UNIT 7 DATA PRESENTATION

Structure

7.0 OBJECTIVES

After going through this unit you will be able to:

- arrange discrete data in a tabular form;
- identify class intervals and arrange continuous data in tabular form;
- present data in the form of histogram and frequency curves;
- arrange data in the form of bar diagrams; and
- represent data in the form of pie chart.

7.1 INTRODUCTION

Let us imagine a situation where you have to collect data on the number of visitors to a library and their age, income level, gender, education level and areas of interest. Your objective is to revamp the library policy. For example, you may like to decide on a membership fee keeping in view the ability to pay of users, or to build stock of books depending on the age profile and interest of users, or to increase the sitting capacity on the basis of the number of visitors, or to provide for separate facilities on the basis of gender. In order to carry on such a project you requested each member visiting the library to enter the above particulars on a register and compiled the same on a daily basis. No doubt you end up with a large volume of data very soon. Unless these data are subjected to statistical analysis it would not be of much help in policy formulation.

Once data collection is completed, your efforts should be geared towards bringing these raw data into a presentable form. You may think of presentation of the data in the form of a table or a chart. The task before you is to decide on the structure of the table or the shape of the graph. Let us look in to the options available.

In the previous Unit we had classified variables into nominal, ordinal and numerical types. Let us begin with tabulation of data pertaining to different types of variables and later on we will move on to graphical presentation. The first step in the analysis and interpretation of data is its classification and tabulation. The process of arranging data into groups according to their common characteristics is known as its classification. On the other hand tabulation implies a systematic presentation of data in rows and columns according to some salient features or characteristics.

In this Unit we will first discuss tabulation of data pertaining to discrete and continuous variables. Subsequently we will take up graphical presentation. There are basically three forms in which data can be presented: i) graphs such as line graph, histogram, frequency polygon and frequency curve, ii) bar diagrams (single bars, component bars, multiple bars), and iii) pie charts. We explain each of these through appropriate examples.

7.2 PREPARATION OF A TABLE

There are certain things we should take care of while preparing a table.

- It is required to give a table number for identification of the particular table.
- There should be a title of the table that indicates the type of information contained in the table. Title should be brief and precise. Avoid expressions like 'Table presents…' or Table contains….'
- If necessary give a head note. It should be given in parentheses and should appear on the right side top just below the title. See, for example, the expression (in Rupees) given in Table 7.4.
- Stub head describes the nature of stub entry, e.g., 'class interval' in Table 7.4
- Stub entries describe the rows.
- Caption describes the nature of data presented in columns followed by column heads and sub-heads. In certain tables it may not be necessary to give sub-heads.
- The main body of the table contains numerical information.
- Below the table there may be footnote. The purpose of footnote is to caution the readers about the limitations of the table.
- Source of the table may be the last component. It is quite important in the case of secondary data. It provides opportunity to the readers to check the data if they desire and get more of it.
- Remember that you have to design your own table, keeping your requirements in view. In Table 7.1 we have summarised different parts of a table.

Table No. 7.1: (————————TITLE————————)

(Head note)

Stub Head	←———————— Caption ————————→			
	Column Head I		Column Head II	
	Sub-head	Sub-head	Sub-head	Sub-head
Stub Entries	MAIN BODY OF THE TABLE			
Totals				

Footnote:
Source:

7.3 TABULAR PRESENTATION

7.3.1 Nominal and Ordinal Data

You may recall that nominal and ordinal data can be classified into categories. Thus we can count the number of observations in a category, note them down and present in a table. Such an arrangement of data is called 'frequency distribution' because we are counting the frequency with which each category is repeated.

Let us take a concrete example. Suppose, in order to collect data on areas of interest of visitors to the library you identified 5 subject areas, viz., economics, history, political science, sociology and public administration. Thus there are five categories. Very often the number of observations may be large enough. In such situations we use tally bars. In this method all the categories (in our example, five) are written in a column. For every observation, a tally bar denoted by (|) is noted against its corresponding category. Every fifth repetition is marked by crossing the previous four bars as (𝍸). In this way, we get blocks of five, which simplify counting at the end. Thus a category repeated fourteen times will be marked as (𝍸 𝍸 IIII).

