3. Processing of Data

Classification and Tabulation

The collected data after their scrutiny need to be classified in order to make the data fit for analysis and interpretation. The first step in the analysis and interpretation of data is classification and tabulation. Classification is the first step in tabulation, eventhough the phrase 'classification and tabulation' is used. Proper classification helps proper tabulation.

Classification

It is the process of arranging the data on the basis of some common characteristics possessed by them. Eg. If sex is the basis of classification, then all the male population will be grouped together on one side and the female population on the other side. Likewise if age is the basis of classification, persons of the same age will be grouped together and so on.

Objects of Classification

The chief objects of classification are:

1. To condense the mass of data.
2. To present the facts in a simple form.
3. To bring out clearly the points of similarity and dissimilarity.
4. To facilitate comparison.
5. To bring out the relationship.
6. To prepare data for tabulation.
7. To facilitate the statistical treatment of the data.

Types of Classification

There are numerous ways of classifying the data. T important types are

1. *Geographical*
2. *Chronological*
3. *Qualitative*
4. *Quantitative*

1. Geographical Classification (Regionwise Classificatio

This type of classification is based on geographic region like countries, states, districts, taluks, etc. For e the yield of agricultural output per hectare for differe countries in a particular year is given below:

Table 3.1: The yield of agricultural output.

Country	Average output (in kg per hectare)
U.S.A	600
China	300
Pakistan	250
India	150

2. Chronological Classification

This type of classification is based on time of it occurrence such as years, months, weeks, days, hours, etc.

Eg. The fish (catla) production in a particular farm over 5 years is given below:

Table 3.2: Fish production.

Year	Fish production (in kg per hectare)
1987	1400
1988	1500
1989	1450
1990	1550
1991	1600

3. Qualitative Classification (Descriptive Classification)

This type of classification is based on the quality or attributes such as sex, literacy, marital status, etc. So it is also called ***descriptive classification.***

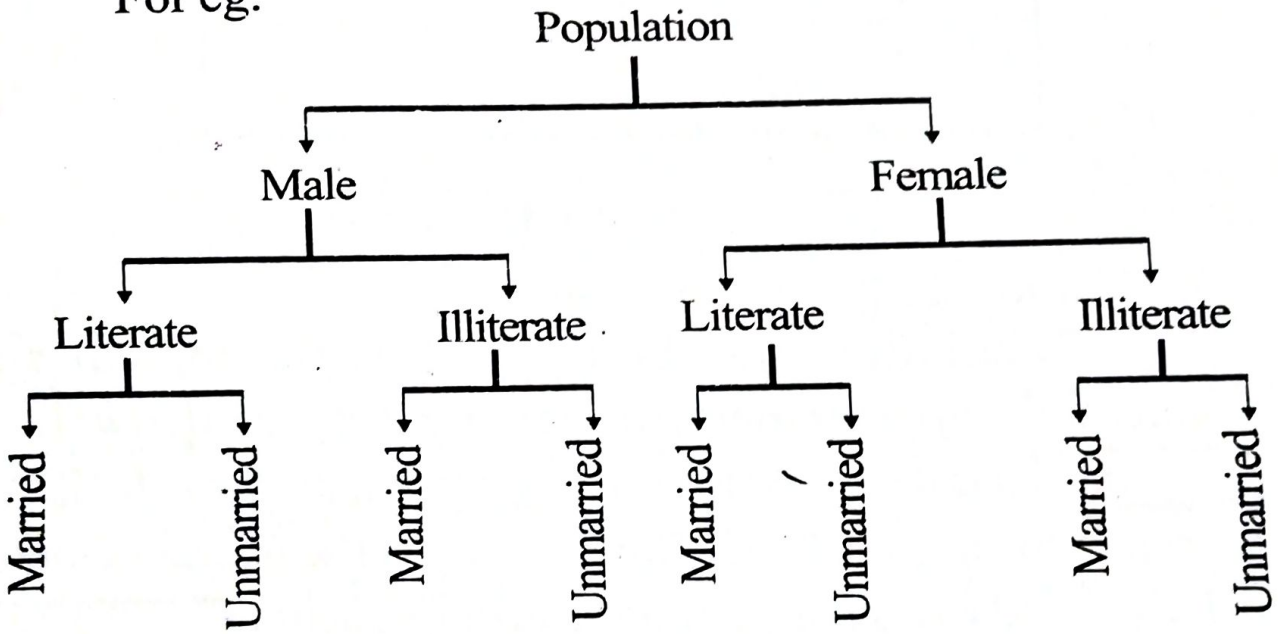
It is further divided into two types

1. *Simple classification*
2. *Manifold classification*

i. Simple Classification : The data are classified into only two classes. It is dichotomy or two fold. Eg.

ii. Manifold Classification : Here the data are classified into many classes.

