Government of Tamilnadu

Department of Employment and Training

Course : TNPSC Group I Mains Material

Subject : General Aptitude & Mental Ability

Topic : **Conversion of Information to Data**

Commissioner,
Department of Employment and Training.

Conversion of Information to Data

Collection of data

We need to collect data having the specific information in our mind. Suppose the specific information needed by us is to know the height of class VII students, then we should collect specific data related to their heights and ages rather than the data related to health record of students.

From the above discussion, we conclude that the purpose for which a data is to be collected has to be kept in mind before starting the process of data collection. Then only we can get the desired information, which is appropriate to the purpose. Let us look into a few situations that are given below.

Data can be generated in many situations around us. For example,

- The height of class VII students.
- The performance of class VII students in the term II examinations.
- The number of trees planted in your locality.
- The highest temperatures recorded in all the major cities of India during the year 2018.
- The least amount of rainfall recorded in all the districts of Tamil Nadu during the year 2018.

Organisation of Data

We first collect data, record it and organise them. To understand this, consider an example which deals with the weights of 10 students of class VII. These are collected to know whether the weights of students are appropriate to their heights. The data is given below.

- Anbu-20 kg; Nambi-19 kg; Nanthitha- 20 kg; Arul- 24 kg;
- Mari-25 kg; Mathu-22 kg; Pavithra – 23 kg; Beeman- 26 kg;
- Arthi-21 kg; Kumanan-25 kg.

Let us try to answer the following questions.

(i) Who is the least weight of all?

(ii) How many students weigh between 22 kg to 24 kg?

(iii) Who is the heaviest of all?

(iv) How many children are above 23 kg and how many are below 23 kg?

The data mentioned above is not easy to understand.

If the data is arranged according to the order of weights, it will be easy for answering the questions. Observe the following table.

S. No	Name	Weight (kg)
1	Nambi	19
2	Anbu	20
3	Nanthitha	20
4	Arthi	21
5	Mathu	22

S. No	Name	Weight (kg)
6	Pavithra	23
7	Arul	24
8	Kumanan	25
9	Mari	25
10	Beeman	26

Now we can answer the above questions easily. Hence it is essential to organise the data to obtain any kind of inferences from the data.

Organisation of data is helpful to understand quickly and get an overall view of data. It becomes easy to understand and interpret which in turn also helps to take decision accordingly.

Data

In our daily life, we come across many situations where we need to collect information in the form of Facts or Numbers.

For example,

- Number of students in your class using calculators.
- Number of brothers and sisters in your family.
- Number of different types of houses in a village.
- Number of girls wearing bangles.
- Number of buses crossing a certain road junction at a particular time.
- Number of persons in a street who watch T.V. for more than 2 hours a day.
- Number of shops in a shopping mall selling textiles.
- Speeds of bikes, cars and other vehicles passed along a specific highway.

Thus, the numerical information or facts collected are known as Data.

The word 'data' was first used in 1640's. In 1946, the word 'data' also meant for "transmittable and storable computer information". In 1954, a term called 'data processing' was introduced. The plural form of 'datum' is 'data'. It also means "given" or "to give" in Latin.

Types of Diagrams

Example: 1

The production cost of the company in lakhs of rupees is given below.

(i) Construct a simple bar diagram.

(ii) Find in which year the production cost of the company is

(a) maximum (b) minimum (c) less than 40 lakhs.

(iii) What is the average production cost of the company?

(iv) What is the percentage increase from 2014 to 2015?

Year	Production Cost
2010	55
2011	40
2012	30
2013	25
2014	35
2015	70

Solution:

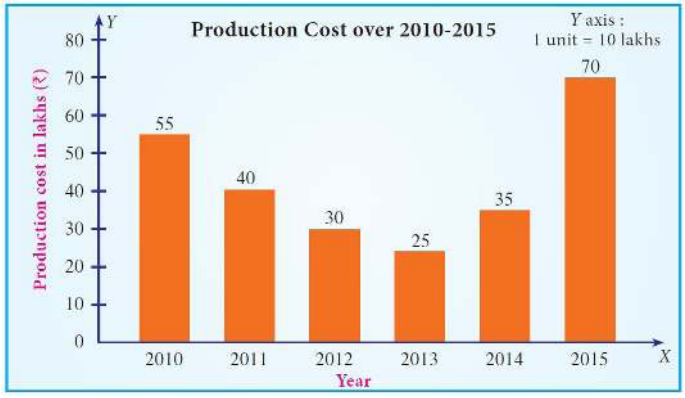
(i) We represent the above data by simple bar diagram in the following manner:

Step 1: Years are marked along the X-axis and labelled as 'Year'.

