Data Handling

Frequency Distribution Table and Terminology Related to It

Observe the information given in the following cases.

(1) The weights (in kg) of 15 students in the same class are as follows:

45, 50, 48, 47, 58, 52, 49, 54, 48, 51, 46, 57, 56, 50, 44

(2) Minimum temperature (in °C) of a city for each day of a week is given as follows:

1.5, 2, 0, 2.5, 3.5, 1, 1

(3) Runs scored by 8 players of a cricket team in a match are as follows:

Player	Runs
Harry	21
Venkat	16
Robin	74
Dinesh	09
Vikram	81
Laxmipati	42
Jairaj	36
Ysuf	27

It can be seen that in each case, we have some numeric information.

Numeric information collected for a particular purpose is known as raw data and each number involved in this raw data is known as score.

For example, in case (1), weight of each student is a score.

Similarly, in case (2), temperature of each day and in case (3), runs scored by each player are scores.

Raw data is found in unorganized form and in many real life situations, we have to deal with it.

Data can be of two types such as **primary data** and **secondary data**.

Primary data: When the data is collected by an investigator according to a plan for a particular objective then the collected data is called primary data.

For example, if a person collects information about the people using a particular mobile phone network in a particular locality, then the data collected by the person will be called primary data.

Secondary data: When the required data is taken from the data already collected by other private agency, government agency, an organization or any other party then the data is called secondary data.

For example, if an organization extracts the data from the records of census published by the government, then the data is called secondary data.

To draw meaningful inference, we organize the data into systematic pattern in the form of frequency distribution table.

Let us understand this with the help of an example.

Heights (in cm) of 30 students of a class are given as follows:

152, 160, 154, 151, 158, 165, 152, 160, 160, 152, 152, 161, 158, 160, 152, 165, 165, 155, 158, 158, 154, 158, 160, 161, 158, 161, 158, 155, 161, 160

Now, it can be seen that the lowest score is 151 and the highest score is 165.

The difference between the highest observation and lowest observation in a given data set is called the **range**. Range of the above data is 165 – 151 = 14.

It can be observed that few scores occur more than once in the data.

Number of times by which a score occurs in the data is called the frequency of that score.

Score 151 occurs just once, so its frequency is 1.

Similarly,

Score 152 occurs five times, so its frequency is 5.

Score 154 occurs two times, so its frequency is 2.

Score 155 occurs two times, so its frequency is 2.

Score 158 occurs seven times, so its frequency is 7.

Score 160 occurs six times, so its frequency is 6.

Score 161 occurs four times, so its frequency is 4.

Score 165 occurs three times, so its frequency is 3.

The sum of all frequencies or total frequency is 30 which gives us the total number of scores in the data. Total frequency is denoted by N.

Now, we can arrange these scores in a table according to their respective frequencies and such a table is known as frequency distribution table.

Frequency distribution table for the given data is as follows:

Height	Tally Mark	Frequency
151	𝍸	1
152	𝍸𝍸	5
154	𝍸	2
155	𝍸	2
158	𝍸𝍸𝍸	7
160	𝍸𝍸𝍸	6
161	𝍸𝍸	4
165	𝍸	3
	Total (N)	30

The bars in the second column are known as tally marks which are used to represent the numbers.

In tally marks representation, 1 is represented by one bar i.e., 𝍸, 2 is represented by the group of two bars i.e., 𝍸 and 5 is represented by 𝍸𝍸 (four vertical bars are intersected by one bar diagonally). Similarly, each number is represented by putting that many of bars in a group.

We can make frequency distribution table by arranging the data in small groups or intervals also.

To understand the concept with the help of an example, look at the following video.

The table obtained in the video can be represented using tally marks as follows:

Group	Tally Mark	Frequency
0 – 10	||	2
10 – 20	𝍸||||	14
20 – 30	𝍸||||	14
30 – 40	𝍸	10
40 – 50	𝍸|||	8

Class mark is the arithmetic mean of the upper and lower limits of a class. It is also known as the mid value of the class interval.

Therefore,

$$\text{Class mark} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

Let us now go through the given examples to understand this concept better.

Example 1:

The marks obtained by 10 students out of 100 are given below:
55, 79, 68, 85, 96, 48, 39, 67, 80, 72

Find the range of marks.

Solution:

From the given marks, we observe that the highest mark is 96 and the lowest mark is 39.

∴ Range of the marks = Highest mark – Lowest mark = 96 – 39 = 57

Example 2:

The number of runs scored by a cricket player in 25 innings are given below:

64, 94, 26, 35, 46, 49, 107, 56, 3, 36, 41, 73, 8, 63, 128, 17, 33, 68, 5, 11, 23, 77, 28, 85, 117

Prepare a frequency distribution table, taking the size of the class interval as 20, and answer the following questions:

(i)What are the class intervals of highest and lowest frequency.

(ii)What does the frequency 2 corresponding to the class interval (100 – 120) indicates?

(iii)What is the class mark of the class interval (100 – 120).

(iv)What is the range of the runs scored by the player?

Solution:

The frequency distribution for the given data is as follows:

Class interval (Runs scored)	Tally marks	Frequency
0 – 20	𝍸	5
20 – 40	𝍸 |	6
40 – 60	||||	4
60 – 80	||||	4
80 – 100	|||	3
100 – 120	||	2
120 – 140	|	1

(i) Class interval with the highest frequency is 20 – 40 whereas the class interval with the lowest frequency is 120 – 140.

(ii) The frequency 2 in the class interval 100 – 120 indicates that the player has scored runs in the range 100 to 120 twice in 25 innings.

(iii)Class mark of the interval 100 – 120 $= \frac{100+120}{2} = 110$

(iv) Range of the runs scored = Highest run – Lowest run = 128 – 3 = 125

Example 3:

Observe the given frequency distribution table and answer the following questions:

Salary per month(Rupees in thousands)	Number of employees(Frequency)
15	20
20	35
25	30
30	25

35	20
40	20
45	18
50	12

I. How many employees are involved in the survey?

II. How many employees earn Rs 25,000 per month?

III. What is the difference between the number of employees getting the highest salary and the number of employees getting the lowest salary?