For eg.



4. Quantitative Classification

This type of classification is based on some quantitative phenomenon such as age, height, weight, etc. Here variable is the quantitative phenomenon under study. Hence this classification is also called ***classification by variables.***

For example, the weight of 100 fishes reared in a pond are given below:

Table 3.3: Weight of 100 fishes.

Wt (in gms)	No. of fishes
0-100	8
100-200	16
200-300	20
300-400	12
400-500	15
500-600	13
600-700	6
700-800	6
800-900	2
900-1000	2
Total	100

Tabulation of Data

Definition

Tabulation may be defined as the logical and systematic arrangement of statistical data in rows and columns. It is designed to simplify presentation and facilitate comparison and analysis. Columns are vertical arrangements and rows are horizontal arrangements

Objectives

1. To clarify the object of investigator.
2. To simplify the complex data.
3. To present the facts in the minimum space.
4. To facilitate comparison.

5. To detect errors and omissions in the data.
6. To facilitate statistical processing.
7. To help reference.

Parts of a Table

Arranging values in columns is called ***tabulation.*** *A column of values is called a* ***table.*** *Tabulation is a* ***presentation*** *of data.*

A table contains boxes called ***cells.*** The cells are arranged in horizontal ***rows*** and ***vertical columns.***

A typical table has the following parts:

1. *Table number*
2. *Title*
3. *Head note*
4. *Caption*
5. *Stub*
6. *Body*
7. *Foot note*
8. *Source*

Table 3.4: Food habits of III B.Sc .Zoology students.

Eggtarians : Egg-eaters

Table number — Title — Caption — Head note

Food habits	Number of students	
	Boys	Girls
Vegetarians	2	7
Eggtarians (Egg-eaters)	1	2
Non-vegetarians	7	11

Stub → Vegetarians

Body → 2

Foot Note → Census was made during 2000-2001

Source → Data collected by the class teacher

Fig.3.4: Parts of a table.

The ***table*** has a number and it is given at the top.
The name of the table is called ***title***. It is given at the top.

Head note refers to the units of values given below the title.

The headings of vertical columns are called ***caption***.
The headings of horizontal columns are called ***stub***.
The values given in the horizontal and vertical columns are called ***body***.

Foot note *is given below the table. It gives explanations on the values.*

Source *refers to source of information. It is given at the bottom of the table.*

1. Table Number

A table should always be numbered for easy identification and reference in future. The table number may be placed at the top of the table either in the centre above the title or in the left side of the title.

2. Title

Every table must be given a title, which usually appears at the top of the table. It should be clear, brief and self-explanatory.

3. Head-note

It is actually a part of the title. It explains certain points relating to the whole table that have not been included in the title captions or stubs.

4. Caption

It refers to the headings of vertical columns. It has usually a main heading and a sub-heading. It should be clear, brief and self-explanatory.

5. Stub

It refers to the headings of horizontal rows.

6. Body

It contains numerical informations arranged in accordance with caption and stub. The arrangement is

generally from left to right in the horizontal rows and from top to bottom in the vertical columns.

7. Foot-note

Anything in a table which the reader may find difficult to understand can be explained in foot-notes.

8. Source

It refers to the source from which information has been taken. It should preferably include the name of the author, title, volume, number, page, publisher's name and the year of publication.

After collecting the data, the investigator has to undertake the task of its organization. By organization, the classification and presentation of data in such a way that the data becomes easy and convenient to use and handle.

Tables are broadly classified into two types.

1. *Simple tables*
2. *Complex tables*

1. Simple Tables

In a simple table, only one characteristic is shown. Hence this type of table is also known as ***one-way table.*** It has two factors placed in relation to each other. The following table shows the marks secured by students in a class test.