Step 2: Values of Production Cost are marked along the Y-axis and labelled as 'Production Cost (in lakhs of ₹).

Step 3: Vertical rectangular bars are erected on the years marked and whose height is proportional to the magnitude of the respective production cost.

Step 4: Vertical bars are filled with the same colours.



(ii) (a) The maximum production cost of the company was in the year 2015.

(b) The minimum production cost of the company was in the year 2013.

(c) The production cost of the company during the period 2012- 2014 is less than 40 lakhs.

(iii) Average production Cost of the company

$$= \frac{55+40+30+25+35+70}{6}$$

= ₹ 42.5 Lakhs

(iv) Percentage increase in the production cost of the company is

$$= \frac{70}{35} \times 100$$

$$= 200\%$$

Example: 2

Administration of a school wished to initiate suitable preventive measures against breakage of equipment in its Chemistry laboratory. Information collected about breakage of equipment occurred during the year 2017 in the laboratory are given below:

Equipment	No. of breakages
Burette	45
Conical flask	75
Test tube	150
Pipette	30

Draw Pareto Diagram for the above data. Which equipment requires more attention in order to reduce breakages?

Solution:

Since we have to find the vital few among the several, we draw Pareto diagram.

Step 1 : Arrange the equipment according to the descending order of the number of breakages.

Step 2 : Find the percentage of breakages for each equipment using the formula

$$= \frac{\text{No. of Breakages}}{\text{Total No. of Breakages}} \times 100$$

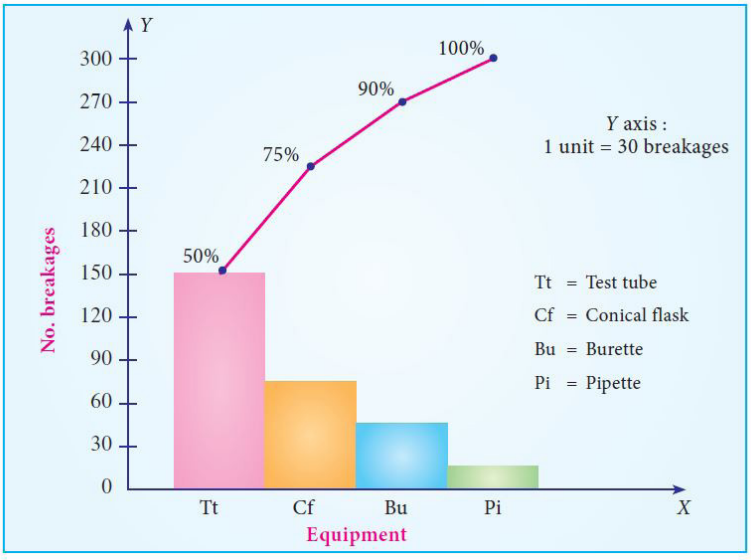
Step 3 : Calculate cumulative percentage for each equipment.

Step 4 : Mark the equipment along the X-axis and the number of breakages along the Y-axis. Construct an attached simple bar diagram to this data. In an attached simple bar diagram, the vertical bars are erected adjacently.

Step 5 : Mark the cumulative no. of breakages for each equipment corresponding to the mid-point of the respective vertical bar.

Step 6 : Draw a free hand curve joining those plotted points.

Equipment	No. of Breakages (Frequency)	No. of Breakages in percentage	Cumulative No. of Breakages in percentage
Test tube	150	$\frac{150}{300}\times 100 = 50$	50
Conical flask	75	$\frac{75}{300}\times 100 = 25$	75
Burette	45	$\frac{45}{300}\times 100 = 15$	90
Pipette	30	$\frac{30}{300}\times 100 = 10$	100
Total	**300**	**100**	



Pareto Diagram for No. of Breakages in the Chemistry Laboratory.

It can be found that 50% of breakages is due to Test tube, 25% due to Conical Flask. Therefore, the School Administration has to focus more attention on reducing the breakages of Test Tubes and Conical Flasks.

Example: 3

The table given below shows the profit obtained before and after tax payment (in lakhs of rupees) by a business man on selling cars from the year 2014 to 2017.