Table 7.2: Frequency Distribution of Areas of Interest of Visitors

Area of Interest	Tally Sheet	Frequency
Economics	𝍸 𝍸 𝍸	15
History	𝍸 𝍸 II	12
Political Science	𝍸 𝍸 𝍸 𝍸 II	22
Public Administration	𝍸 𝍸 𝍸 III	18
Sociology	𝍸 𝍸 𝍸 𝍸 III	23
Total		**90**

In Table 7.2 we have constructed a frequency distribution of 90 visitors to a library according to their area of interest. Here we have given the tally sheet also. But while making the final presentation we delete the second column, that is the tally sheet, and provide only the first and third columns so that the table does not look cluttered.

At times we are interested in 'relative frequency'. In this case the percentage share of

each category is given in addition to actual frequency. For example, for the data given in table 7.2 we obtain relative frequency by dividing each frequency by the total and then multiplying by 100. For economics we obtain relative frequency as $\frac{15}{90} \ 100 =$ 16.7. The total of all frequencies is 100. Relative frequency gives the percentage share of each category in the total.

Table 7.3: Relative Frequency Distribution

Area of Interest	Frequency	Relative Frequency (%)
Economics	15	16.7
History	12	13.3
Political Science	22	24.4
Public Administration	18	20.0
Sociology	23	25.6
Total	**90**	**100.0**

7.3.2 Numerical Data

In the case of numerical data we have two types: discrete and continuous. The frequency distribution of discrete data is not much different from the method discussed above for nominal and ordinal data. Here we make a list of all possible values that the characteristic is likely to assume and count the frequency of occurrence for each value. In place of categories in Table 7.2 we write down the discrete values and construct a similar table. For example, if you have to prepare a frequency distribution of 'number of books issued' to 100 members of the library you may obtain a frequency distribution as given in Table 7.3.

Table 7.4: Books Issued to Borrowers

Number of books issued	Number of borrowers
0	10
1	23
2	25
3	17
4	15
5	10
Total	**100**

In the case of continuous data preparation of a frequency distribution is somewhat different because of the following: i) Recall that an observation can assume any value (implies infinite number of values) within a range in the case of continuous data. Hence it may not be possible to list out all the possible values. ii) Repetition of the same value for two or more observations may be a rare coincidence.

In order to resolve the above issue we divide the range of values that is the difference between the highest and the lowest values, into certain 'class intervals'. In each class

interval we count the number of observations and report it. For example, let us consider the monthly expenditure on purchase of books by 175 persons. We find that the lowest monthly expenditure is Rs. 135 and the highest is Rs. 750. Thus the range is Rs. 750 - Rs. 135 = Rs. 615. We divide this range into 7 class intervals and prepare a frequency distribution as given in Table 7.5.

Table 7.5: Monthly Expenditure on Purchase of Books (in Rupees)

Class Interval	Frequency	Relative Frequency
100-200	21	12.00
200-300	32	18.29
300-400	49	28.00
400-500	33	18.86
500-600	23	13.14
600-700	12	6.86
700-800	5	2.86
Total	**175**	**100.00**

At this point two questions may be shaping up in your mind:

- How many class intervals should be taken?
- What should be the width of each class interval?

Let us discuss some of the issues that we should be careful about.

Number of class intervals: There is no hard and fast rule regarding the number of class intervals. However, it should not be too small, neither too large. If the number of class intervals is too small, then there is a chance of losing valuable information due to grouping. For example, in Table 7.5, we do not know the exact amount spent by the 21 persons whose monthly expenditure is between Rs. 100 and Rs. 200. Note that when we have lesser number of class intervals, the width of class intervals will increase. On the other hand, if the number classes is very large, the distribution may appear to be too fragmented and may not reveal any pattern of behaviour. Based on experience, it has been observed that the minimum number of classes should not be less than 5 and in any case, there should not be more than 20 classes. A decision on the number of class intervals should also take into account the number of observations - higher the number of observations, higher the number of class intervals.

Width of Class Intervals: As far as possible, all the classes should be of equal width. However, when a frequency distribution, based on equal class intervals, does not reveal a regular pattern of behaviour, it might become necessary to re-group the observations into class intervals of unequal width. By a regular pattern of behaviour we mean that observations should not be distributed among classes in an erratic manner. In other words, there should not be situations where frequency is zero in one class and very high in the adjoining class.

Open-ended Class Intervals: In many cases a few observations may be very high or very low in value. For example, in Table 7.5, suppose one person has a monthly expenditure of Rs. 1150 while others have less than Rs. 800. Here if we provide four extra class intervals, namely, 800-900, 900-1000, 1000-1100, and 1100-1200, then the frequency in the classes 800-900, 900-1000, 1000-1100 classes will be zero each

and in the class 1100-1200 will be one. In order to manage such cases we often resort to open-ended class intervals. In Table 7.5, instead of having class limits for the last class as 700-800 we may modify it as 'more than 700'. This class may include any observation above Rs. 700. There may be another situation where we need to modify the first class interval as 'less than 200' if one or two observations are lees than Rs. 100.