IV. How many employees earn more than Rs 35,000 per month?

V. What is the monthly salary that is being paid to the maximum number of employees?

Solution:

I. Number of employees involved in the survey = Sum of all frequencies

∴ Number of employees involved in the survey = 20 + 35 + 30 + 25 + 20 + 20 + 18 + 12

⇒ Number of employees involved in the survey = 180

II. 30 employees earn Rs 25,000 per month.

III. Number of employees getting the highest salary = 12

Number of employees getting the lowest salary = 20

∴ Required difference = 20 – 12

⇒ Required difference = 8

IV. Number of employees earning more than Rs 35,000 per month = Sum of number of employees earning Rs 40,000, Rs 45,000 and Rs 50,000 per month

∴ Number of employees earning more than Rs 35,000 per month = 20 + 18 + 12

⇒ Number of employees earning more than Rs 35,000 per month = 50

V. Highest frequency in the table is 35 which represents the maximum number of employees in any salary group. Also, each employee in this group earns Rs 20,000 per month.

Thus, Rs 20,000 is the monthly salary that is being paid to the maximum number of employees.

Example 4:

Observe the given frequency distribution table and then answer the following questions.

Class interval (height in cm)	Frequency (number of students)
0 – 12	2
12 – 24	3
24 – 36	5
36 – 48	10
48 – 60	3

1. What is the size of the class intervals?

2. Which class interval has the highest frequency?

3. Which two classes have the same frequency?

4. How many students have height less than 36 cm?

5. What is the lower limit of the class interval 24 – 36?
6. What is the class mark of the class interval 48 – 60?

Solution:

1. The difference between the upper and lower class limits for each class interval is 12. Therefore, the class size is 12.

2. The class 36 – 48 has the highest frequency. 10 students height belong to this category.

3. The classes 12 – 24 and 48 – 60 have the same frequency.

4. The number of students having height less than 36 cm is 2 + 3 + 5 = 10.

5. The lower limit of the class interval 24 – 36 is 24.

6. $$\text{Class mark} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

$$\Rightarrow \text{Class mark} = \frac{48 + 60}{2}$$

$$\Rightarrow \text{Class mark} = \frac{108}{2}$$

$$\Rightarrow \text{Class mark} = 54$$

Representation Of Data Using Tally Marks

Data is defined as a collection of numbers which give the required information. For example, marks scored by the students in a class, number of members in a family, number of books sold etc.

Data are of two types:

(i) **Primary data:** It is the data collected by the person directly for a specific purpose without referring any source. Primary data is collected through surveys, local sources etc.

(ii) **Secondary data:** It is the data collected through other sources like research organizations, financial institutions etc.

The original form of data is called **raw data**. But when the data is arranged in ascending or descending order, it is referred to as **array**.

Example: The marks obtained by 10 students in a test out of 30 are as follows:

10, 15, 25, 22, 12, 18, 28, 29, 17, 18

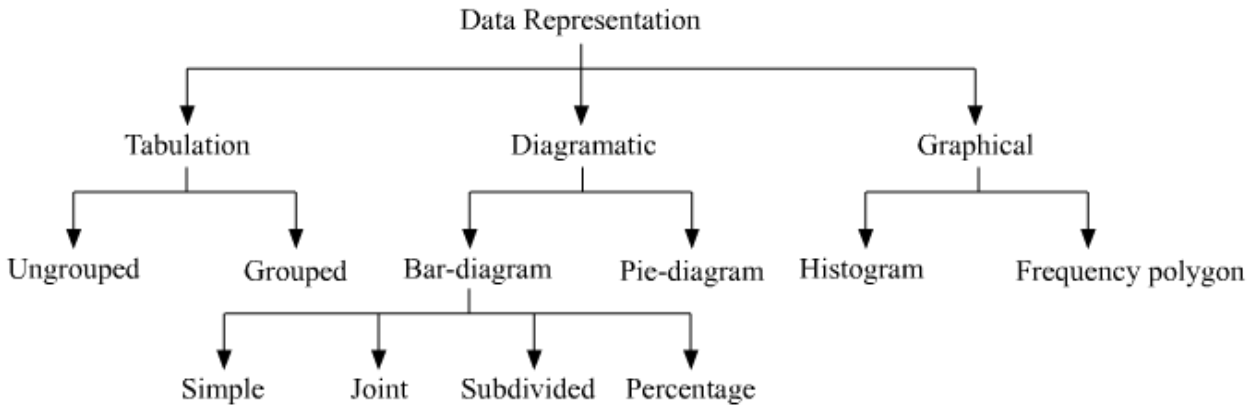
This data is in the form of raw data or ungrouped data.

Now, after arranging the given data in ascending order we get 10, 12, 15, 17, 18, 18, 22, 25, 28, 29 and on arranging them in descending order we get 29, 28, 25, 22, 18, 18, 17, 15, 12, 10.

Such arrangement of the data in ascending or descending order is called an array.

There are different ways of representing the data.

Below given chart explain the same.



Let us now discuss some examples based on representation of a data through tally marks.

Example 1:

A survey was conducted in a village of 80 people. The blood groups of 80 people are as follows.

A, AB, A, B, O, AB, A, AB, O, B, A, O, O, B, A, A, B, AB, O, AB, O, B, A, O, B, O, A, O, A, AB, A, B, AB, O, B, O, O, B, AB, O, A, A, B, AB, AB, O, B, O, O, B,

O, O, B, A, A, O, O, B, O, O, AB, O, B, O, AB, O, A, O, O, B, A, AB, AB, A, B, AB, O, B, O, O.

Represent this data using tally marks.

Solution:

The given data is represented by using tally marks as follows.