Year	Profit before Tax	Profit after Tax
2014	195	80
2015	200	87
2016	165	45
2017	140	32

(i) Construct a multiple bar diagram for the above data.

(ii) In which year, the company earned maximum profit before paying the tax?

(iii) In which year, the company earned minimum profit after paying the tax?

(iv) Find the difference between the average profit earned by the company before paying the tax and after paying the tax.

Solution:

Since we are comparing the profit earned before and after paying the tax by the same Company, the multiple bar diagram is drawn. The diagram is drawn following the procedure presented below:

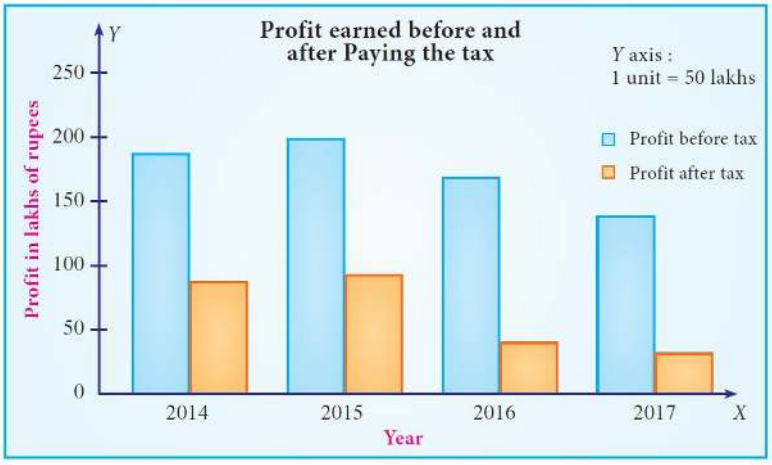
Step 1 : Years are marked along the X-axis and labeled as "Year".

Step 2 : Values of Profit before and after paying the tax are marked along the Y-axis and labeled as "Profit (in lakhs of `)".

Step 3 : Vertical rectangular bars are erected on the years marked, whose heights are proportional to the respective profit. The vertical bars corresponding to the profit earned before and after paying the tax in each year are placed adjacently.

Step 4 : The vertical bars drawn corresponding to the profit earned before paying the tax are filled with one type of colour. The vertical bars drawn corresponding to the profit earned after paying the tax are filled with another type of colour. The colouring procedure should be applied to all the years uniformly.

Step 5 : Legends are displayed to describe the different colours applied to the bars drawn for profit earned before and after paying the tax.



Multiple Bar Diagram for Profit by the Company earned before and after paying the Tax.

(i) The company earned the maximum profit before paying the tax in the year 2015.

(ii) The company earned the minimum profit after paying the tax in the year 2017.

(iii) The average profit earned before paying the tax = $\frac{700}{4}$ = ₹ 175 lakhs

The average profit earned after paying the tax = $\frac{244}{4}$ = ₹ 61 lakhs

Hence, difference between the average profit earned by the company before paying the tax and after paying the tax is = 175 – 61 = ₹ 114 lakhs.

Example: 4

Total expenditure incurred on various heads of two schools in an year are given below. Draw a suitable bar diagram.

Expenditure Head	Amount (in lakhs)	
	School I	School II
Construction/Repairs	80	90
Computers	35	50
Laboratory	30	25
Watering plants	45	40
Library books	40	30
Total	**230**	**235**

Which school had spent more amount for (a) construction/repairs (b) Watering plants?

Solution:

Since we are comparing the amount spent by two schools in a year towards various expenditures with respect to their total expenditures, a component bar diagram is drawn.

Step 1 : Schools are marked along the X-axis and labeled as "School".

Step 2 : Expenditure Head are marked along the Y-axis and labeled as "Expenditure (₹ in lakhs)".

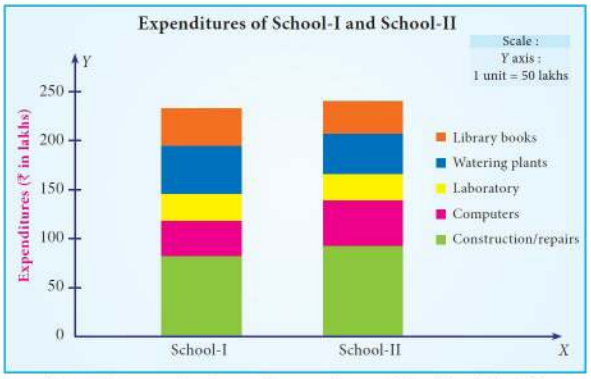
Step 3 : Vertical rectangular bars are erected for each school, whose heights are proportional to their respective total expenditure.

