FORMATION OF FREQUENCY DISTRIBUTION

1) Discrete frequency distribution

In discrete frequency distribution, values of the variable is arranged individually. The frequencies of the various values are the number of times each value occurs.

The weekly wages in Rs. paid to the workers are given below. Form a discrete frequency distribution

300,240,240.150,120,240,120,120,150,150,150,240,150,150,120,300,120,150,240,150,150,120,240,150,240,150,120,120,240,150.

Weekly wages in Rs.(x)	Tally marks	No. of workers
120	HHI III	8
150	HHI HHI II	12
240	HHI III	8
300	II	2

Weekly wages in Rs.(x)	120	150	240	300	TOTAL
No. of workers(f)	8	12	8	2	30

2) Continuous frequency distribution

Continuous frequency distribution is also known as grouped frequency distribution. This is a form in which class intervals and respective class frequencies are given.

We can find frequency distribution by the following steps:

- First of all, calculate the range of the data set.

- Next, divide the range by the number of the group you want your data in and then round up.

- After that, use class width to create groups

- Finally, find the frequency for each group.

Let us consider the following example regarding daily maximum temperatures in $°C$ in a city for 50 days.

28, 28, 31, 29, 35, 33, 28, 31, 34, 29,

25, 27, 29, 33, 30, 31, 32, 26, 26, 21,

21, 20, 22, 24, 28, 30, 34, 33, 35, 29,

23, 21, 20, 19, 19, 18, 19, 17, 20, 19,

18, 18, 19, 27, 17, 18, 20, 21, 18, 19.

- Minimum Value= 17
- Maximum Value=35
- Range=35-17=18
- Number of classes=5 (say)
- $\therefore$ width of each class=4

Table Showing frequency distribution of temperature in a city for 50 days.

Class Intervals (Temperatures in $°C$)	Frequency
17-21	17
21-25	7
25-29	10
29-33	9
33-36	7
Total	50

Terms used in continuous frequency distribution

Class Interval: The whole range of variable values is classified in some groups in the form of intervals. Each interval is called a class interval.

Class Frequency: The number of observations in a class is termed as the frequency of the class or class frequency.

Class limits and Class boundaries:

Class limits are the two endpoints of a class interval which are used for the construction of a frequency distribution. The lowest value of the variable that can be included in a class interval is called the lower class limit of that class interval. The highest value of the variable that can be included in a class interval is called the upper-class limit of that class interval. These are not the real limits or endpoints of a class interval. Hence, class limits are called apparent limits of a class.

In the previous data, the class intervals are 17-21, 21-25, 25-29, 29-33 and 33-36. Here, say for the class 17-21, the lower-class limit is 17 and the upper-class limit is 21. If there was an observation of 17, it would not be included in this class. An observation of 21, would be included in the class 21-24 and not in 17-21. This type is called as exclusive type.

Table Showing frequency distribution of temperature in a city for 50 days.

Class Intervals (Temperatures in $^\circ C$)	Frequency
17-21	17
21-25	7
25-29	10
29-33	9
33-36	7
Total	50

Class Intervals (Temperatures in $^\circ C$)	Frequency
17-20	17
21-24	7
25-28	10
29-32	9
33-36	7
Total	50

The second table is inclusive type of data, i,e,, if a value is equal to 17 or 20 it will be included in the first class itself.

The inclusive type of data can be converted to exclusive type as follows

We obtain class boundaries from class limits by dividing the difference between the upper limit of a class and the lower limit of the next higher class into two equal parts. Say, we are considering the classes 17-20 and 21-24. 21-20=1. Again, we have $1 \div 2 = 0.5$. We add 0.5 to the upper-class limit of each class and subtract 0.5 from the lower-class limit of each class. So, the class boundaries are 16.5-20.5, 20.5-24.5 and so on. For the class 16.5-20.5, 16.5 is the lower-class boundary and 20.5 is the upper-class boundary. It should be noted that the upper-

class boundary of the lower class coincides with the upper-class boundary of the next higher class.

Open-end classes: It may be so that some values in the data set are extremely small compared to the other values of the data set and similarly some values are extremely large in comparison. Then what we do is we do not use the lower limit of the first class and the upper limit of the last class. Such classes are called open end classes.

Class Intervals (Temperatures in $^\circ C$)	Frequency
below-21	17
21-25	7
25-29	10
29-33	9
Above 33	7
Total	50

Size of the Class: The length of the class is called the class width. It is also known as class size.

Class interval or size of the class = U.C.B. – L.C.B.

U.C.B. is Upper Class Boundary

L.C.B. is Lower Class Boundary

Mid-point of the Class: The midpoint of a class interval is called Mid-point of the Class. It is the representative value of the entire class.

Mid-point of the class = (U.C.B. +L.C.B.)/2

Cumulative frequency distribution

There are two kinds of cumulative frequency distribution.

i) Less than cumulative frequency distribution and

ii)   More than cumulative frequency distribution

i) Less than cumulative frequency distribution

Frequency distribution both discrete and continuous are to be taken in ascending order. The total of the frequencies from the beginning up to and including each frequency is found. That cumulative frequency shows how many items are less than or equal to the corresponding value of the class interval.

ii) More than cumulative frequency distribution:

Frequency distribution both discrete and continuous are to be taken in ascending order. The total of the frequencies from the end up to and including each frequency is found. That cumulative frequency shows how many items are more than or equal to the corresponding value of the class interval.