Mid-value of a Class: If we look into Table 7.5 we find that for each class, there are two class limits - lower limit and upper limit. We assume that the observations are uniformly distributed within the class. Thus we can say that the average value of the observations in a class is equal to the mid-value of the class. In Table 7.5 the mid-value of the first class (100-200) is 150 while that of the second class (200-300) is 250 and so on. Remember that class limits are usually kept as multiple of 5 or 10 so that it is convenient to locate the mid-value of a class.

You may observe that in Table 7.5 the upper limit of the second class - interval (200-300) is equal to the lower limit of the third class interval (300-400). In which class interval do you include a person having monthly expenditure of exactly Rs. 300? We should note that the second class - interval is defined as 'monthly expenditure of Rs. 200 or more but less than Rs. 300'. Similarly, the third class interval is defined as 'monthly expenditure of Rs. 300 or more but less than Rs. 400'. Naturally Rs. 300 would be included in the third class interval (300-400) and not in the second class - interval (200-300).

As in the case of nominal data, we can present the relative frequency for each class interval. It is obtained by presenting the frequencies as percentage of the total so that total frequency is 100 percent (or equal to 1).

Self-Check Exercise

1) Define the following terms.

   a) Class interval

   b) Open-ended class

   c) Frequency distribution

2) What are the factors one should keep in mind while preparing a frequency distribution for continuous data?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

.....

.....

Cumulative Frequency

Many times we are interested to know the number of persons below certain value. For example, for the data given in Table 7.5, we are interested to know the number of persons whose monthly expenditure on books is less than Rs. 500. We obtain it by adding the frequencies of preceding classes and find it to be 21 32 49 33 135.

Table 7.6: Cumulative Frequency (monthly expenditure in Rupees)

Class Interval	Frequency	Cumulative Frequency (less than type)
Less than 200	21	21
Less than 300	32	53
Less than 400	49	102
Less than 500	33	135
Less than 600	23	158
Less than 700	12	170
Less than 800	5	175
Total	**175**	

We can construct cumulative frequencies for the number of persons having a monthly expenditure of more than a particular value. For example, suppose we have to find out the number of persons having monthly expenditure of more than Rs. 400. We can obtain it by adding the frequencies of the succeeding classes and find it to be . Similarly, you can find out cumulative frequency for other classes.

33 23 12 5 73

7.4 GRAPHICAL PRESENTATION

Collected data are very often presented through graphs and diagrams for greater clarity.

7.4.1 Line Graph

Suppose you are provided with data on number of books issued in a library (month-wise for the year 2004) as given in Table 7.1.

Table 7.4.1: Number of Books Issued in a Library

Month	Number of visitors	Month	Number of visitors
January	76	July	105
February	85	August	108
March	86	September	110
April	90	October	115
May	82	November	118
June	98	December	106

Line graph is appropriate when we need to present the movement or variation in a variable. It is quite simple to draw and indicates the increase or decrease in a variable over time or across observations. Line graphs can be used for discrete data. Recall that in the case of continuous data we assumed that the average value of each class is its mid-point. Thus we can plot the frequencies for each class against its mid-point and join these points to obtain a line graph.

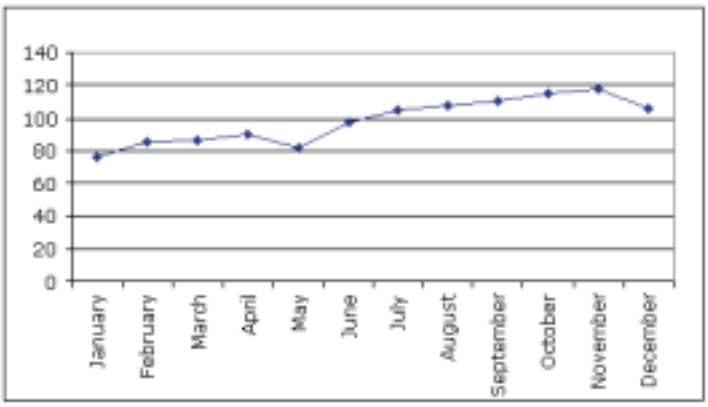


Fig. 7.1: Number of Books Issued in a Library

In Fig. 7.1 we take the variable 'months' on x-axis and the number of books issued on the y-axis and plot it as line graph. You may observe that the number of books issued has increased over time except for the months of May and December.