Step 4 : Each vertical bar is split into components in the order of the list of expenditure heads. Area of each rectangular box is proportional to the frequency of the respective expenditure head/component. Rectangular boxes for each school are coloured with different colours. Same colours are applied to the similar expenditure heads for each school.

Step 5 : Legends are displayed to describe the colours applied to the rectangular boxes drawn for various expenditure heads.

The component bar diagram is presented

Expenditure Head	Amount (in lakhs)			
	School I		School II	
	Amount Spent	Cumulative Amount Spent	Amount Spent	Cumulative Amount Spent
Construction/Repairs	80	80	90	90
Computers	35	115	50	140
Laboratory	30	145	25	165
Watering plants	45	190	40	205
Library books	40	230	30	235



Component Bar diagram for expenditures of School I and School II

(i) School- II had spent more amount towards Construction/Repairs.

(ii) School- I had spent more amount towards Watering plants.

Example: 5

Draw a pie diagram for the following data (in hundreds) of house hold expenditure of a family.

Items	Expenditure
Food	87
Clothing	24
Recreation	11
Education	13
Rent	25
Miscellaneous	20

Also find

(i) The central angle of the sector corresponding to the expenditure incurred on Education

(ii) By how much percentage the recreation cost is less than the Rent.

Solution:

The following procedure is followed to draw a Pie diagram for a given data:

(i) Calculate the total expenditure, say, N.

(ii) Compute angles for each component food, clothing, recreation, education, rent and

miscellaneous using the formula $\frac{\text{class frequency}}{N} \times 360$

Items	Expenditure	Angle of the circle
Food	87	$\frac{87}{180} \times 360 = 174$
Clothing	24	$\frac{24}{180} \times 360 = 48$
Recreation	11	$\frac{11}{180} \times 360 = 22$
Education	13	$\frac{13}{180} \times 360 = 26$
Rent	25	$\frac{25}{180} \times 360 = 50$
Miscellaneous	20	$\frac{20}{180} \times 360 = 40$
Total	**N=180**	**360**

(iii) Draw a circle with radius of sufficient length as a horizontal line.

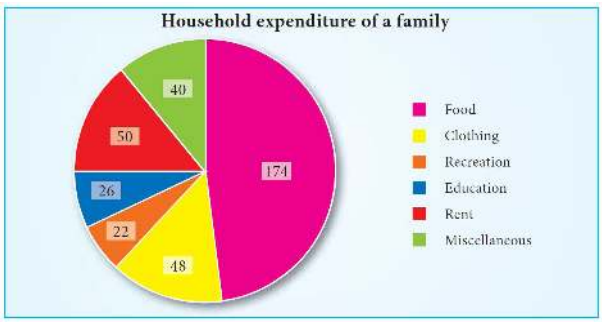
(iv) Draw the first sector in the anti-clockwise direction at an angle calculated for the first component food.

(v) Draw the second sector adjacent to the first sector at an angle corresponding to the second component clothing.

(vi) This process is continued for all the components namely recreation, education, rent and miscellaneous.

(vii) Shade/colour each sector with different shades/colours.

(viii) Write legends to each component.



Example: 6

Draw the histogram for the 50 students in a class whose heights (in cms) are given below.

Height	111 – 120	121 – 130	131 – 140	141 – 150	151 – 160	161 – 170
Number of students	4	11	15	9	8	3

Find the range, whose height of students are maximum.

Solution:

Since we are displaying the distribution of Height and Number of students in visual form, the histogram is drawn.

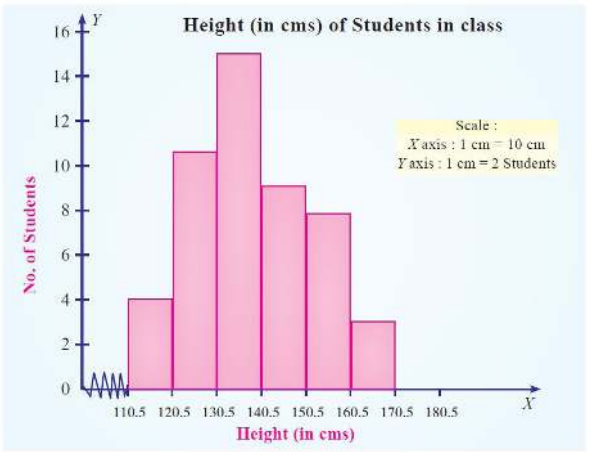
Step 1 : Heights are marked along the X-axis and labeled as "Height(in cms)".