Weekly wages in Rs.(X)	No. of workers(f)	Weekly wages in Rs.	Less than cumulative frequencies	Weekly wages in Rs.	more than cumulative frequencies
120	8	120	8	120	30
150	12	150	20	150	22
240	8	240	28	240	10
300	2	300	30	300	2

Marks(x)	No. of students(f)	Marks below	No. of students	Marks above	No, of students
		Upper limit	Less than c.f.	Lower limit	More than c.f.
0 - 20	2	20	2	0	40
20 - 40	7	40	2+7=9	20	38
40 - 60	15	60	9+15=24	40	31
60 - 80	9	80	24+9=33	60	16
80 - 100	7	100	33+7=40	80	7
Total	40				

Graphical presentation

Graphs are charts consisting of points, lines and curves. Charts are drawn on graph sheets. Scales are to be chosen suitably in both X and Y axes so that entire data can be presented in the graph sheet.

Statistical measures such as quartiles, median and mode can be found from the appropriate graph.

Graphs are useful for analysis of timeseries, regression analysis, business forecasting, interpolation, extrapolation, and inverse interpolation.

Types of graphs

Graphs are broadly divided into two

   i)   Graphs of time series or Historigrams and
   ii)   Graphs of frequency distribution

Graphs of frequency distribution is further divided into

   a) Frequency lines used to represent discrete frequency distribution
   b) Histogram
   c) Frequency polygon   used to present continuous frequency distribution
   d) Frequency curve
   e) Ogive curve used to represent cumulative frequency distribution

Graphs of time series or Historigrams

Time series are the values of the variables observed at different points of time.

A historigram is a graph to show a time series. It shows the fluctuation of a variable over a given period. X axis is used to denote the time and Y axis the value of the variable. Each pair of (time, variable) is denoted by a point on the graph. After plotting all such points, successive points are joined by straight lines. The resulting curve is historigram.

year	sales in crores
2003	4.2
2004	3.9
2005	4.8
2006	5.3
2007	5.8
2008	4.9
2009	6.2
2010	5.9
2011	4.9
2012	5.6

Graphs of frequency distribution

i) **Frequency lines:** Are used to represent discrete frequency distribution.
Variable is taken in the X axis and the frequencies in the Y axis, the pair of points (variable, frequency) is plotted on the graph. From each point a vertical line is drawn to the X axis. The resulting graph is the frequency line.

Weekly wages in Rs.	No. of workers
120	8
150	12
240	8
300	2
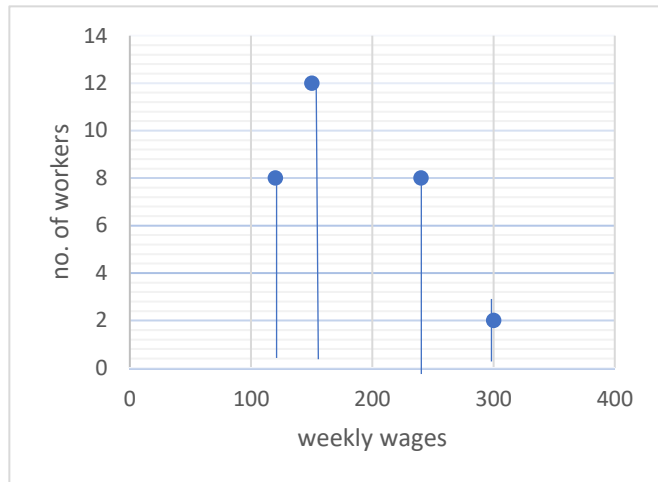
Histogram, frequency polygon and frequency curve

All the three graphs are used to represent continuous frequency distribution.

Histograms are similar to bar diagram without any gap in between. The variable is taken in the X axis and the frequency in the Y axis, The width of the bars or rectangles are equal to the class interval and the height is equal to the frequency.

Frequency polygon and frequency curves are drawn along with the histogram or separately,

If it is drawn along with the histogram, the midpoint of the horizontal line on top of the bar is plotted and the successive points are joined by straight lines or a smooth curve to form the frequency polygon or frequency curve respectively.

If drawn separately the pair of points (midpoint, frequency) are plotted and the successive points are joined by straight lines or smooth curves to form the frequency polygon or frequency curve respectively.