Table 3.5: Simple Table.

Marks	No. of students
0 - 5	2
5 - 10	5
10 - 15	10
15 - 20	11
20 - 25	9
25 - 30	20

2. Complex Tables

In a complex table, more than two characteristics are shown. If there are two co-ordinate factors, the table is called a ***double table;*** if the number of co-ordinate group is three, it is called as ***treble table.*** If it contains more than three co-ordinate factors, then it is called as ***multiple table.***

The marks of students can be classified according to the sex to get a ***double table.***

Table 3.6: Double Table.

Marks	No . of Students	
	Male	Female

The males and females are further classified according to their residence into hostellers or day scholars. It is a case of ***treble table.***

Table 3.7: Treble Table.

Marks	No . of Students			
	Male		Female	
	Hostellers	Day scholars	Hostellers	Day scholars

If they are again classified as belonging to different religions, nationalities, states, etc. It will be an example for ***multiple table.***

Table 3.8: Multiple Table.

Marks	Number of Students											
	Male						Female					
	Hostellers			Day Scholars			Hostellers			Day Scholars		
	Hindu	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*	*Hindu*	*Christian*	*Muslim*

Raw Data

The statistical information collected from the investigation is known as ***raw data.*** Suppose we are interested in the weight measurements of students of final B.Sc. class. There are 40 students in this class.

Their weight measurements are recorded as follows in Table 3.9.

Table 3.9: Weight measurements of 40 students.

161	156	153	146	163	152	147	164
145	145	135	168	154	144	140	150
135	128	147	126	120	157	158	125
142	135	142	138	144	148	146	135
150	140	173	176	165	136	138	150

In the above table we have 40 observations relating to the weight measurements of 40 students. This data is known as ***raw data.***

4. Diagrammatic Presentation of Data

Presentation of data** is the visual form of presenting the data.* Presentation of data helps ***easy understanding; It facilitates ***comparison*** of data; it helps for the quick arrival of ***inferences.***

Data are presented in two methods. They are ***diagrammatic presentation*** and ***graphic presentation.***

Presentation of Data

Diagrammatic Presentation ***Graphic Presentation***

In ***diagrammatic presentation,*** the data are presented in the form of ***diagrams*** such as bars, circles, maps, cartograms, etc. In ***graphic presentation***, the data are presented in the form of ***graphs.***

Diagram

One of the main functions of statistics is to simplify complex data. A ***diagram is a drawing or a design to explain (illustrate) something.***

It is a visual form for presentation of statistical data. It is an acknowledged fact that visual aids are more useful and interesting than dry numbers.

Diagram refers to the various types of devices such as ***bars, circles, maps, cartograms***, etc.

An ordinary man can understand pictures and diagrams more easily than the figures. The use of diagrams is becoming

more and more popular in the present time because they ar more appealing, attractive and meaningful.

Significance and Utility of Diagrams

1. Diagrams are more attractive.
2. They create more effects on the minds of th readers.
3. They simplify complexity.
4. They save time and labour.
5. They provide more information.
6. They make comparison.

Limitations of a Diagram

1. A diagram shows only approximate values.
2. It is a supplement to the tabular presentation, b not an alternative to it.
3. It cannot be analyzed further.
4. It cannot represent all details.
5. Small differences in large measurements cann be studied. For example, the differences between 1000 a 1005 shown in diagram, cannot be apparent.

Rules for Drawing a Diagram

1. Each diagram should have a suitable title.
2. A proper scale must be chosen for the diagram make more attractive.
3. Drawing should be neat and clean.
4. Foot notes should be given at the bottom of t diagram.
5. Index must be given for identifying a understanding the diagram.
6. The most important one is the selection of proper diagram to represent the data.

Kinds of Diagrams

There are different types of diagrams by wh statistical data can be presented. The common types are

1. *Line diagram* 3. *Pie diagram* 5. *Cartogram*
2. *Bar diagram* 4. *Pictogram*

1. Line Diagram

A line diagram is a diagram where *the statistical data are represented in the form of straight lines.* This is the simplest of all the diagrams. On the basis of size of the figures, heights of lines are drawn. The distance between lines is kept uniform. It makes comparison easy. This line diagram is not attractive so it is less important.