Step 2 : Number of students are marked along the Y-axis and labeled as "No. of students".

Step 3 : Corresponding to each Heights, a vertical attached bar is drawn whose height is proportional to the number of students.

For drawing a histogram, the frequency distribution should be continuous. If it is not continuous, then make it continuous as follows.

Height (in Cm)	No.of Students
110.5 - 120.5	4
120.5 - 130.5	11
130.5 - 140.5	15
140.5 - 150.5	9
150.5 - 160.5	8
160.5 - 170.5	3



Histogram for heights of students in a class.

The tallest bar shows that maximum number of students height are in the range 130.5 to 140.5 cm

Data in Tabular form: To make the given data easily understandable, we tabulate the data in the form of tables or charts. A table has three columns that contains
(i) Variable /Class (ii) Tally Marks (iii) Frequency

Variable / Class: Arrange the given data from the lowest to the highest in the first column under the heading variable or class.

Tally Marks: A vertical line(|) which is marked against each item falling in the variable /class is called tally marks.

Frequency: The number of times an observation occurs in the given data is called the frequency of the observation. This is easily counted from the tally marks column.

For example:

Variable (Marks)	Tally marks	Frequency (f) (no. of students)
10	|||	3
14	𝍸	5
17	𝍸 |||	8
20	||||	4
	Total	20

From the table, we understand that three students got 10 marks, five students got 14 marks and so on.

Ungrouped data or Discrete Data:
An ungrouped data can assume only whole numbers and exact measurement. These are the data that cannot have a range of values. A usual way to represent this is by using Bar graphs .

Examples: 1. The number of teachers in a school.
2. The number of players in a game.

Grouped data or Continuous Data:
A grouped data is any value within a certain interval. The data can take values between certain range with the highest and the lowest value. Continuous data can be tabulated in what is called as frequency distribution. They can be graphically represented using Histograms.

Example: 1. The age of persons in a village.
2. The height and the weight of the students of your class.

FREQUENCY DISTRIBUTION TABLE

Frequency distribution:

A frequency distribution is the arrangement of the given data in the form of the table showing frequency with which each variable occurs.

If we have more number of students in the class , it would be very difficult to understand and to get information unless it is organised. For this reason, we organise larger data into a table called the frequency distribution table. Therefore, the tabular arrangement which shows the observations and their frequency of occurrences is called the frequency distribution table.

There are two types of distribution table namely

(i) frequency distribution table for ungrouped data and
(ii) frequency distribution table for grouped data.

Construction of frequency distribution table for ungrouped data.

Example: 1

Form an ungrouped frequency distribution table for the weight of 25 students in STD IV given below and answer the following questions.

25, 24, 20, 25, 16, 15, 18, 20, 25, 16, 20, 16, 15, 18, 25, 16, 24, 18, 25, 15, 27, 20, 20, 27, 25.

(i) Find the range of the weights.
(ii) How many of the students has the highest weight in the class?
(iii) What is the weight to which more number of students belong to?
(iv) How many of them belong to the least weight?

Solution:

To form a distribution table, arrange the given data in ascending order under Weight column then, put a vertical mark against each variable under Tally marks column and count the number of tally marks against the variable and enter it in Frequency column as given below. Hence, the distribution table is

Weight	Tally marks	Frequency
15	|||	3
16	||||	4
18	|||	3
20	𝍸	5
24	||	2
25	𝍸 |	6
27	||	2
	Total	25

Thus, we can tabulate the above table as follows.

Weight (kg)	15	16	18	20	24	25	27
Frequency	3	4	3	5	2	6	2

(i) The range of the given data is the difference between the largest and the smallest value. Here, the range = 27–15 =12.

(ii) From this table, two of the students have the highest weight of 27 kg.

(iii) 6 students belong to 25 kg weight.

(iv) 3 students belong to the least weight of 15 kg.

So, when we tabulate the given data, it is easy to get the information at a glance, Isn't it?

Construction of frequency distribution table for grouped data:

Now, we will consider a situation, if we collect data of marks for 50 students, it becomes very diffi cult to put tally for each and every marks of all the 50 students. Because if we arrange the marks in a table, it will be very large in length and not understandable at once. In this case, we use class intervals. In this table, consider the groups of data in the form of class intervals to tally the frequency for the given data.