Blood Group	Tally Marks	No. of people
A	𝍸 𝍸 𝍸 ||	17
B	𝍸 𝍸 𝍸 |||	18
AB	𝍸 𝍸 𝍸	15
O	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
Total		80

Example 2:

Neha threw a dice 150 times and noted the number appearing each time. This data is represented by the following table containing tally marks. Fill in the blanks in this table.

Observation	Tally Marks	No. of Observations
1		28
2	𝍸	
3		42
4	𝍸𝍸	
5		13
6		
Total		150

Solution:

We easily fill the first four blanks by counting the number of tallies.

Observation	Tally Marks	No. of Observations
1	𝍸𝍸	28
2	𝍸	15
3	𝍸𝍸𝍸	42
4	𝍸𝍸	22
5	𝍸	13
6		
Total		150

Now, let us find the last two blanks.

Total number of observations = 150

Number of observations excluding the 6^{th} observation = 28 + 15 + 42 + 22 + 13 = 120

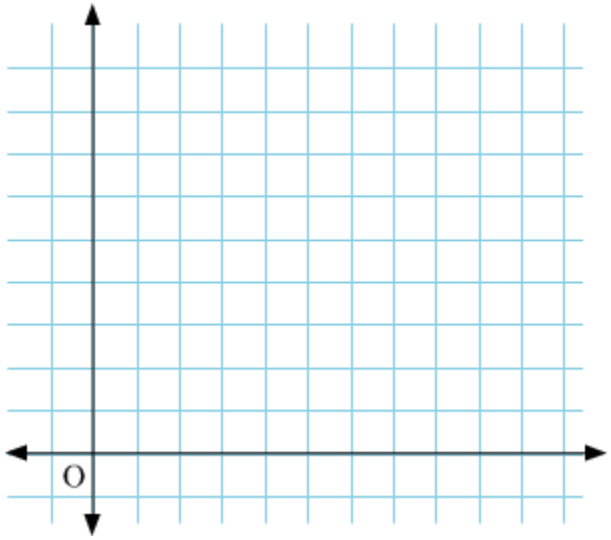
Hence, number of observations for the 6^{th} observation 6 = 150 – 120 = 30

Observation	Tally Marks	No. of Observations
1	𝍸𝍸	28
2	𝍸	15
3	𝍸𝍸𝍸	42

4	𝍸𝍸𝍸𝍸𝍸𝍸||	22
5	𝍸𝍸𝍸|||	13
6	𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸𝍸	30
Total		150

Construction of Bar Graphs

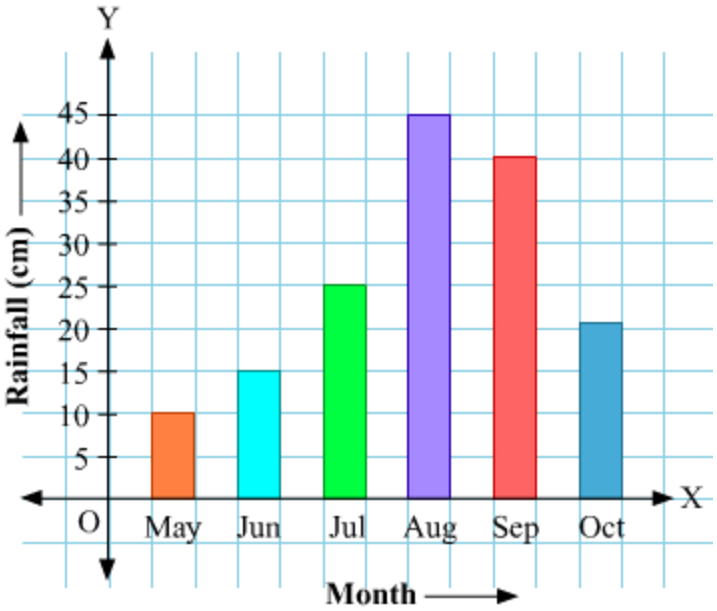
Look at the following figure.



This is a graph paper. A **graph paper** has various vertical and horizontal lines intersecting each other. Also, distance between every two lines is always equal. Thus, we get many small squares on the graph paper.

Graph paper is used to draw bar diagrams. A bar diagram drawn on such a graph paper is known as **bar graph**.

Look at the following graph given by meteorological department.



The above graph shows the rainfall in a place from the month of May to October of a year.

Here, the rainfall is shown by vertical bars of uniform width and with equal spaces between them. This type of representation of a data is known as **bar graph.**

Let us now look at one more example to understand this concept better.

Example:

The following information represents the amount of money earned by a trader in different months.

Month	Amount of money earned (Rs)
January	**9000**
February	**5000**
March	**7000**

April	**11000**
May	**10500**
June	**7000**

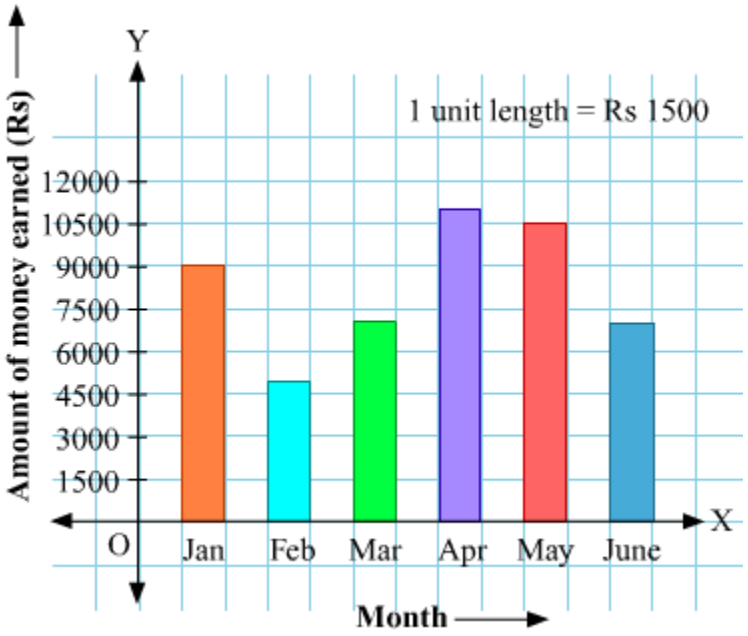
Represent this information with the help of a bar graph.