Illustration : *A traffic survey shows the following flow of vehicles passing a particular point during an hour. Draw a line diagram.*

Table 4.1: A traffic survey.

Vehicles	Frequency
Cars	45
Lorries	22
Motor Cycles	6
Buses	3
Total	76

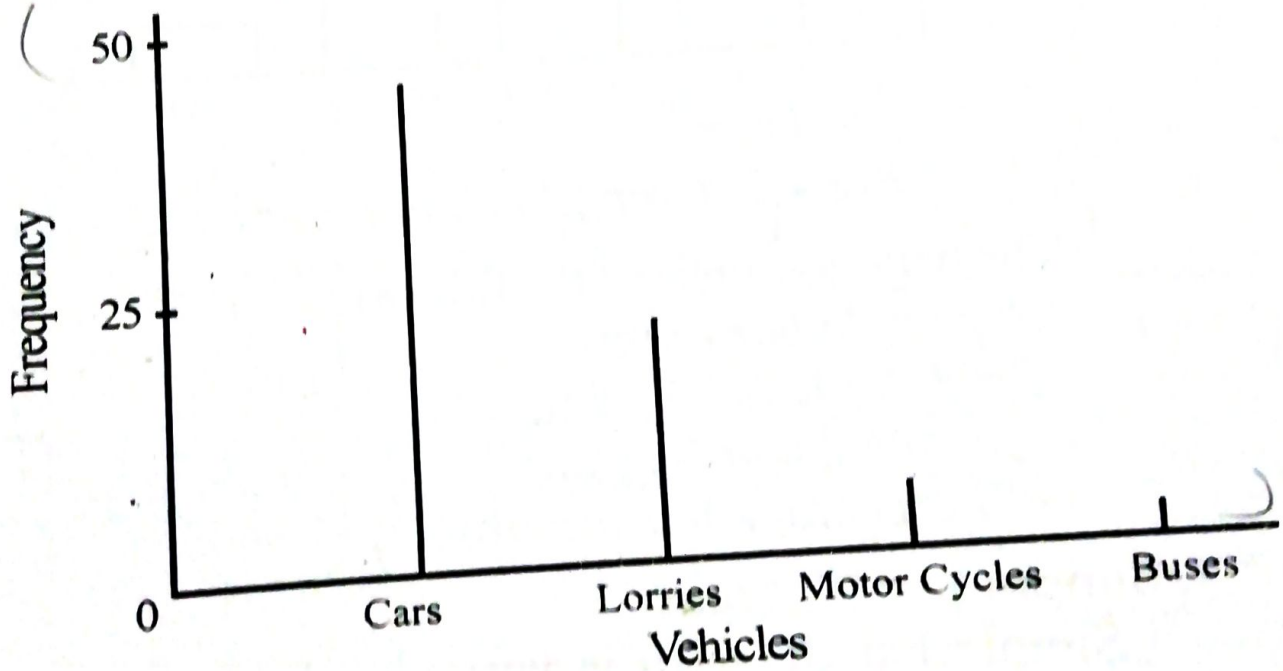


Fig.4.1: A line diagram.

The length or height of each line represents the frequency.

2. Bar Diagram

*In **bar diagram**, the data are drawn in the form of **rectangles** or **bars.***

The Y - axis is marked with frequency on a scale.

The X - axis is marked with variable on a scale.

Vertical rectangles are drawn as per the height of frequency. Equal space is given between the rectangles. The width of each rectangle should be also same.

Colours or shades may be given for the rectangles.

It is a ***one dimensional*** diagram as height alone is considered. The width is not considered.

The value may be written at the top of the rectangle.