Class Interval:

The range of the variable is grouped into number of classes, and each group is known as class interval (C.I). The difference between the upper limit (U) and the lower limit (L) of the class is known as class size.

i.e. C.I = Upper limit – Lower limit

Construction of grouped frequency distribution table – Continuous Series

Example:

The EB bill(in ₹) of each of the 26 houses in a village are given below. Construct the frequency table.

215	200	120	350	800	600	350	400	180	210	170	305	204
200	425	540	315	640	700	790	340	586	660	785	290	300

Solution:

Maximum bill amount = ₹ 800

Minimum bill amount = ₹ 120

Range = maximum value – minimum value

Range = 800 – 120 = ` 680

Suppose if we want to take class size as 100, then

$$\text{The number of possible class intervals} = \frac{\text{Range}}{\text{Class size}}$$

$$= \frac{680}{100} = 6.8 \simeq 7$$

Class Intervals	Tally marks	Frequency
100-200	|||	3
200-300	𝍸 |	6
300-400	𝍸 |	6
400-500	||	2
500-600	||	2
600-700	|||	3
700-800	||||	4
	Total	26

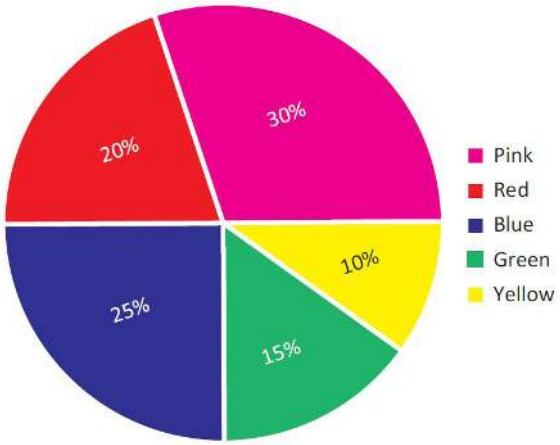
Graphical representation of the frequency distribution for ungrouped data

A graphical representation is the geometrical image of a set of data. It is a mathematical picture. It enables us to think about a statistical problem in visual terms. A picture is said to be more effective than words for describing a particular thing. The graphical representation of data is more effective for understanding. In the previous classes, we have studied some graphical representations of ungrouped data such as Line graph, Bar graph, and Pictograph. Now, we are going to represent the given ungrouped data in the circular form namely the pie diagram or the pie chart.

Pie chart (or) Pie diagram

A pie chart is a circular graph which shows the total value with its components. The area of a circle represents the total value and the different sectors of the circle represent the different components. The circle is divided into sectors and the area of the sectors is proportional to the information given. In the 'pie chart' the data are mostly expressed in percentage. Each component is expressed as percentage of the total value.

The Pie diagram is so called because the entire graph looks like an American food 'pie' and the components resemble slices cut from 'pie'.



Method of constructing a pie chart:

In a pie chart, we know that the various components are represented by the sectors of a circle and the whole circle represents the sum of the value of all the components. Therefore, the total angle of 360° at the centre of the circle is divided into different sectors according to the value of the components.

$$\textbf{The central angle of a component} = \frac{\text{Value of the component}}{\text{Total value}} \times 360°$$

Sometimes, the value of the components are expressed in percentage. In such cases,

$$\textbf{The central angle of a component} = \frac{\text{Percentage value of the component}}{100} \times 360°$$

Steps for construction of the pie chart:

1) Calculate the central angle for each component using the above formula and tabulate it.
2) Draw a circle of convenient radius and mark one horizontal radius in it.
3) Draw radius making central angle of first component with horizontal radius. This sector represents the first component. From this radius, draw next radius with central angle of the second component and so on, until the completion of all components.
4) For identification of each sector, shade with different colours.
5) Label each sector.

Here are given some examples, let us draw the pie chart for the given data.

Example

The number of hours spent by a school student on various activities on a working day is given below. Construct a pie chart.