Solution:

We will represent the months on the horizontal line and the amount of money earned on the vertical line.

Here, we take the scale as 1 unit = Rs 1500

The bar graph of the given information is as follows:



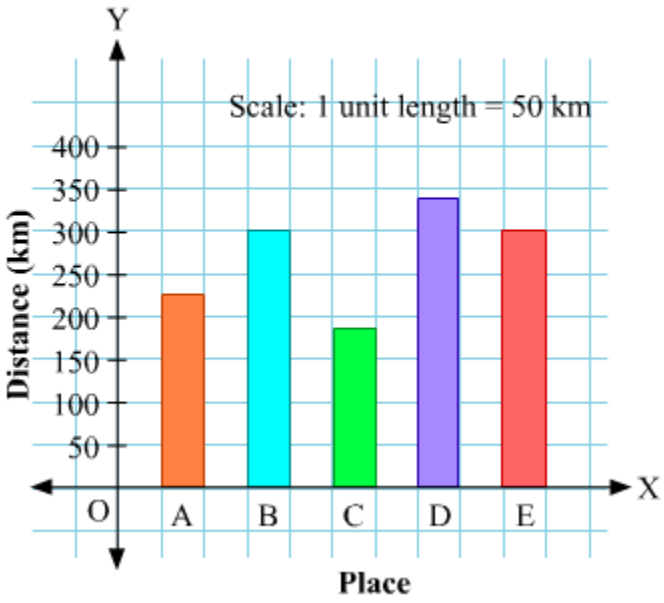
Interpretation of Bar Graphs

Bar graphs are one of the most common tools used to represent data. Therefore, to read bar graphs and extract the required information from them is an important part of data analysis.

Let us now look at one more example to understand this concept better.

Example:

The following bar graph shows the distance of a mall from five different places. Read the graph and answer the questions asked below.



1. **Which place is the farthest from the mall?**

2. **Which place is the nearest to the mall?**

3. **How far is place B from the mall?**

4. **The distance of which two places is the same from the mall?**

Solution:

1. The bar corresponding to place D is the highest. Thus, place D is the farthest from the mall.

2. The bar corresponding to place C is the shortest. Thus, place C is the nearest to the mall.

3. Place B is at a distance of 300 km from the mall.

4. The height of the bars corresponding to the places B and E are the same. Thus, the places B and E are at the same distance from the mall.

Construction Of Double Bar Graphs

Many times in daily life, we come across such situations where we have certain data which we have to compare. In such conditions, it is always better to compare the data graphically.

We know how to draw a bar graph with the given data. Now, one method can be to draw the graphs of each data separately and then compare it. But this is not a fruitful method as one has to look at both the graphs separately and the comparison is also not easy.

Thus, to solve such problems, we will now study the concept of double bar graphs. In double bar graphs, we draw the graphs for both the data on the same axis and then comparison becomes easier.

Let us solve another example.

Example 1:

The following table shows the number of boys and girls in a class who like different kinds of fruits.

Name of fruit	Number of boys	Number of girls
Apple	25	15
Banana	35	25
Mango	12	25
Pineapple	15	15
Orange	20	25

Draw a double bar graph representing the given data.

Solution:

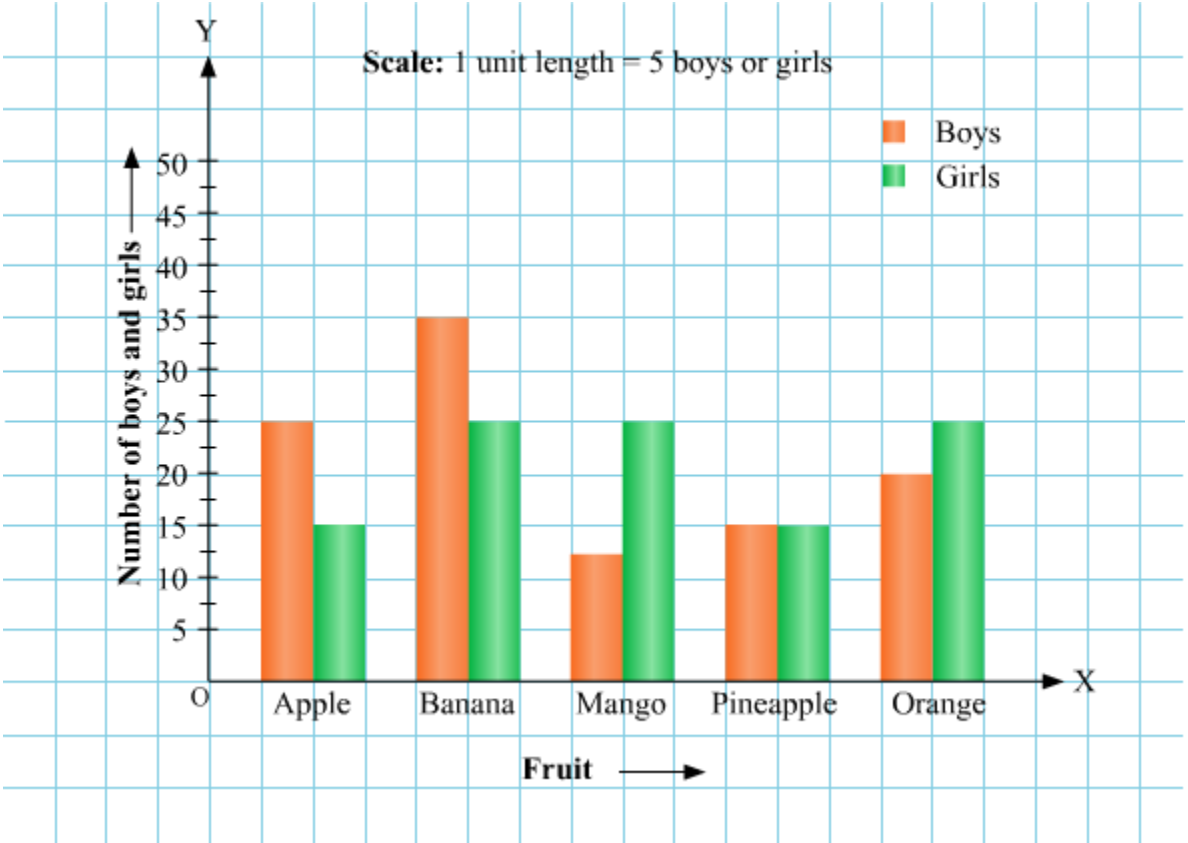
In this case, we take the scale as

1 unit = 5 boys or girls.