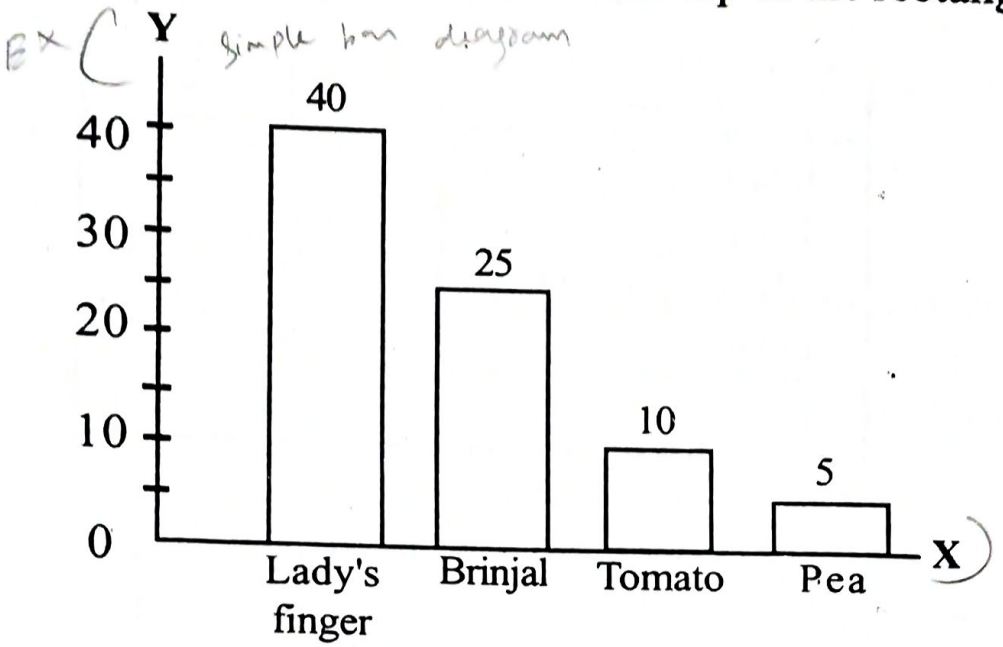


Fig.4.2: Crop plants in a garden.

The bar diagram is of the following types:

1. *Simple bar diagram*
2. *Multiple bar diagram*
3. *Component bar diagram*
4. *Percentage bar diagram*

1. Simple Bar Diagram

Simple bar diagram is very simple. It has ***vertical rectangles*** at regular intervals. The height of each rectangle

is corresponding to the data. The width of each rectangle should be the same. The rectangles may be given colours or shades. It is ***one - dimensional.*** The diagram given above is a simple bar diagram. *(Fig.4.2)*

2. Multiple Bar Diagram

The multiple bar diagram contains ***two*** or ***more bars*** drawn side by side. It is also called ***compound bar diagram.***

Height is proportional to the data. Width should be same. Equal spacing should be given.

It is a ***one - dimensional*** diagram.

It helps comparison.

Table 4. 2: Blood groups of boys and girls in 3 different classes.

Class	No. of Students	
	Boys	Girls
I B.Sc	40	35
II B.Sc	30	20
III B.Sc	20	15

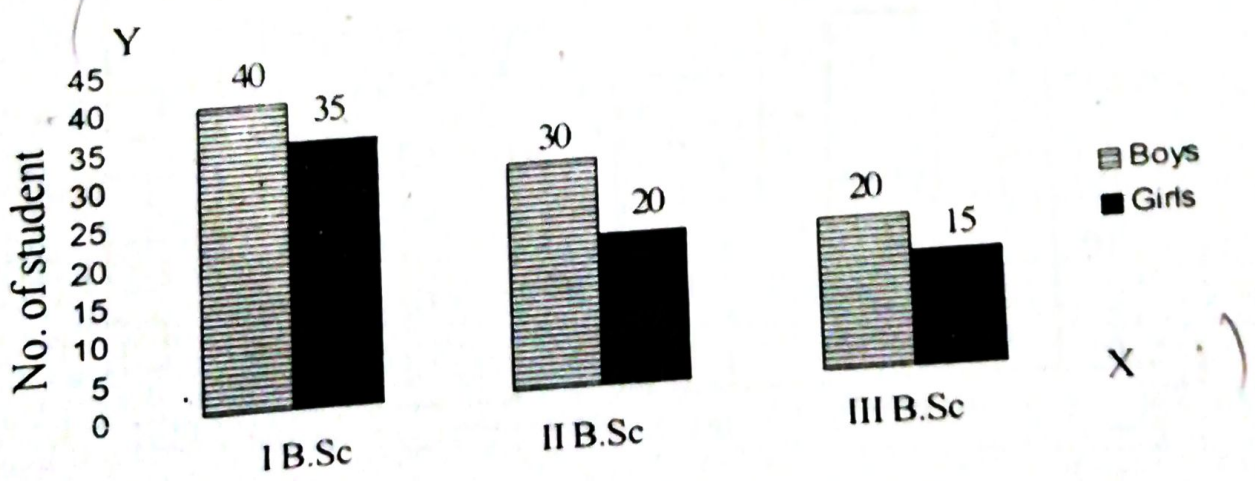


Fig.4.3: Multiple bar diagram.