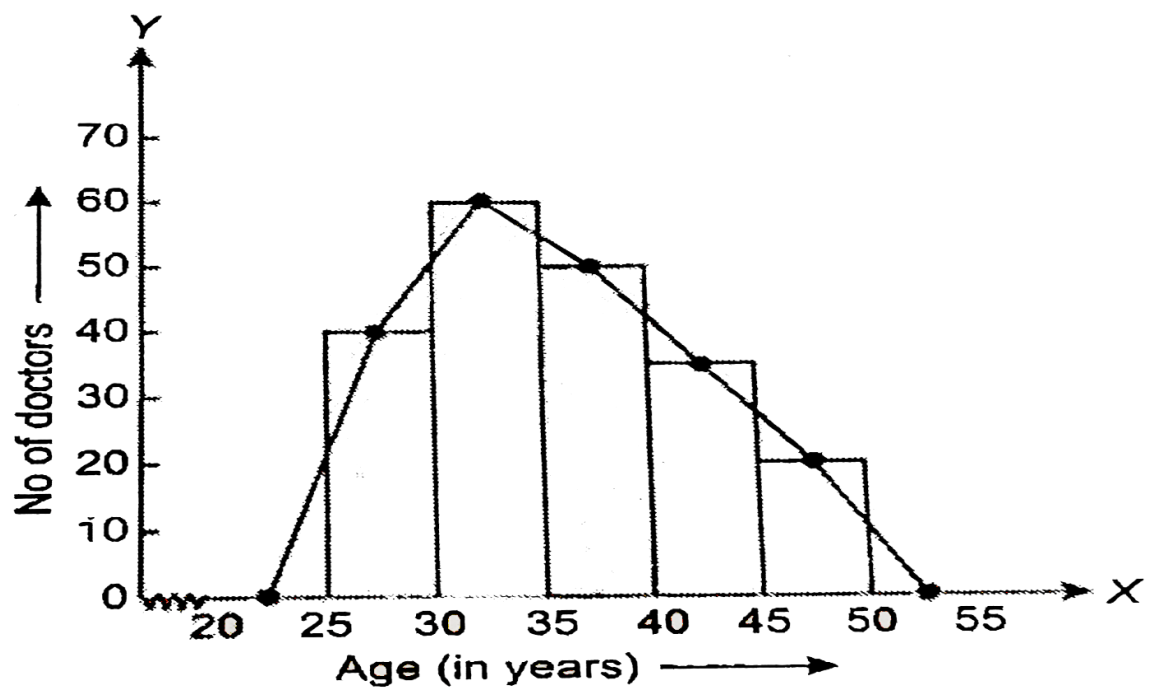| Age in years	Number of people	midpoints
0 - 5	4	(0+5)/2=2.5
5 – 10	28	7,.5
10 - 15	43	12.5
15 - 20	54	17.5
20 - 25	38	22.5
25 - 30	24	27.5

Draw a Histogram for the above data
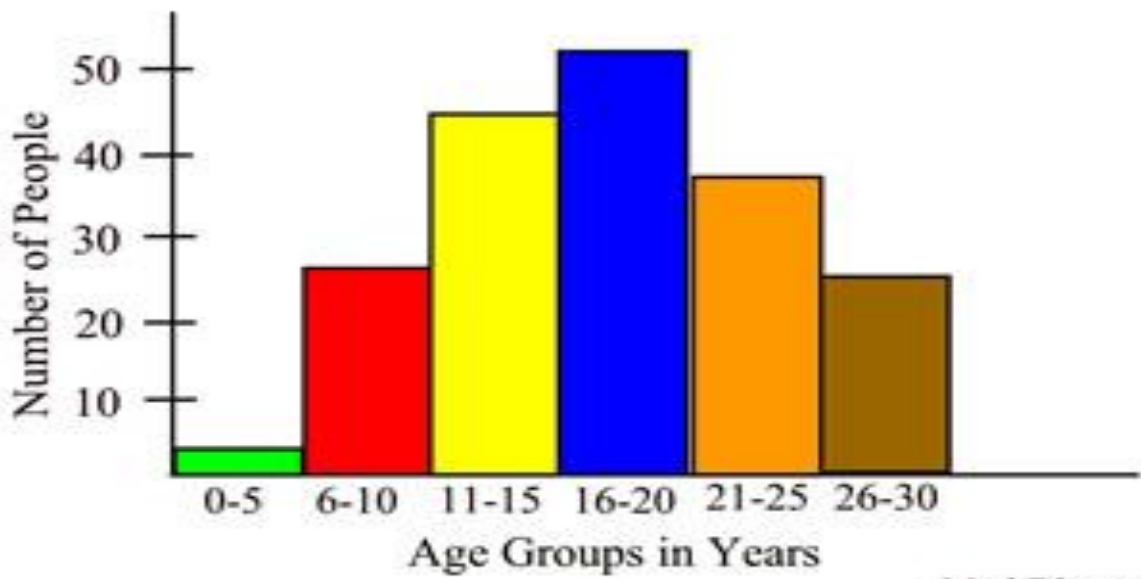
M&M as Favorite Candy
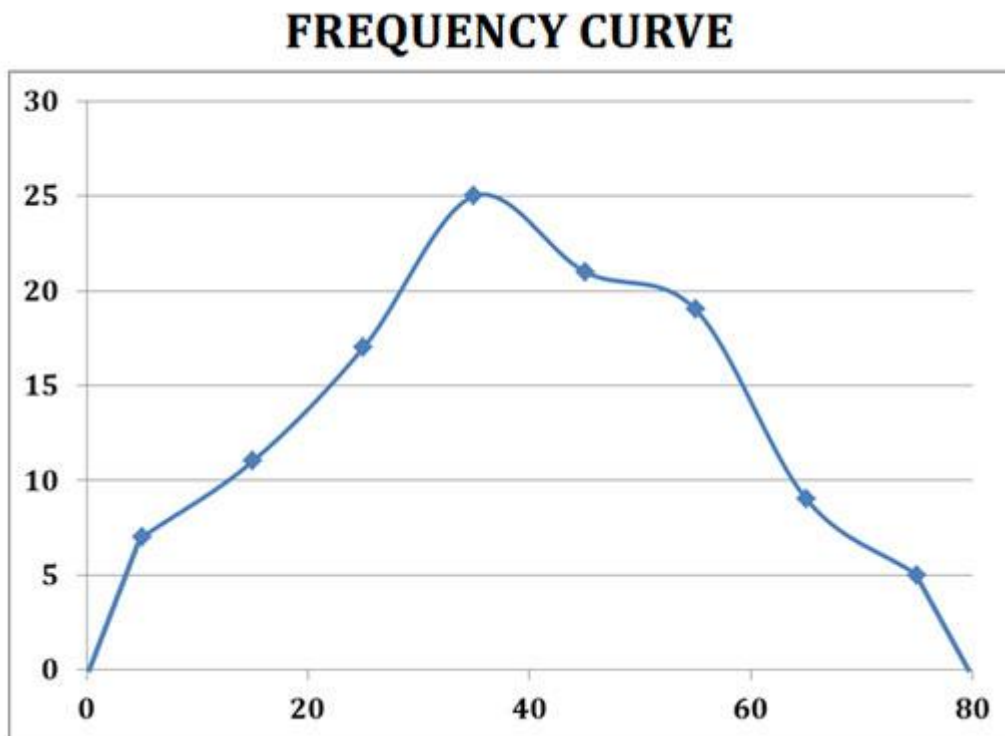
MathBits.com