7.4.2 Histogram

Histogram is a rectangular diagram where the area of each rectangle is proportional to the frequency of the respective class. Remember that histogram is appropriate for continuous data arranged into class intervals. It is not used for discrete data.

The steps followed are:

- On a graph paper we mark class intervals such as 100-200, 200-300, etc. on the horizontal axis.
- Similarly we mark frequencies on the vertical axis.
- We draw rectangles as shown in Fig. 7.2.
- When class intervals are equal the height of rectangles are equal to the frequency of classes.
- When class intervals are not equal the frequencies are adjusted so that area of rectangle is proportional to class frequency. For example, if the interval of one class is double that of other classes, then we need to divide the frequency of the former by two.

Let us construct histogram for the data given in Table 7.5.

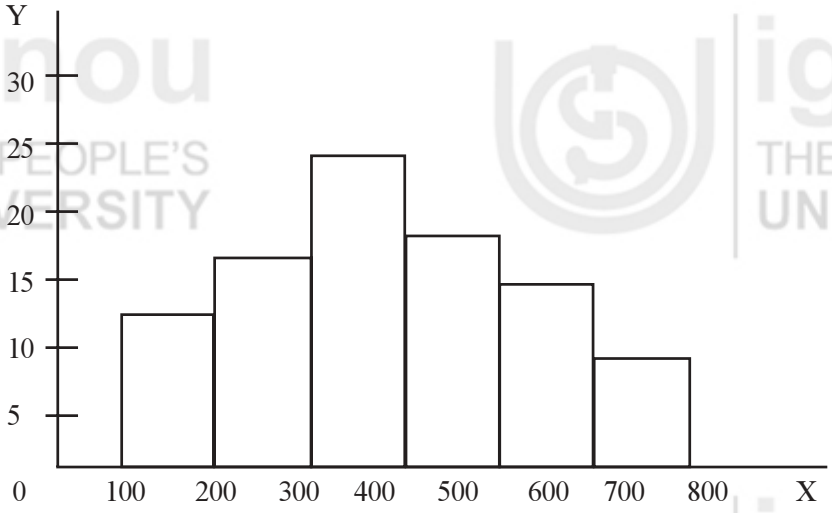


Fig. 7.2: Histogram

7.4.3 Frequency Polygon

It is obtained from a histogram by joining the mid-value of the top of the rectangles with the help of straight lines as shown in Fig. 7.3. Remember that the area under the frequency polygon should be same as the area under the histogram. Hence, we draw two additional class intervals, one on each end of the histogram. For the histogram given in Fig. 7.2 we follow the steps given below.

- Draw two class intervals, viz., 0-100, and 800-900.
- Take the frequency for these two classes to be 0.
- Join the mid-values of all the classes, including 0-100, and 800-900.
- The frequency polygon obtained is as given in Fig. 7.3.

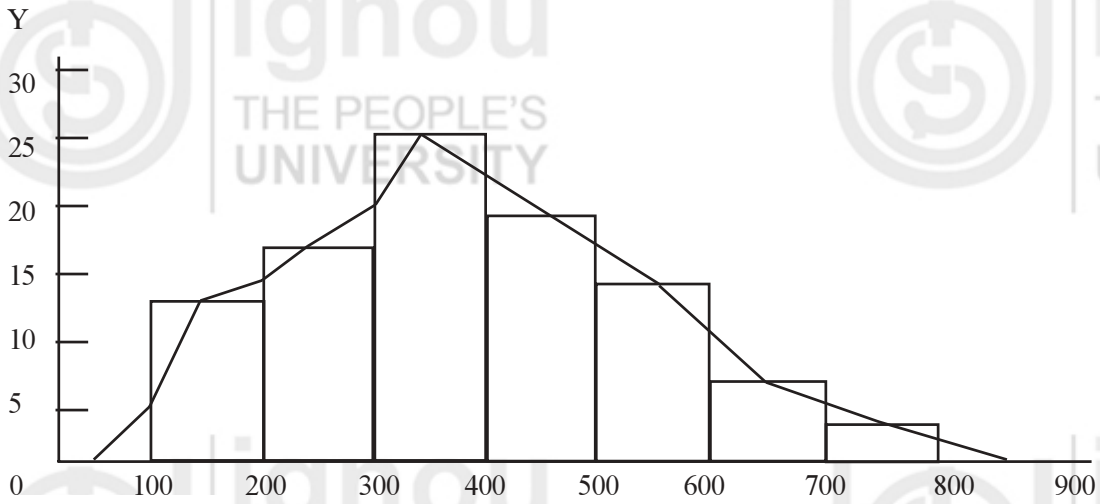


Fig. 7.3: Frequency Polygon

y Curve

frequency polygon by smoothening the edges of the frequency
n Fig. 7.4.