Activity	Sleep	School	Play	Home work	Other
No of hours	8	6	2	3	5

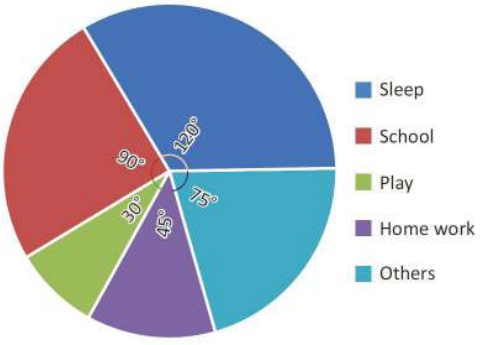
1. Find the percentage of sleeping hours.
2. By what angle is home work more than play?
3. By what angle are other activities less than sleep?

Solution:

Number of hours spent in different activities in a day of 24 hours are converted into components parts of 360°.

Activity	Duration in hours	Central angle
Sleep	8	$\frac{8}{24} \times 360° = 120°$
School	6	$\frac{6}{24} \times 360° = 90°$
Play	2	$\frac{2}{24} \times 360° = 30°$
Home work	3	$\frac{3}{24} \times 360° = 45°$
Other	5	$\frac{5}{24} \times 360^{\circ} = 75°$
Total	**24**	**360°**

g a day (24 hours)



1. The percentage of sleeping hours = 8/24 × 100 = 33.33%
2. Home work is 45°–30°=15° more than play
3. Other activities are 120°–75°=45° less than sleep.

Uses of pie chart:

1. Pie charts are widely used by the business and the media people.
2. With the help of Pie charts, one can show how the expenditure of the Government or Industry is distributed over different heads.
3. Research people use these type of charts to show their results.

MERITS AND DEMERITS:

Merits:

1. Simple to create.
2. Pie charts are visually simple than other types of graphs.
3. Pie charts are easy to understand information quickly.

Demerits:

1. Pie charts are inconvenient for comparing more than one sample.
2. Separate Pie charts have to be used for different samples.
3. It becomes less effective if there are more components in the data.

Graphical representation of the frequency distribution for grouped data

The Line graph, Bar graph, Pictograph and the Pie chart are the graphical representations of the frequency distribution for ungrouped data. Histogram, Frequency polygon, Frequency curve, Cumulative frequency curves (Ogives) are some of the graphical representations of the frequency distribution for grouped data.

Frequency Polygon

A frequency polygon is a line graph for the graphical representation of the frequency distribution. If we mark the midpoints on the top of the rectangles in a histogram and join them by straight lines, the figure so formed is called a frequency polygon. It is called a polygon as it consists of a number of lines as the sides of a polygon.

A frequency polygon is useful in comparing two or more frequency distributions. A frequency polygon for a grouped frequency distribution can be constructed in two ways.

i) Using a histogram ii) Without using a histogram

To construct a frequency polygon using a histogram:

1. Draw a histogram from the given data.
2. Join the consecutive midpoints of the upper sides of the adjacent rectangles of the histogram by the line segments.
3. It is assumed that the class interval preceding the first rectangle and the class interval succeeding the last rectangle exists in the histogram and the frequency of each extreme class interval is zero. These class intervals are known as imagined class intervals.
4. To get frequency polygon, join the midpoints of these imagined classes with

the corresponding midpoints of the upper sides of the first and last rectangles of the histogram.

Example

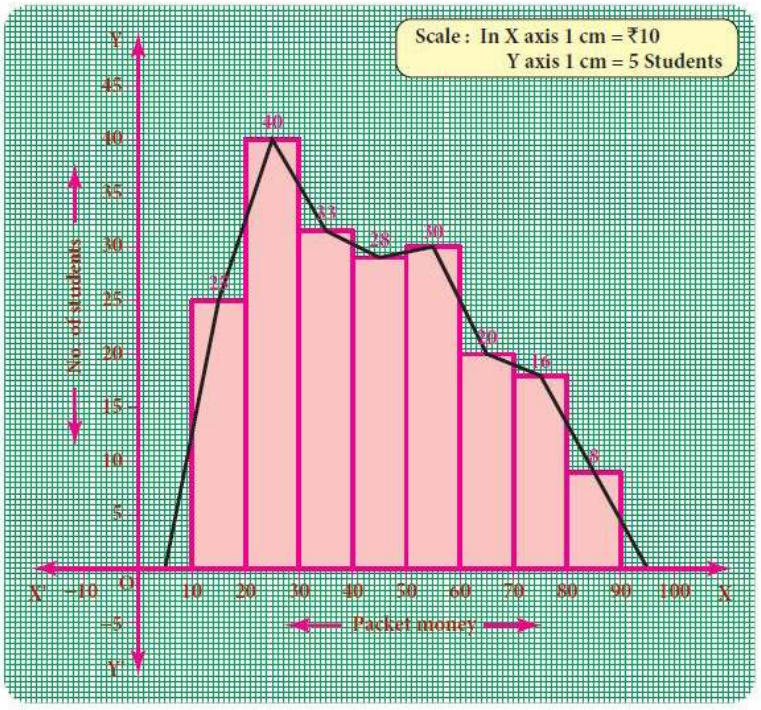
The following is the distribution of pocket money of 200 students in a school.