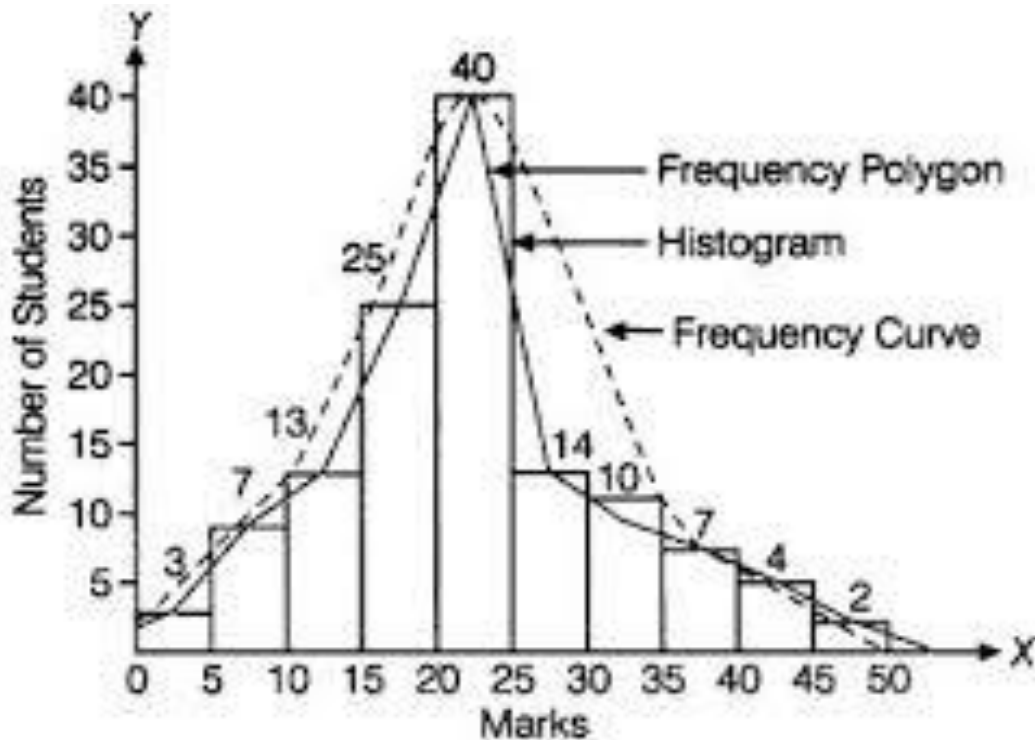**Frequency polygon**

Test Scores
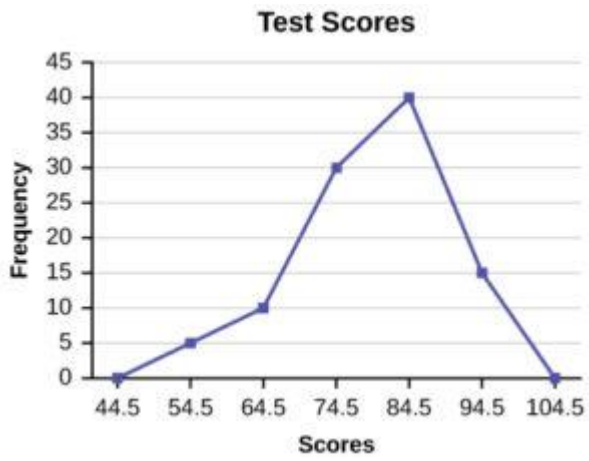
Frequency curve

FREQUENCY CURVE

Ogive curve or cumulative frequency curves

An Ogive Chart is a curve of the cumulative frequency distribution or cumulative relative frequency distribution. Below are the steps to construct the less than and greater than Ogive.