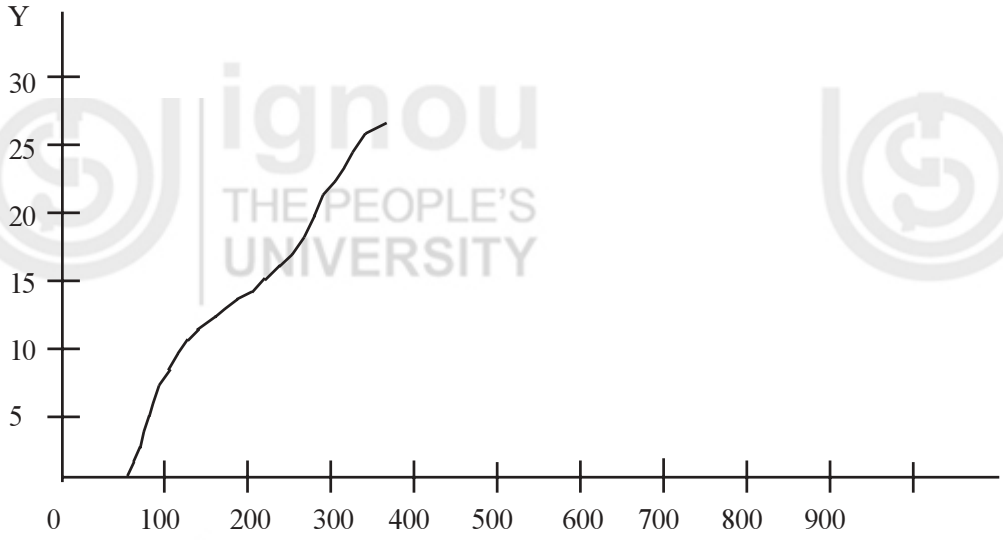


Fig. 7.4: Frequency Curve

Self Check Exercise

3) Construct a more than type cumulative frequency distribution for the data given in Table 7.5.

Note: i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

7.4.5 Bar Diagrams

Diagram is a visual form of presentation data. Diagrams can be bars, squares, circles, maps and pictures. In this section we discuss the procedure of drawing bar diagrams. Bar diagrams are more appropriate for nominal data where certain categories are distinguished.

A bar is defined as a thick line, often made thicker to attract the attention of a reader. The height of the bar highlights the value of the variable. Remember that width of the bars does not mean anything. Moreover, bars are separated from each other with equal gaps. Thus it is different from histogram, which is more appropriate for quantitative data and area of the bars is important. Finally, in histogram the bars are always vertically placed whereas in bar diagram they can be placed both vertically as well as horizontally. Let us take a simple example to demonstrate the construction of a bar diagram.

Table 7.4.2: Number of Students in a College

Stream	Number of students
Arts	800
Science	460
Commerce	520

The bar diagram of the above data is drawn in Fig. 7.5. To make the bar diagrams beautiful we can either fill in colour in the bars or shade them in different ways. This is left to the aesthetic taste of the investigator.

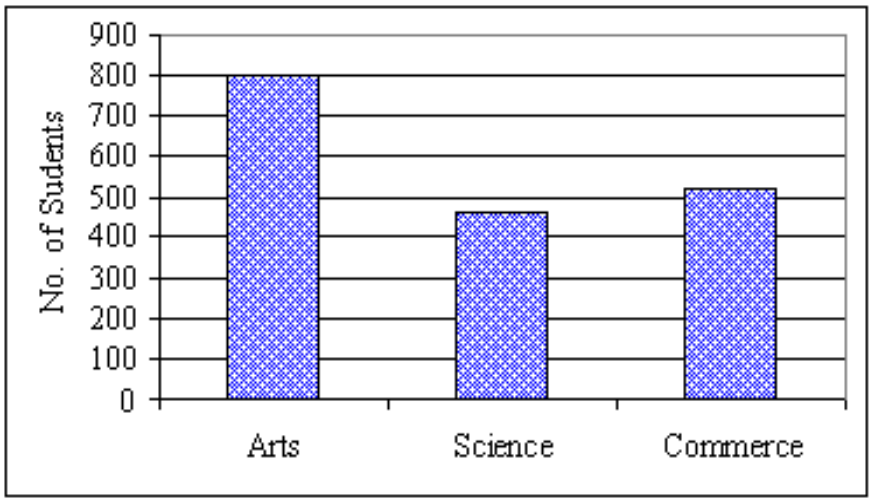


Fig. 7.5: Simple Bars

Component Bar Diagram

A component bar diagram presents the components of a phenomenon so that a comparison can be made. It is more appropriate for qualitative data. The bars corresponding to each category is divided into various components. The portion of the bar occupied by each component denotes its share in the total.