3. Component Bar Diagram

In component bar diagram, each ***rectangle is subdivided into segments*** as per data. It is also called ***subdivided bar diagram.***

The height is proportional to the data. The width should be same. Equal space should be given between rectangles.

The largest component should be given at the base of data.

It helps comparison.

Table 4.3: Blood groups of three classes.

Blood Group	No. of Students		
	I B.Sc.	II B.Sc.	III B.Sc.
A	5	3	4
B	19	15	17
AB	9	6	9
O	7	4	7

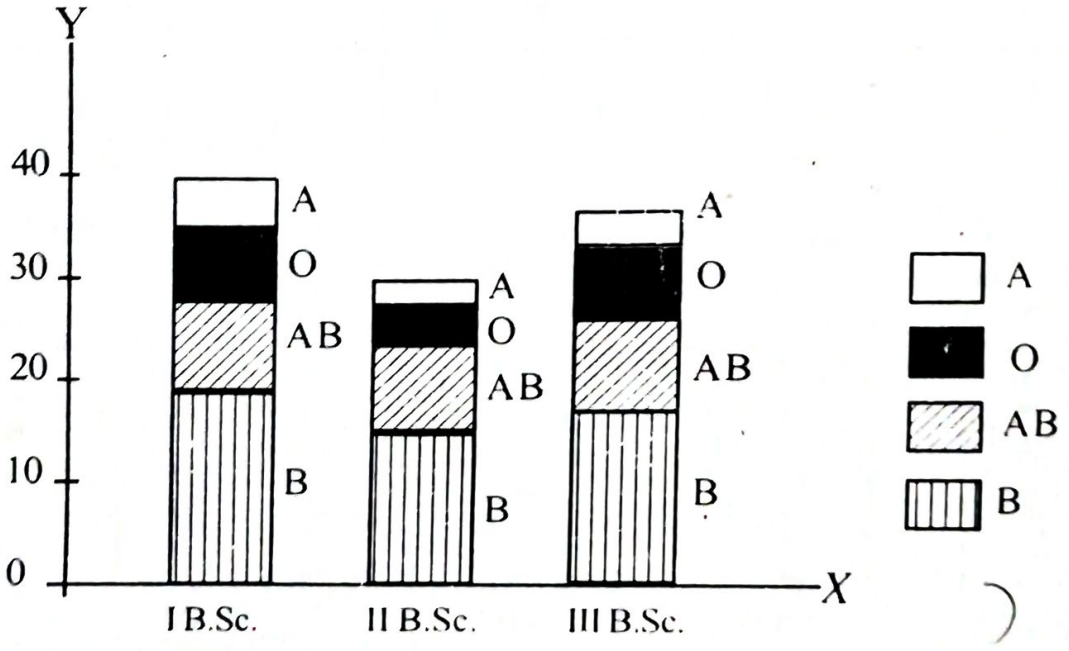


Fig.4.4: Sub divided bar diagram.

4. Percentage Bar Diagram

In percentage bar diagram, the ***rectangles are drawn as per the percentage of the data.***

All the rectangles are equal in height. Equal spacing is given between the bars.

Table 4.4: Blood groups of 3 classes.