How to Draw Less Than Ogive Curve?

- Draw and mark the horizontal and vertical axes.
- Take the cumulative frequencies along the y-axis (vertical axis) and the upper-class limits on the x-axis (horizontal axis).
- Against each upper-class limit, plot the cumulative frequencies.
- Connect the points with a continuous curve.

How to Draw Greater than or More than Ogive Curve?

- Draw and mark the horizontal and vertical axes.
- Take the cumulative frequencies along the y-axis (vertical axis) and the lower-class limits on the x-axis (horizontal axis).
- Against each lower-class limit, plot the cumulative frequencies
- Connect the points with a continuous curve.

Draw a less than and more than Ogive for the following data

Marks	No. of students	Marks below	No. of students	Marks above	No, of students
		Upper limit	Less than c.f.	Lower limit	More than c.f.
0 - 20	2	20	2	0	40
20 - 40	7	40	9	20	38
40 - 60	15	60	24	40	31
60 - 80	9	80	33	60	16
80 - 100	7	100	40	80	7
Total	40				

Marks below	No. of students
Upper limit	Less than c.f.
0	0
20	2
40	9
60	24
80	33
100	40
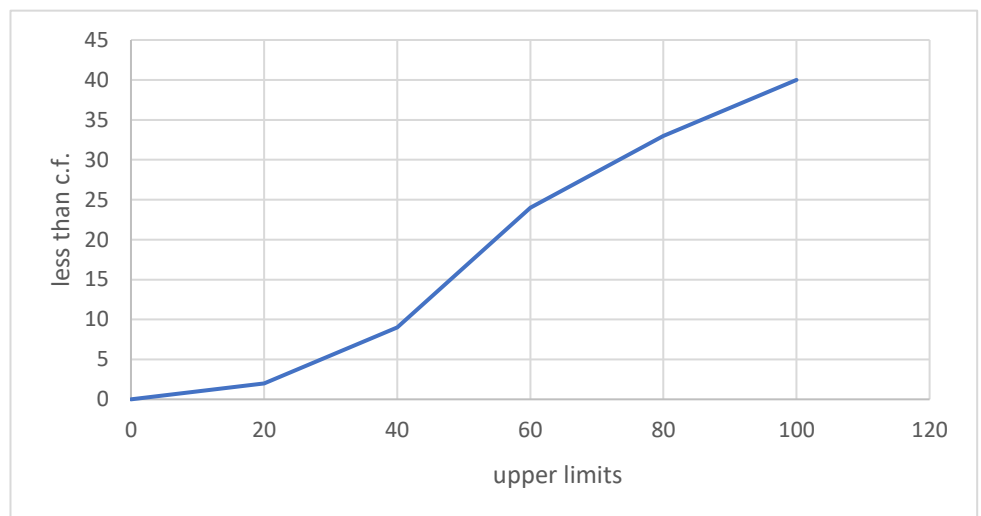
Less than ogive

| Marks above	No, of students
Lower limit	More than c.f.
0	40
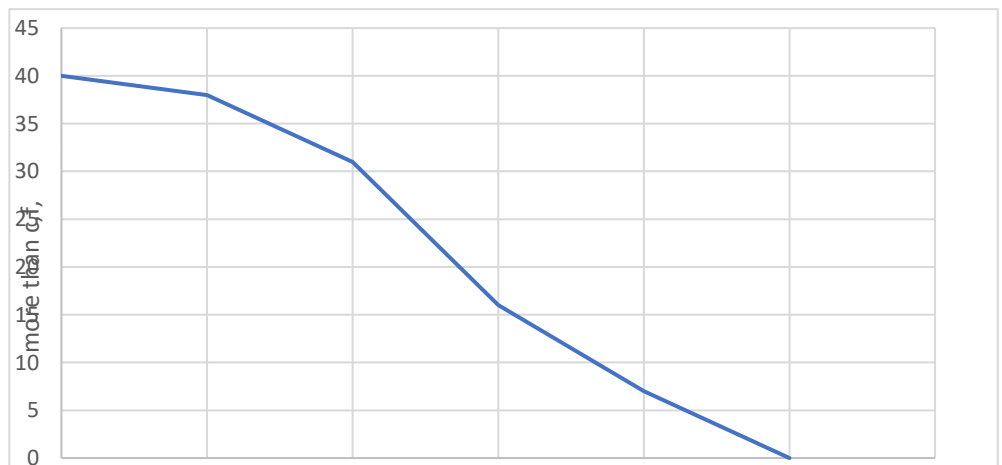
20	38
40	31
60	16
80	7
100	0

More than ogive

Uses of Ogive Curve

Ogive Graph or the cumulative frequency graphs are used to find the median of the given set of data. If both, less than and greater than, cumulative frequency curve is drawn on the same graph, we can easily find the median value. The point in which, both the curve intersects, corresponding to the x-axis, gives the median value. Apart from finding the medians, Ogives are used in computing the percentiles of the data set values.