The number of boys and girls are represented on the *y*-axis and the fruits are represented on the *x*-axis.

The green coloured bars represent the number of girls and the orange coloured bar represents the number of boys.

The double bar graph can be drawn as follows.

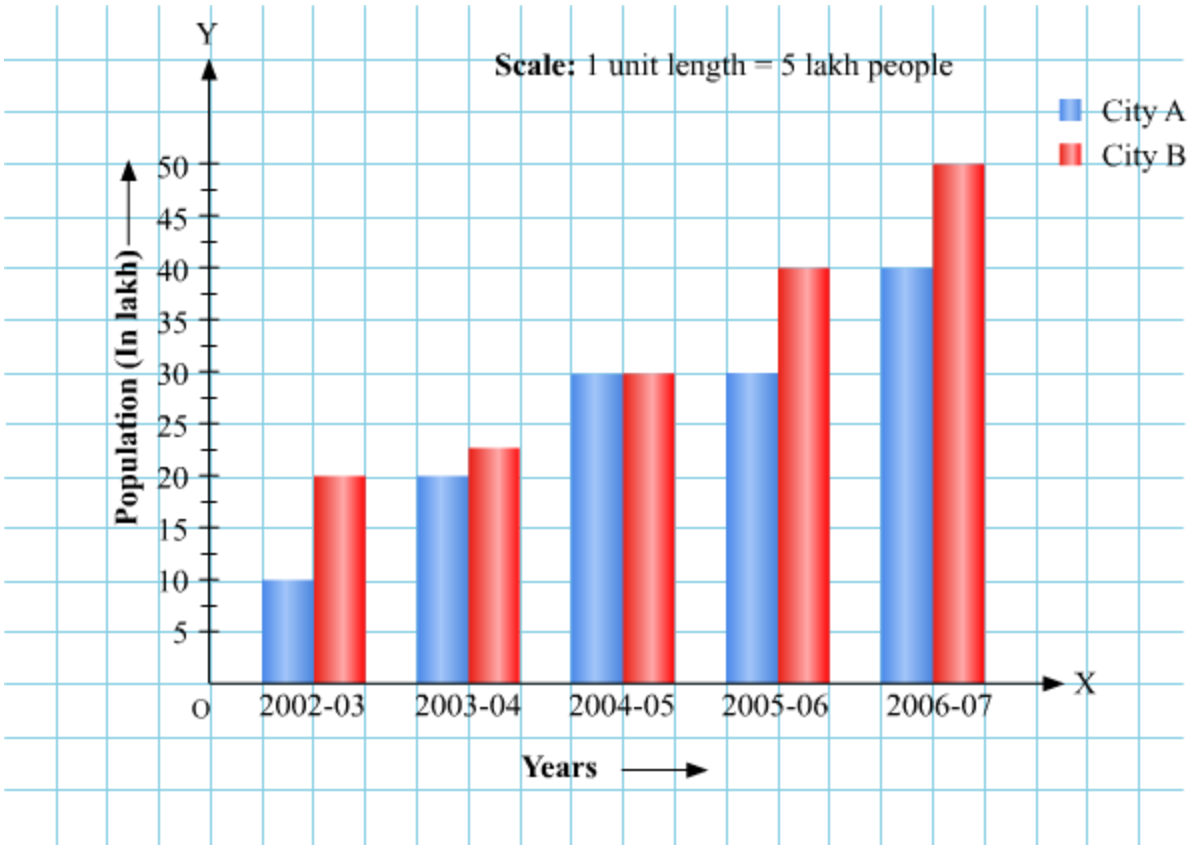


Interpretation of Double Bar Graphs

Let us now look at some more examples.

Example 1:

Look at the double bar graph given below and answer the following questions.