Suppose we have additional information on the number of boys and girls admitted in a college (Table 7.9). We present component bar diagram for such data in Fig. 7.6. Remember that the sub-divisions of different bars must always be done in the same order and these should be distinguished from each other by using different colours or shades.

Table 7.4.3: Number of Students in a College

	Boys	Girls
Arts	400	400
Science	260	200
Commerce	350	170

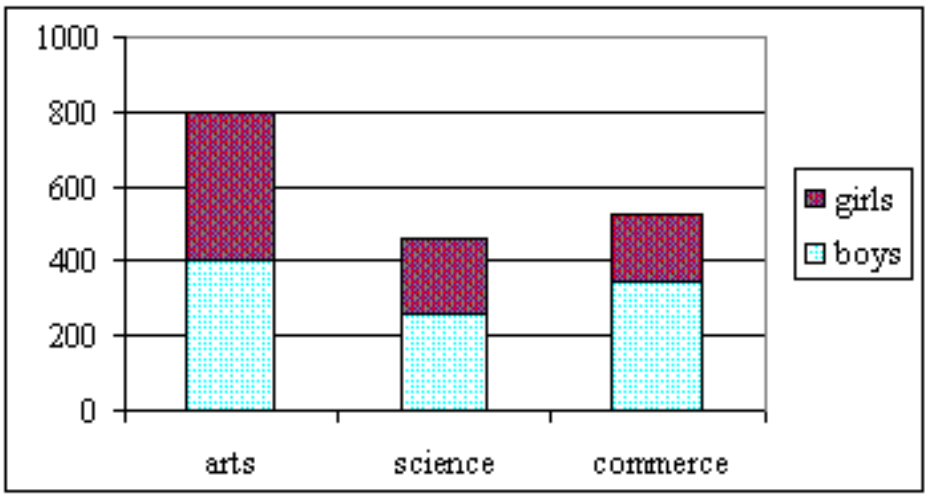


Fig. 7.6: Component Bar Diagram

Multiple Bar Diagram

It is used when comparisons are to be shown between two or more sets of data. A set of bars for a period or place or a related phenomenon is drawn side by side without gap. Different bars are distinguished from one another by different shades or colours. A multiple bar diagram for the hypothetical data given in Table 7.10 is drawn in Fig. 7.7. Suppose our purpose is to show the changes in the number of girl students admitted to different of a college over the years.

Table 7.4.4: Number of Girls Students during past three years

Stream/Year	2001	2002	2003
Arts	130	135	125
Science	45	57	65
Commerce	20	26	30

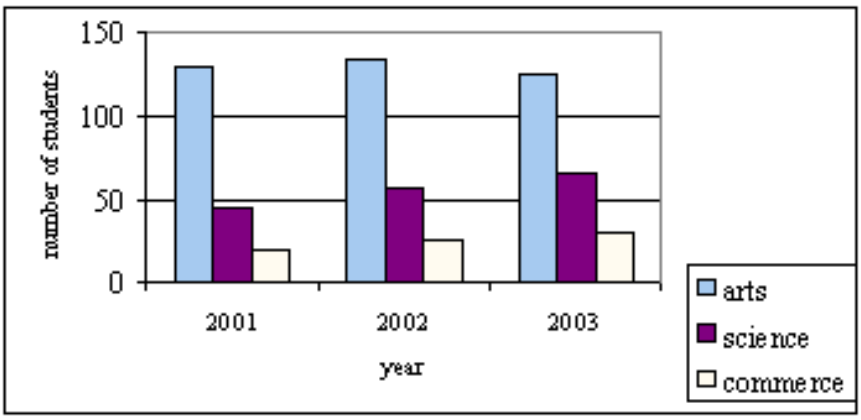


Fig. 7.7: Multiple Bar Diagram

7.4.6 Pie Chart

Pie chart is widely used to show share of different components in a variable. For example, expenditure of a library on different heads can be shown in the form a pie chart. Suppose for the financial year 2004-05 you have budget data of a library as given in Table 7.11. Recall that a circle has 360^0. This area is divided into different components according respective share. Therefore, we first calculate the share of each component and convert it to degrees.