Calculating Median from Ogive
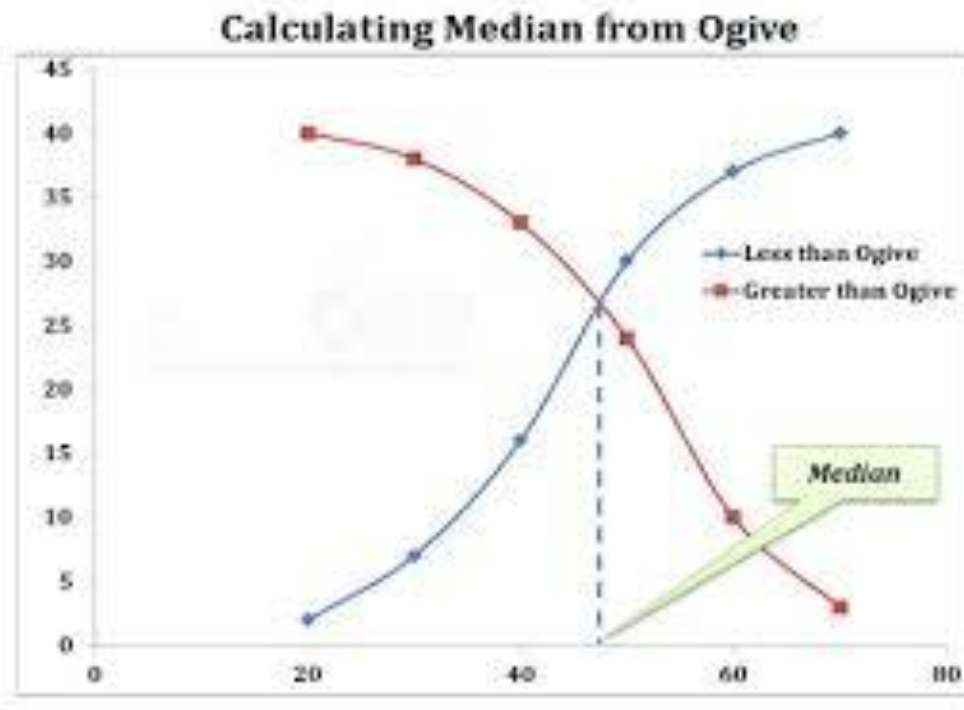
General Principles of Diagrammatic Presentation of Data

A diagrammatic presentation is a simple and effective method of presenting the information that any statistical data contains. Here are some general principles of diagrammatic presentation which can help you make them a more effective tool of understanding the data:

- Write a suitable title on top which conveys the subject matter in a brief and unambiguous manner. If you want to provide more details about the title, then you can mention them in the footnote below the diagram.

- You must construct a diagram in a manner that immediately impacts the viewer. Ensure that you draw it neatly with an appropriate balance between its length and breadth. Further, make sure that the diagram is neither too large nor too small. You can also use different colors or shades to emphasize different aspects of the problem.

- Draw the diagram accurately using proper scales of measurement. You should never compromise accuracy for attractiveness.

- Select the design of the diagram carefully keeping in view the nature of the data and also the objective of the investigation.

- If you use different shades or colors to depict the different characteristics in the diagram, then ensure that you provide an index explaining them.

- If you are using a secondary source, then ensure that you specify the source of data.

- Try to keep your diagram as simple as possible.

Types of Diagrams

There are many types of diagrams which are used for data presentation. Some popular types of diagrams are explained below:

One dimensional diagram-bar diagram

Two dimensional dimensional – circles, squares

Three dimensional diagrams - cubes, spheres

Pictograms and cartograms

Bar diagram

1. They are rectangles placed on a common baseline
2. The width of the bars are equal and height varies according to the value of the variable
3. There should be equal gap in between the bars

There are four types of bar diagrams

Simple bar diagram

Multiple bar diagram

Subdivided or component bar diagram

And percentage subdivided bar diagram

Line Diagram

In a line diagram, you can represent different values using lines of varying lengths. Further, these lines are either horizontal or vertical. Also, there is a uniform gap between successful lines. You can use this when the number of items is very large. Here is an example:

Example 1

The income of 10 workers in a particular week was recorded as given below. Represent the data by a line diagram.

Sr. no. of workers	1	2	3	4	5	6	7	8	9	10
Income (Rs.)	240	350	290	400	420	450	200	300	250	200

The diagram is as follows:
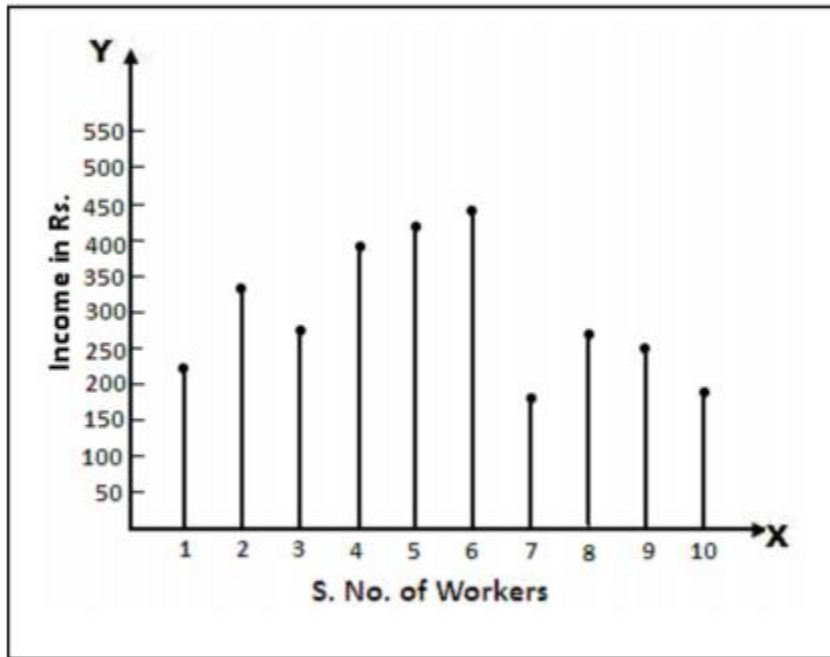
Line Diagram of Income of 10 workers