Group	I B.Sc	II B.Sc	III B.Sc
A	5 %	10 %	15 %
B	55 %	50 %	45 %
AB	35 %	30 %	25 %
O	5 %	10 %	15 %

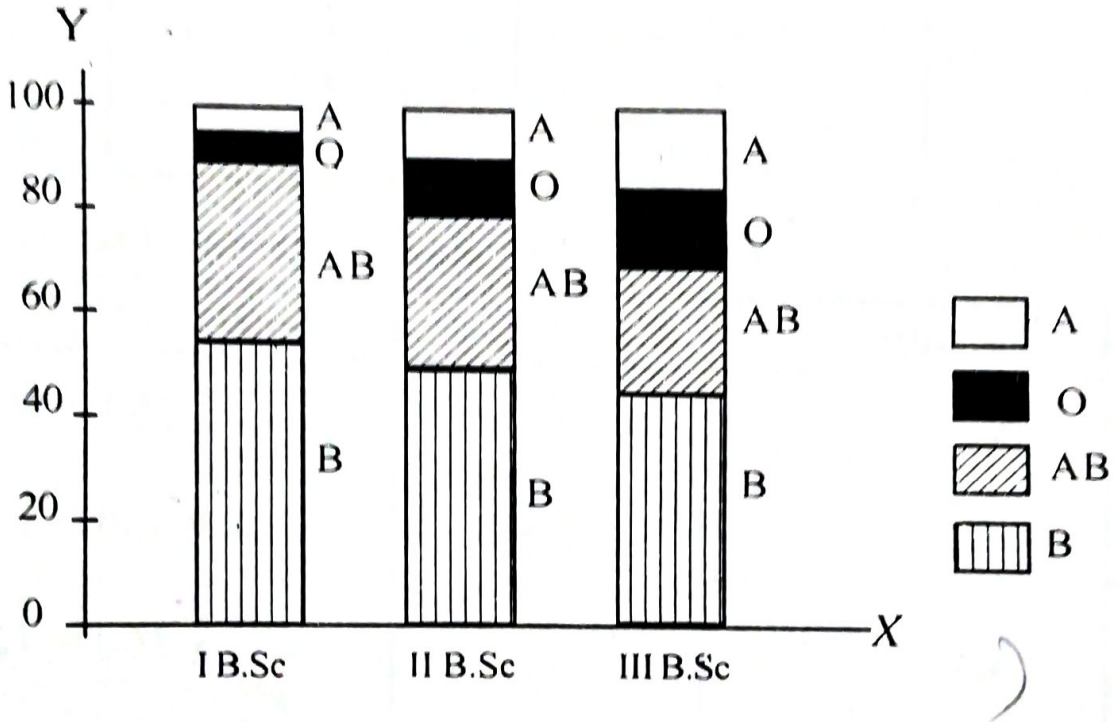


Fig.4.5: Percentage bar diagram.

Percentage bar diagram helps in comparison on relative basis.

3. Pie Diagram

In pie diagram, the data are presented in the form of a ***circle.***

Pie diagram is also called a ***circle diagram***.

The total area of a circle is 360^0. So the data is converted into degrees. Accordingly the circle is partitioned. As the data is drawn as per angles (degrees), pie diagram is also called ***angular diagram.***

The partitions are given different colours.

It is an ***area diagram.*** It is a ***two-dimensional diagram.***

The values are added to get the total.

Then each value is converted into degrees.

Table 4.5: Blood group of 50 students.

Group	Students
A	5
B	20
AB	10
O	15
Total	50

Table 4.6: Blood group of 50 students given in degrees.

Group	**Students**	**Degrees**
A	5	$\frac{5}{50} \times 360 = 36^\circ$
B	20	$\frac{20}{50} \times 360 = 144^\circ$
AB	10	$\frac{10}{50} \times 360 = 72^\circ$
O	15	$\frac{15}{50} \times 360 = 108^\circ$
Total	50	360°

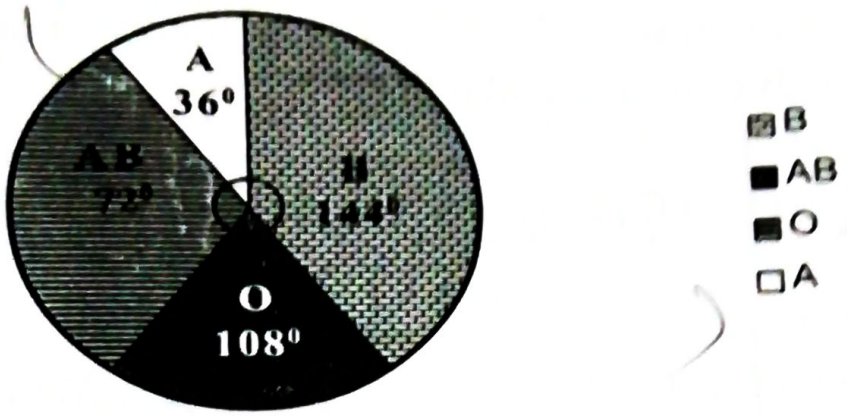


Fig.4.6: A Pie diagram.

