph.

31. Draw a suitable diagram to represent the following information :

	Company A	*Company B*
Selling price	12,000	8,000
Raw Materials	5,000	6,000
Direct Wages	4,000	3,200
Factory and office on cost	1,000	800

32. A company dealing in 60 products, in the course of establishing an inventory control system, classified products according to price as shown in the frequency table below :

Unit cost (in hundreds of Rs.) :	3–5	6–8	9–11	12–14	15–17	18–20	21–23
No. of items :	6	8	10	20	8	5	3

Prepare more than ogive for the distribution on a graph paper. Use this graph to determine 20th and 80th percentiles.

33. Discuss the following terms with illustrations :
   (*i*) Classification and Tabulation.
   (*ii*) Frequency, cumulative frequency and frequency polygon.
   (*iii*) Histogram, line and bar diagrams.
   (*iv*) Pie chart. (*MBA, Vikram Univ., 1998*)

34. (*a*) Diagrams help us to visualise the whole meaning of a numerical complex at a single glance". Comment.

   (*b*) Draw a suitable diagram to represent the following:

	Selling Price Per Unit (Rs.)	*Qty Sold*	*Wages*	*Materials*	*Others*
Factory A	400	20	3200	2400	1600
Factory B	600	30	6000	6000	9000

Show also profit or loss as the case may be. (*MBA, HPU, 2001*)

35. From the following frequency distribution, prepare the less than and more than cumulative frequency curve (ogive curve)

Class-Interval :	0–10	10–20	20–30	30–40	40–50	50–60
Frequency :	8	12	30	25	18	17

(*MBA, K.U., 2003*)

*****