Simple Bar Diagram

In order to draw a simple bar diagram, you construct horizontal or vertical lines who have heights proportional to the value of the item. You choose an arbitrary width of the bar but keep it constant. Also, ensure that the gaps between the bars are constant. This diagram is suitable to represent individual time-series or a spatial series. Here is an example:
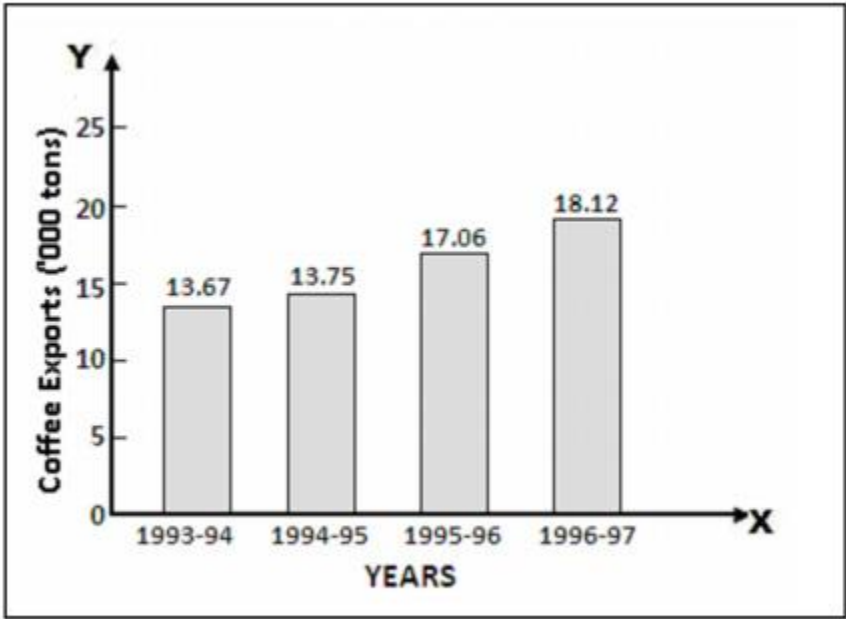
Example 2

Represent the following data using a bar diagram:

Years	1993-94	1994-95	1995-96	1996-97

Coffee Exports ('0000 tonnes)	13.67	13.73	17.06	18.12

The diagram is as follows:

Bar Diagram of Coffee Exports
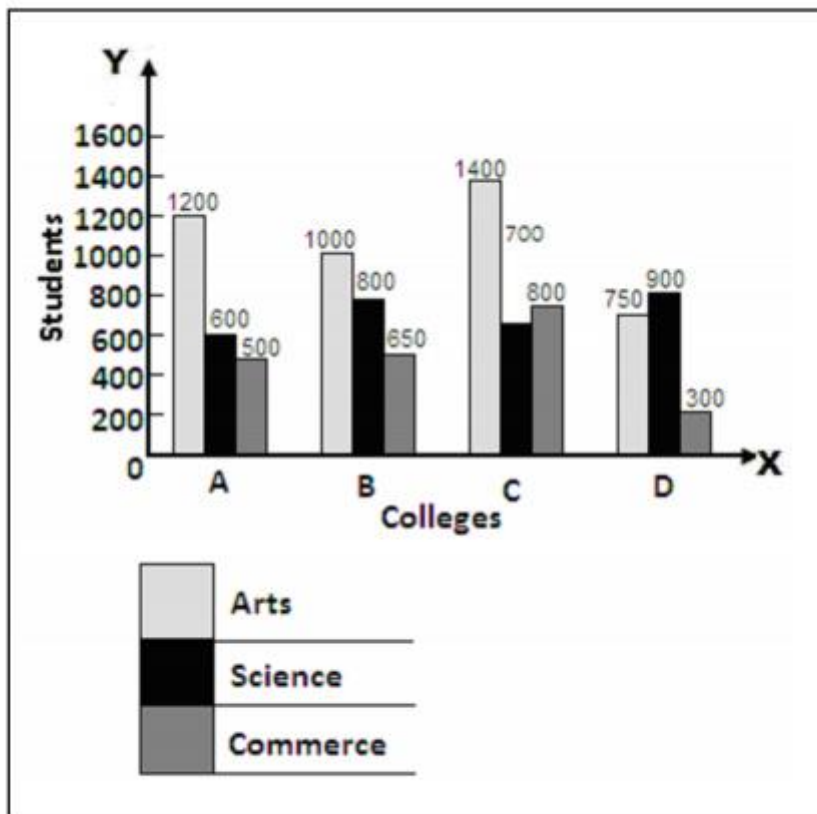
Multiple Bar Diagram