This is done by dividing each value by the total and nultiplying it by 360 (360 is the total degree of a circle).

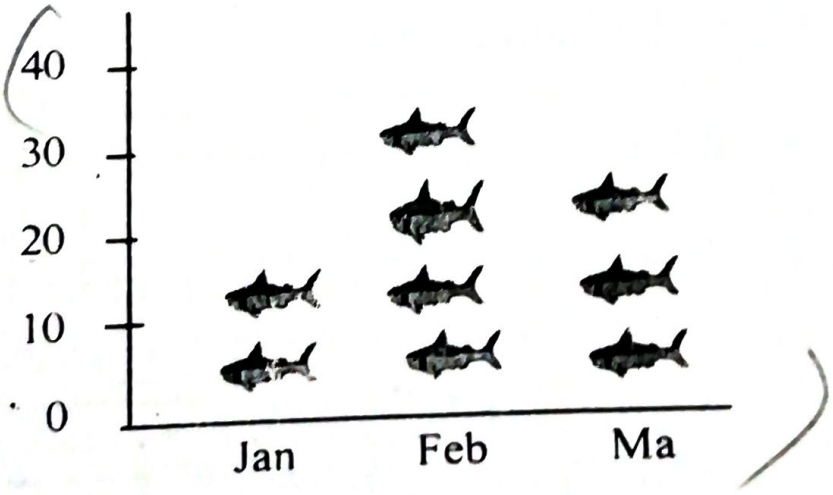
A circle is drawn. The degree for each value is measured vith a protractor.

Usually, the value with the highest degree is marked at 12'O clock position on the circle. Other values are marked n the decending order in the clockwise direction.

4. Pictogram

In pictogram, data are drawn in the form of ***pictures.***

It is simple and it helps lay man. *Eg. The data of fishes re represented in the form of fish pictures.*



g.4.7: Pictogram showing monthly fish landing in a pond.

5. Graphic Presentation of Data

Presenting data in the form of graphs is called ***graphic presentation of data.***

Graph

A *graph* is *the geometrical image of a data.*

A *graph* is *a diagram consisting of lines of statistical data.*

The graph is drawn on a ***graph paper.***

The graph has two intersecting lines called ***axes.***

The horizontal line is called ***X - axis.*** The vertical line is called ***Y - axis.***

The point of intersection is called ***'O'.***

The ***'O'*** point is common to both X and Y axis. Hence the ***X*** axis is also called ***OX line*** and the ***Y*** axis is also called ***OY line.***

A suitable scale is given for each axis.

Usually independent variables are marked on the X-axis and dependent variables are marked on the Y - axis.

A ***title*** is given to a graph.

The values corresponding to X and Y axis are plotted on the paper.

The points are joined with straight or curved lines.

Table 5.1: Monthly fish landing in a pond.

Month	Jan	Feb	Mar	Apr	May	June
Catch in tonnes	20	25	30	10	35	40

10. After the choice of the scale is made the last step in constructing a graph is to plot the given data by taking the corresponding values of X and Y. The various points so obtained are then joined by straight lines.

Table 5.2: Difference between Diagram and Graph.

	Diagram	Graph
1.	Ordinary paper can be used	Graph paper can be used.
2.	It is attractive and is easily understandable	It needs some effort to understand.
3.	Median and Mode cannot be estimated	Median and Mode can be estimated.

Graphs can be classified into ***two*** groups namely.

A. Graphs of time series or line graphs
B. Graphs of frequency distribution.

A. Graphs of Time Series (Line Graphs)

In a line graph the data is represented in ***straight lines.***

The line graph is divided into ***four*** types. They are :

1. Graph of one variable
2. Graph of two or more variables
3. Range chart
4. Band graph

1. Graph of One Variable

When only one variable is to be represented the desired graph is obtained by plotting the ***time variable*** along the X-axis and the ***value of variables*** on Y-axis on a suitable scale. The points are joined by straight lines.