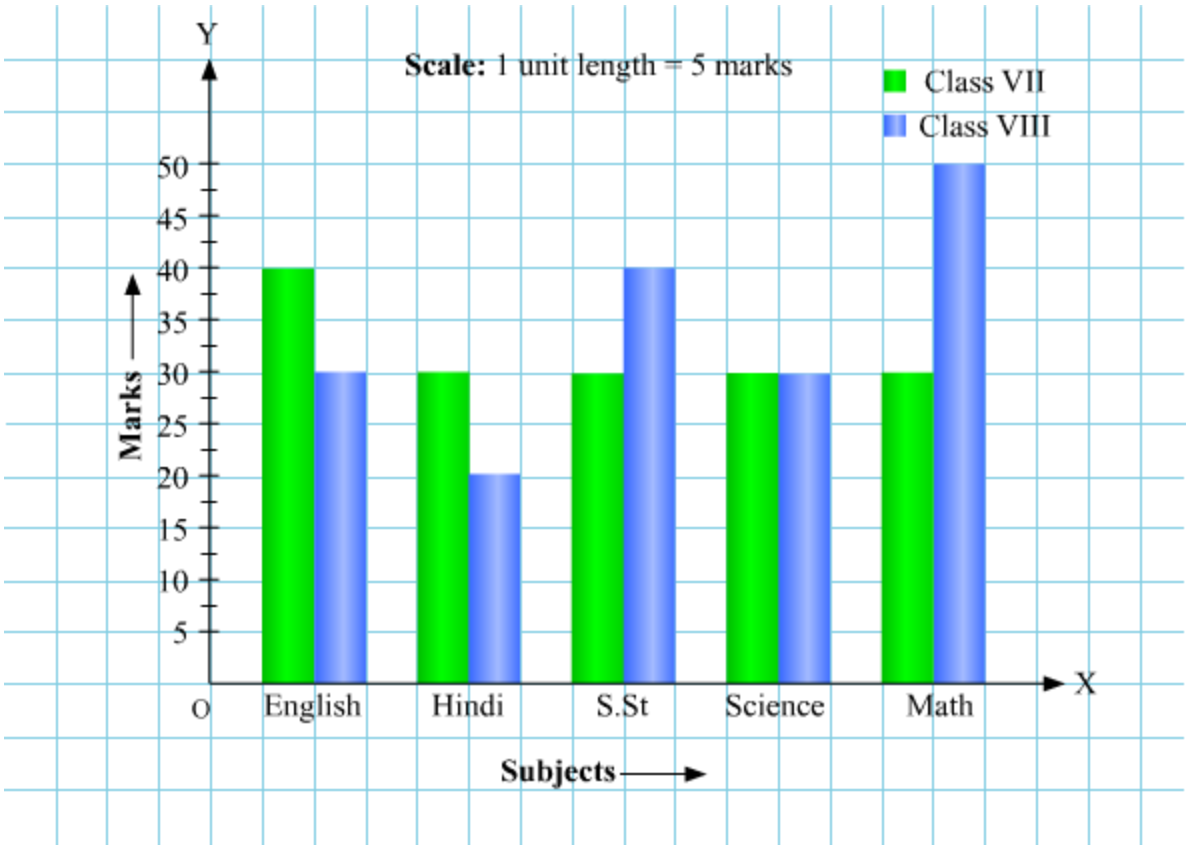
1. **What information is represented in the bar graph?**
2. **In which year was the population of city B the largest?**
3. **In which year was the population of city A same as that of city B?**
4. **Can you estimate the population of city A in the year 2003-04 and also in the year 2005-06?**

Solution:

1. The bar graph represents the population of two cities A and B for five consecutive years from 2002-03 to 2006-07.
2. Since the bar representing city B is biggest in the year 2006-07, the population of city B was the largest in the year 2006-07.
3. The population of the two cities was same in the year 2004-05, since the length of both the bars are the same in this year.
4. As seen from the graph in the year 2003-04, the population of city A was 20 lakhs and in the year 2005-06, the population of city A was 30 lakhs.

Example 2:

The following graph shows the marks obtained by Ritesh in class VII and VIII in five subjects.



Observe the graph and answer the following questions.

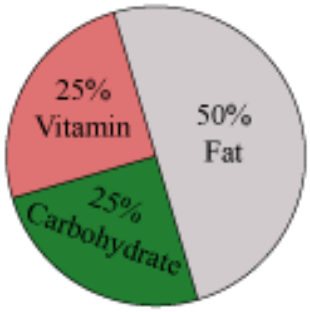
1. **In which subject was the performance of Ritesh same in both classes?**
2. **In which subject did the performance deteriorate?**
3. **In which subject was there the maximum difference in marks?**
4. **How many marks did Ritesh score in English in both the classes?**

Solution:

1. The performance of Ritesh in Science was same in both the classes, since the length of the bar is same for both classes.
2. The performance deteriorated in Hindi and English both, since the blue bar is smaller than the green bar in these two subjects.
3. The maximum difference of marks can be seen in Maths as we can see that the difference between the heights of the blue and green bars is maximum for Maths.
4. Ritesh got 40 marks in English in class VII and 30 marks in class VIII.

Construction of Circle Graphs

Sometimes, the data is represented using circles. For example, the circle given below shows various nutrients present in a chocolate.



The representation of data in this form is called **pie-chart** or **sector graph** or **circle graph.**

A circle graph shows the relationship between a whole circle and its parts. The whole circle is divided into sectors and the size of each sector is proportional to the information it represents.

Now consider the following data that represents the numbers of people who watch channels 1, 2, and 3.

Channels	Percentage of people preferring the channel
1	30%
2	25%
3	45%

Can we draw a pie-chart of the given data?

Let us now go through the given video and see how to draw the pie-chart.

Let us solve some examples now.

Example 1:

The choice of food for a group of people is given below.

Favourite Food	Number of people
North Indian	50
South Indian	40
Others	30

Total	120

Draw a pie-chart for the given data.

Solution:

Firstly, we will find the central angle of each sector. Here, total number of people = 120. The central angle has been calculated in the following table.

Favourite Food	Number of people	In Fraction	Central Angle
North Indian	50	$\frac{50}{120}=\frac{5}{12}$	$\frac{5}{12}\times 360^\circ = 150^\circ$
South Indian	40	$\frac{40}{120}=\frac{1}{3}$	$\frac{1}{3}\times 360^\circ = 120^\circ$
Others	30	$\frac{30}{120}=\frac{1}{4}$	$\frac{1}{4}\times 360^\circ = 90^\circ$

Draw a circle of any radius. Then, draw the angle of sector for the north Indian food, which is 150°. Use the protractor to draw the angle of 150°. Then continue making the remaining angles (120° and 90°). The pie chart has been shown as follows.

Example 2:

Data regarding the maximum leaves taken in a month by the employees of a company is given in the following table.

Number of leaves	Number of employees

0 – 2 2 – 4 4 – 6 6 – 8	120 45 25 10
Total	200

Draw a pie-chart for the given data.