You can use a multiple bar diagram or a compound bar diagram when you want to show a comparison between two or more sets of data. You can draw a set of bars side-by-side, without gaps and separate the sets of bars with a constant gap. Further, you must color or shade different bars in a different manner. Here is an example:

Represent the following data on the faculty-wise distribution of students using a multiple bar diagram:

College	Students

	Arts	Science	Commerce	
A	1200 (1200\2300)x100	600 (600/2300)x100	500 (5000/2300)x100	2300
B	1000	800	650	2450
C	1400	700	800	2900
D	750	900	300	1950

The diagram is as follows:

Multiple Bar Diagram
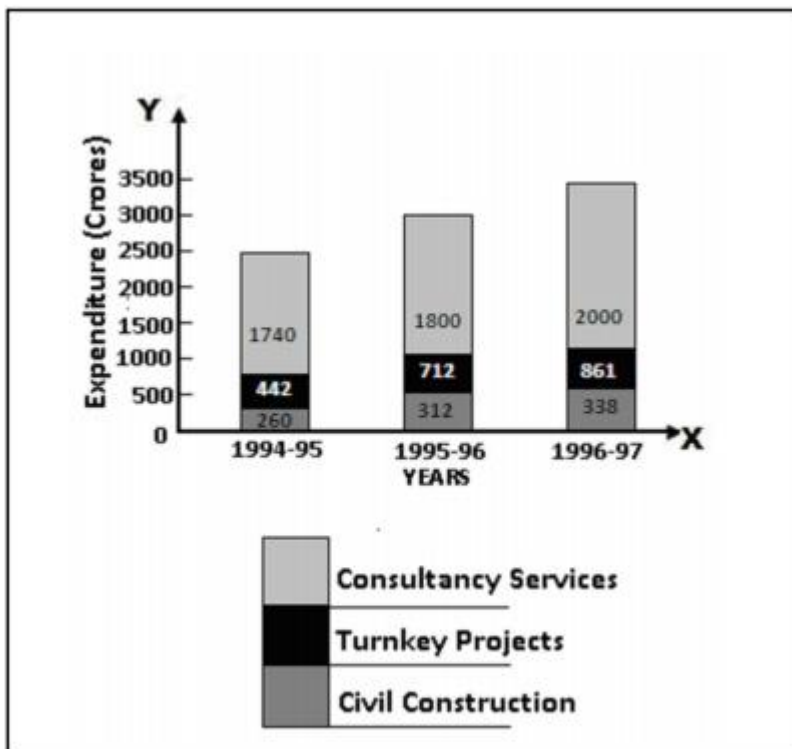
Component or Sub-Divided Bar Diagram

In this diagram, you divide the bar corresponding to each phenomenon into various components. Therefore, the portion that each component occupies denotes its share in the total. You must ensure that the sub-divisions follow the same order and also that you use different colors or shades to distinguish them. You can use this diagram to represent the comparative values of different components of a phenomenon. Here is an example:

The following table gives the value of (A in Crores) of contracts secured from abroad, in respect of Civil Construction, industrial turnkey projects and software consultancy in three financial years. Construct a component bar diagram to denote the share of activity in total export earnings from the three projects.

Years	1994-95	1995-96	1996-97
Civil Construction	260	312	338

Turnkey Projects	442	712	861
Consultancy Services	1740	1800	2000
Total	2442	2824	3199

The diagram is as follows:

Component Bar Diagram

Circular or Pie Chart

A pie chart consists of a circle in which the radii divide the area into sectors. Further, these sectors are proportional to the values of the component items under investigation. Also, the whole circle represents the entire data under investigation.

Steps to draw a Pie Chart step I; convert the individual values to degrees using the formula angle= (individual value/ total)x360

Draw a circle of convenient radius

Divide the circle into different sectors according to the angles calculated using a protractor

Use of Pie Chart

The use of pie charts is quite popular as the circle provides a visual concept of the whole. Pie charts are simple to use and hence are one of the most commonly used charts. However, the pie charts are sparingly used only for the following reasons:

- They are the best chart for displaying statistical information when the number of components is not more than 6. In the case of more components, the chart becomes too complex to understand.

- Pie charts are not useful when the values of the components are similar. This is because in the case of similarly sized sectors the viewer can find it difficult to differentiate between the slice sizes.

Here is an example:

Represent the following data, on India's exports (Rs. in Crores) by regions from April to February 1997.

Region	Europe	Asia	America	Africa
Exports	32699	42516	23495	5133

From the table we have,

Total exports = 32699 + 42516 + 23495 + 5133 = Rs. 103, 843 crores

Europe = $32699 \times \frac{360}{103843} = 32699 \times \frac{360}{103843} = 113°$
Asia = $42516 \times \frac{360}{103843} = 42516 \times \frac{360}{103843} = 147°$
America = $23495 \times \frac{360}{103843} = 23495 \times \frac{360}{103843} = 82°$
Africa = $5133 \times \frac{360}{103843} = 5133 \times \frac{360}{103843} = 18°$

Pie-Chart