Table 7.4.5: Heads of Expenditure in a Library

(in Rs. Thousand)

Heads of Expenditure	Budget	Ratio of the component	Degrees
Salary	246	0.24	86.3^0
Purchase of books	325	0.32	114.0^0
Purchase of journals	175	0.17	61.4^0
Purchase of furniture	200	0.19	70.2^0
Maintenance	50	0.05	17.5^0
Contingency	30	0.03	10.5^0
Total	**1026**	**1.00**	**360^0**

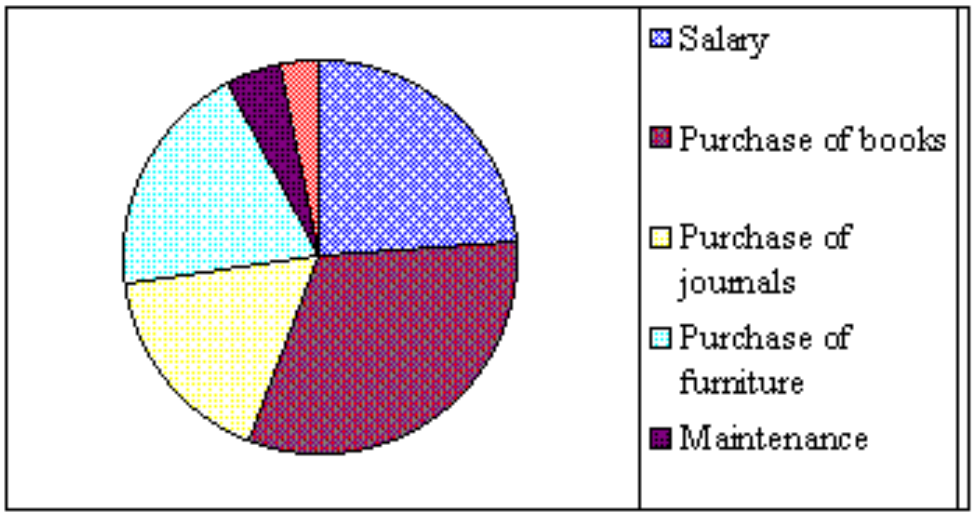


Fig. 7.8: Pie Chart

Steps to be followed in the construction of Pie diagram:

- Find the total of all components (in this case =1026).
- Find ratio of each component to total. For example, for 'salary' the ratio is $\frac{246}{1026}$ 0.24.
- Multiply the ratio of each component by 360^0. For example, for 'salary' it is $0.24 \quad 360^0 \quad 86.3^0$.
- Draw a circle of a suitable size.
- Use protractor to draw the angles you have obtained. Preferably start with the largest one.
- Shade the different segments with different colours.
- Identify the colour or shade for different components on the right hand side.
- At times we write down the share of each component inside the circle.

Self Check Exercise

4) Draw a simple bar diagram for the data given in Table 7.3.

5) For the data given in Table 7.3 draw a pie chart.

Note: i) Write your answers in the space given below.
ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

7.5 SUMMARY

In this unit we discussed the method of presentation of data. We began with preparation of frequency distribution for qualitative and quantitative data. Also we prepared relative frequencies and cumulative frequencies for numerical data.

Data can be presented in the form of tables and graphs. There are certain parts of a table that we should mention while preparing a table. Graphs can be in the form of histogram, frequency polygon and frequency curve. Diagrams could be in the form simple bar, component bar, multiple bar or pie chart.

7.6 ANSWERS TO SELF CHECK EXERCISES

1) a) The difference between the upper limit and the lower limit of a class is called the class interval.

b) A class with no upper limit is called an open- ended class.

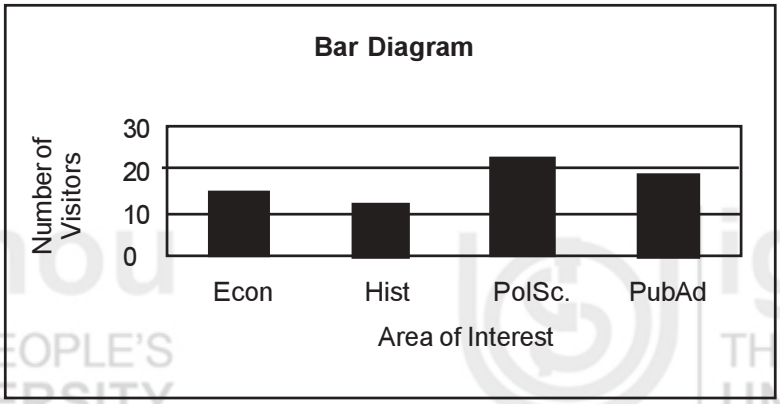
c) A tabular presentation of nominal or ordinal data as classes along with the frequency of their occurrence is called frequency distribution.