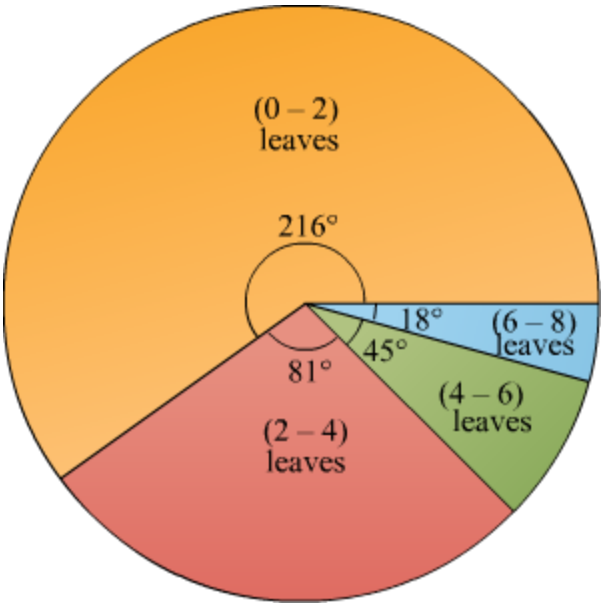
Solution:

Firstly, we will find the central angle of each sector. Here, total number of employees = 200. The central angle has been calculated in the following table.

Number of leaves	Number of employees	In Fraction	Central Angle
0 – 2	120	$\frac{120}{200} = \frac{3}{5}$	$\frac{3}{5} \times 360° = 216°$
2 – 4	45	$\frac{45}{200} = \frac{9}{40}$	$\frac{9}{40} \times 360° = 81°$
4 – 6	25	$\frac{25}{200} = \frac{1}{8}$	$\frac{1}{8} \times 360° = 45°$
6 – 8	10	$\frac{10}{200} = \frac{1}{20}$	$\frac{1}{20} \times 360° = 18°$

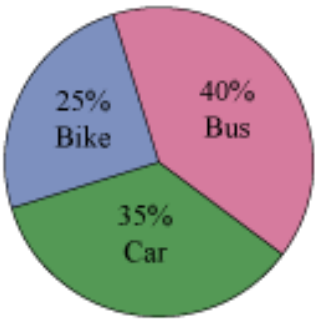
Draw a circle of any radius. Then, draw the angle for each sector.

The obtained pie-chart is as follows:



Interpretation of Circle Graphs

The following pie-chart represents the percentage of number of students who come to school by bus, car, or bike. The number of students studying in the school is 5000.



Can we find out how many students come by bike?

Let us see.

In the graph, the sector representing the number of students who come by bike is given to be 25%.

The total numbers of students are 5000.

Thus, number of students who come by bike = 25% of 5000 = $\frac{25}{100} \times 5000$ = 25 × 50 = 1250

Also find out how many students come by car and bus.

In the graph, the sector representing the number of students who come by car is given to be 35%.

Number of students who come by car = 35% of 5000 = $\frac{35}{100} \times 5000$ = 35 × 50 = 1750

In the graph, the sector representing the number of students who come by bus is given to be 40%.

Number of students who come by bus = 40% of 5000 = $\frac{40}{100} \times 5000$ = 40 × 50 = 2000

Which is the most common mode of transport?

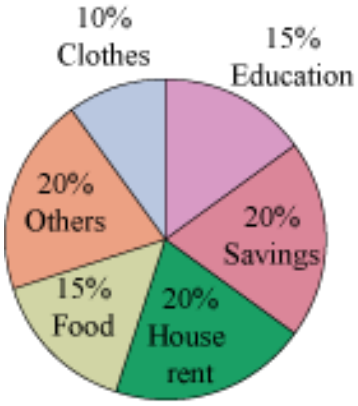
Since the number of students coming by the bus is highest, bus is the most common mode of transport.

In this way, we can interpret the information given in a pie-chart or a circle graph.

Let us now look at one more example.

Example:

Mr. Nair's expenditure on various items and his savings for a particular month has been represented in the following pie-chart.



Observe the given pie-chart and answer the following questions.

1. **On which of the represented items, the expenditure is least?**
2. **If the monthly savings of Mr Nair is Rs 4000, then what is the monthly expenditure on food?**

Solution:

1. The expenditure is least on clothes. It is 10%.
2. It is shown in the circle graph that his savings is 20%. 20% represents Rs 4000.

∴ 15% represents Rs $\frac{4000}{20} \times 15 = \text{Rs } 3000$

Thus, the monthly expenditure on food is Rs 3000.

Construction of Histograms when Class Size is Same

The frequency distribution table of the marks of 26 students in a particular subject is as follows.

Class interval (marks of students)	Frequency (number of students)
0 – 10	4
10 – 20	2
20 – 30	10
30 – 40	8
40 – 50	2

Can we represent this data graphically?

This data can be represented in the form of a histogram.

To understand the concept of histograms and the method used to draw a histogram, let's look at this video.

Let us now look at an example to understand this concept better.

Example 1:

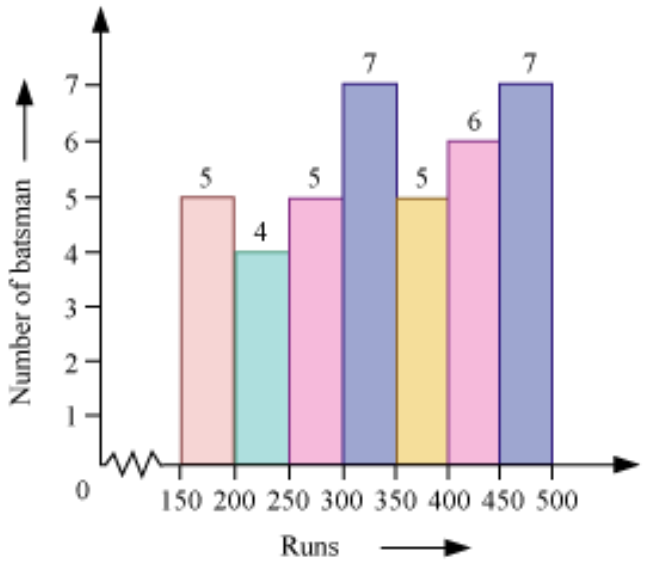
The given tally table represents the total runs scored by 39 batsmen in 10 different test matches.