Pocket money	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Number of Students	25	40	33	28	30	20	16	8

Draw a frequency polygon using histogram.

Solution:

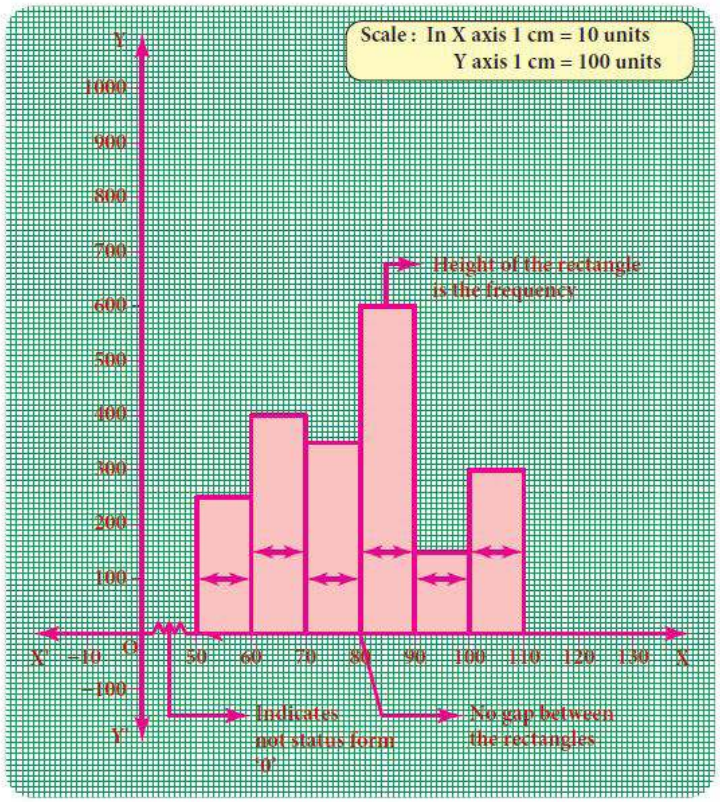
Represent the pocket money along x- axis and number of students along the y–axis. Draw a histogram for the given data. Now, mark the midpoints of the upper sides of the consecutive rectangles. Also mark the midpoints of two imagined class intervals 0-10 and 90-100 whose frequency is 0 on x- axis. Now, join all the midpoints with the help of ruler. We get a frequency polygon imposed on the histogram.



Histogram

A histogram is a graph of a continuous frequency distribution. Histogram contains a set of rectangles, the base of which is the length of the class interval and the frequency in each class interval is its height. i.e the class intervals are represented on the horizontal axis (x- axis) and the frequencies are represented on the vertical axis (y-axis).

The area of each rectangle is proportional to the frequency in the respective class interval and the total area of the histogram is proportional to the total frequency. Because of the continuous frequency distribution, the rectangles are placed continuously side by side with no gap between adjacent rectangles.



Steps to construct a Histogram:

1. Represent the data in the continuous form (exclusive form) if it is in discontinuous form (inclusive form) by converting it using the adjustment factor.
2. Select the appropriate units along the x-axis and y-axis.
3. Plot the lower limits of all class interval on the x –axis.
4. Plot the frequencies of the distribution on the y – axis.
5. Construct the rectangles with class intervals as bases and corresponding frequencies as heights. Each class has lower and upper values. This gives us two equal vertical lines representing the frequencies. The upper ends of the lines are joined together and this process will give us rectangles.

Example

Draw a histogram for the following table which represents the age groups from 100 people in a village.

Ages	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Number of people	11	9	8	20	25	10	8	6	3

Solution:

The given data is a continuous frequency distribution. The class intervals are drawn on x-axis and their respective frequencies on y-axis. Classes (ages) and its frequencies (number of people) are taken together to form a rectangle.

The histogram is constructed as given below.