2) While preparing frequency distribution we should be keep in mind: number of class intervals, their width and whether they have to be closed or open.

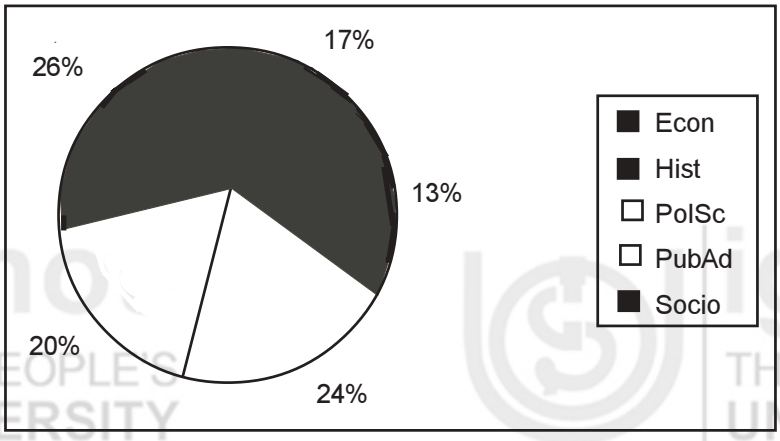
3) The 'more than' type cumulative frequency is given below.

Class Interval	Frequency	Cumulative Frequency
More than 100	21	175
More than 200	32	154
More than 300	49	122
More than 400	33	73
More than 500	23	40
More than 600	12	17
More than 700	5	5
Total	175	

4)



5)



7.7 KEYWORDS

Bar Diagram : It is often defined as a set of thick lines corresponding to various values of the variable. It is different from histogram where width of the rectangle is important.

Class and Class Limits : It is a decided group of magnitudes having two ends called class limits or class boundaries.

Class Range : Also called class interval is the difference of two limits of a class. It is equal to upper limit minus lower limit. It is also called class width.

Continuous Frequency Distribution : A continuous frequency distribution in formed where the variable can take any value between two numbers like height and weight, income, temperatures, etc.

Exclusive Type Class Interval : A class interval which includes all observations that are greater than or equal to the lower limit but less than the upper limit.

Frequency Polygon : It is a broken line graph to represent a frequency distribution and can be obtained either from a histogram or directly from the distribution.

Frequency Curve : It is a smoothened graph of a frequency distribution obtained from frequency polygon through free hand tracing in such a way that the area under both of them is approximately the same.

Inclusive Type Class Interval : A class interval in which all observations lying between and including the class limits are included.

Discrete Frequency Distribution : A discrete distribution or discrete series is formed where the variable can take only discrete values like 1,2,3,..... Number of children in a family, number of students in a university, etc. are examples of discrete variable.

Open-end Class : A class in which one of the limits is not specified.

Mid-value : It is the average value of two class limits. It falls just in the middle of a class.

Relative Frequency Distribution : It is frequency distribution where the frequency of each value is expressed as a fraction or a percentage of the total number of observations.

Histogram : It is a set of adjacent rectangles presented vertically with areas proportional to the frequencies.

Simple and Sub-divided Bar : In the case of simple bar diagram only one diagram variable can be presented. A sub-divided bar diagram is used to show various components of a phenomenon.

Pie Chart : It is a circle sub-divided into components to present proportion of different constituent parts of a total. It is also called pie chart.

7.8 REFERENCES AND FURTHER READING

Chase, Clinton I. (1988). *Elementary Statistical Procedures*. 3rd ed. New York: McGraw Hill.

Elhance, D.N. (1988). *Fundamentals of Statistics*. Allahabad: Kitab Mahal.

Gupta, C.B. (1985). *An Introduction to Statistical Methods*. Delhi: Vikas.

Hraglin, David C. [et.al.] (1985). *Exploring Data Tables, Trends and Shapes*. New York: Wiley.

Simpson, I.S. (1988). *Basic Statistics for Librarians*. London: Library Association.