Runs	Tally marks	Frequency (Number of batsmen)
150 – 200		5
200 – 250		4
250 – 300		5
300 – 350		7
350 – 400		5
400 – 450		6
450 – 500		7

Draw a histogram for the above given distribution table.

Solution:

In order to draw the histogram of the given frequency distribution table, we represent the runs on the horizontal axis and the number of batsmen on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.



Here, we will use a broken line ( ) to indicate that the values between 0 – 150 are not represented.

Example 2:

The given table represents the data related to intelligent quotient (IQ) of the students of a class.

IQ	Number of students
61 – 70	**4**
71 – 80	**3**
81 – 90	**5**
91 – 100	**8**
101 – 110	**15**
111 – 120	**12**
121 – 130	**13**
Total	**60**

Draw a histogram for the above given distribution table.

Solution:

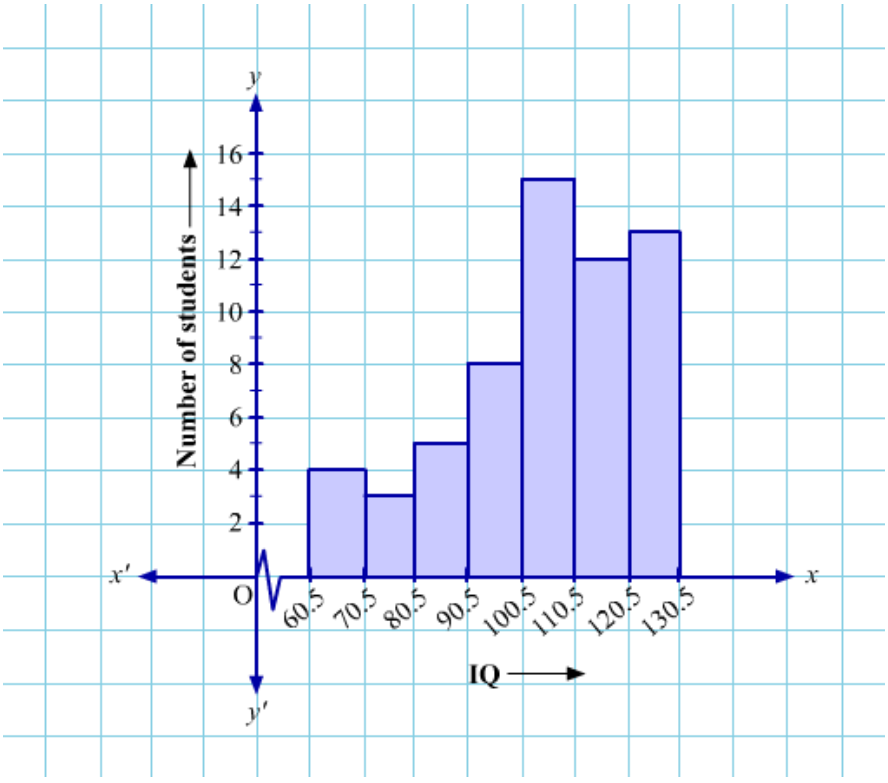
In the given table the class intervals are of inclusive type, so we need to make them of exclusive type. Here, the difference between the upper limit of a class and lower limit of

next class is 1. So, we need to subtract half of this i.e., 0.5 from lower limit and add 0.5 to upper limit of each class. Thus, we will get extended classes according to which, we can draw the histogram.

The modified table consisting extended classes is as follows:

Original Class	Extended Class	Number of students (Frequency)
61 – 70	**60.5 – 70.5**	**4**
71 – 80	**70.5 – 80.5**	**3**
81 – 90	**80.5 – 90.5**	**5**
91 – 100	**90.5 – 100.5**	**8**
101 – 110	**100.5 – 110.5**	**15**
111 – 120	**110.5 – 120.5**	**12**
121 – 130	**120.5 – 130.5**	**13**
Total		**60**

In order to draw the histogram of this frequency distribution table, we represent the IQ on the horizontal axis and the number of students on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.



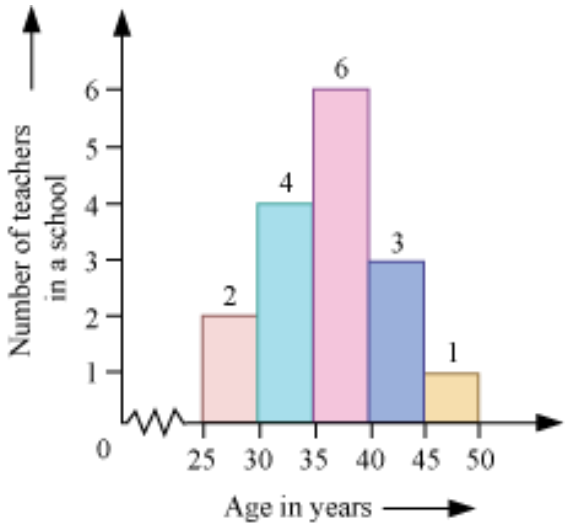
Here, we will use a broken line ( ) to indicate that the values between 0 – 60.5 are not represented.

Interpretation Of Histograms

Let us now look at one more example.

Example 1:

The following histogram shows the age of the teachers in a school.



Observe the histogram and answer the following questions.

1. **How many teachers age is 40 years or more but less than 45 years?**

2. **How many teachers are of age less than 40 years?**

3. **Which group contains the least number of teachers?**

4. **The age of maximum number of teachers lies in which group?**

Solution:

1. In order to find the number of teachers of the age 40 years or more but less than 45, we have to find the number of teachers in the age group 40 − 45. The number of teachers in the age group (40 − 45) is 3.

2. In order to find the number of teachers of age less than 40 years, we take into account the number of teachers of the age groups 25 − 30, 30 − 35, and 35 − 40. Thus, the number of teachers of age less than 40 years is 2 + 4 +6 = 12.

3. The age group 45 − 50 contains the least number of teachers.
4. The maximum numbers of teachers are in the age group 35 − 40.