Chapter 3

Data Handling



3.1 Representative Values

You might be aware of the term *average* and would have come across statements involving the term 'average' in your day-to-day life:

- Isha spends on an average of about 5 hours daily for her studies.
- The average temperature at this time of the year is about 40 degree celsius.
- The average age of pupils in my class is 12 years.
- The average attendance of students in a school during its final examination was 98 per cent.

Many more of such statements could be there. Think about the statements given above.

Do you think that the child in the first statement studies exactly for 5 hours daily?

Or, is the temperature of the given place during that particular time always 40 degrees?

Or, is the age of each pupil in that class 12 years? Obviously not.

Then what do these statements tell you?

By average we understand that Isha, usually, studies for 5 hours. On some days, she may study for less number of hours and on the other days she may study longer.

Similarly, the average temperature of 40 degree celsius, means that, very often, the temperature at this time of the year is around 40 degree celsius. Sometimes, it may be less than 40 degree celsius and at other times, it may be more than 40°C.

Thus, we realise that average is a number that represents or shows the central tendency of a group of observations or data. Since average lies between the highest and the lowest value of the given data so, we say average is a measure of the central tendency of the group of data. Different forms of data need different forms of representative or central value to describe it. One of these representative values is the "**Arithmetic mean**". You will learn about the other representative values in the later part of the chapter.

3.2 Arithmetic Mean

The most common representative value of a group of data is the **arithmetic mean** or the **mean**. To understand this in a better way, let us look at the following example:

Two vessels contain 20 litres and 60 litres of milk respectively. What is the amount that each vessel would have, if both share the milk equally? When we ask this question we are seeking the arithmetic mean.

In the above case, the average or the arithmetic mean would be

$$\frac{\text{Total quantity of milk}}{\text{Number of vessels}} = \frac{20+60}{2} \text{ litres} = 40 \text{ litres.}$$

Thus, each vessels would have 40 litres of milk.

The average or Arithmetic Mean (A.M.) or simply mean is defined as follows:

$$\text{mean} = \frac{\text{Sum of all observations}}{\text{number of observations}}$$

Consider these examples.

Example 1 Ashish studies for 4 hours, 5 hours and 3 hours respectively on three consecutive days. How many hours does he study daily on an average?

Solution The average study time of Ashish would be

$$\frac{\text{Total number of study hours}}{\text{Number of days for which he studied}} = \frac{4+5+3}{3} \text{ hours} = 4 \text{ hours per day}$$

Thus, we can say that Ashish studies for 4 hours daily on an average.

Example 2 A batsman scored the following number of runs in six innings:

36, 35, 50, 46, 60, 55

Calculate the mean runs scored by him in an inning.

Solution Total runs = 36 + 35 + 50 + 46 + 60 + 55 = 282.

To find the mean, we find the sum of all the observations and divide it by the number of observations.

Therefore, in this case, mean = $\frac{282}{6}$ = 47. Thus, the mean runs scored in an inning are 47.



Where does the arithmetic mean lie

Try These

How would you find the average of your study hours for the whole week?

THINK, DISCUSS AND WRITE



Consider the data in the above examples and think on the following:

- Is the mean bigger than each of the observations?
- Is it smaller than each observation?

Discuss with your friends. Frame one more example of this type and answer the same questions.

You will find that the mean lies inbetween the greatest and the smallest observations.

In particular, the mean of two numbers will always lie between the two numbers. For example the mean of 5 and 11 is $\frac{5+11}{2} = 8$, which lies between 5 and 11.

Can you use this idea to show that between any two fractional numbers, you can find as many fractional numbers as you like. For example between $\frac{1}{2}$ and $\frac{1}{4}$ you have their average $\frac{\frac{1}{2}+\frac{1}{4}}{2} = \frac{3}{8}$ and then between $\frac{1}{2}$ and $\frac{3}{8}$, you have their average $\frac{7}{16}$ and so on.



TRY THESE

1. Find the mean of your sleeping hours during one week.
2. Find atleast 5 numbers between $\frac{1}{2}$ and $\frac{1}{3}$.

3.2.1 Range

The difference between the highest and the lowest observation gives us an idea of the spread of the observations. This can be found by subtracting the lowest observation from the highest observation. We call the result the **range** of the observation. Look at the following example:

EXAMPLE 3 The ages in years of 10 teachers of a school are:

32, 41, 28, 54, 35, 26, 23, 33, 38, 40

(i) What is the age of the oldest teacher and that of the youngest teacher?

(ii) What is the range of the ages of the teachers?

(iii) What is the mean age of these teachers?

SOLUTION

(i) Arranging the ages in ascending order, we get:

23, 26, 28, 32, 33, 35, 38, 40, 41, 54

We find that the age of the oldest teacher is 54 years and the age of the youngest teacher is 23 years.

(ii) Range of the ages of the teachers = (54 – 23) years = 31 years

(iii) Mean age of the teachers

$$= \frac{23+26+28+32+33+35+38+40+41+54}{10} \text{ years}$$

$$= \frac{350}{10} \text{ years} = 35 \text{ years}$$

EXERCISE 3.1



1. Find the range of heights of any ten students of your class.

2. Organise the following marks in a class assessment, in a tabular form.

4, 6, 7, 5, 3, 5, 4, 5, 2, 6, 2, 5, 1, 9, 6, 5, 8, 4, 6, 7

(i) Which number is the highest?

(ii) Which number is the lowest?

(iii) What is the range of the data?

(iv) Find the arithmetic mean.

3. Find the mean of the first five whole numbers.

4. A cricketer scores the following runs in eight innings:

58, 76, 40, 35, 46, 45, 0, 100.

Find the mean score.

5. Following table shows the points of each player scored in four games:

Player	Game 1	Game 2	Game 3	Game 4
A	14	16	10	10
B	0	8	6	4
C	8	11	Did not Play	13

Now answer the following questions:

(i) Find the mean to determine A's average number of points scored per game.

(ii) To find the mean number of points per game for C, would you divide the total points by 3 or by 4? Why?

(iii) B played in all the four games. How would you find the mean?

(iv) Who is the best performer?

6. The marks (out of 100) obtained by a group of students in a science test are 85, 76, 90, 85, 39, 48, 56, 95, 81 and 75. Find the:

(i) Highest and the lowest marks obtained by the students.

(ii) Range of the marks obtained.

(iii) Mean marks obtained by the group.

7. The enrolment in a school during six consecutive years was as follows:

1555, 1670, 1750, 2013, 2540, 2820

Find the mean enrolment of the school for this period.

8. The rainfall (in mm) in a city on 7 days of a certain week was recorded as follows:

Day	Mon	Tue	Wed	Thurs	Fri	Sat	Sun
Rainfall (in mm)	0.0	12.2	2.1	0.0	20.5	5.5	1.0

(i) Find the range of the rainfall in the above data.

(ii) Find the mean rainfall for the week.

(iii) On how many days was the rainfall less than the mean rainfall.

9. The heights of 10 girls were measured in cm and the results are as follows:

135, 150, 139, 128, 151, 132, 146, 149, 143, 141.

(i) What is the height of the tallest girl? (ii) What is the height of the shortest girl?

(iii) What is the range of the data? (iv) What is the mean height of the girls?

(v) How many girls have heights more than the mean height.

3.3 Mode

As we have said Mean is not the only measure of central tendency or the only form of representative value. For different requirements from a data, other measures of central tendencies are used.

Look at the following example

To find out the weekly demand for different sizes of shirt, a shopkeeper kept records of sales of sizes 90 cm, 95 cm, 100 cm, 105 cm, 110 cm. Following is the record for a week:

Size (in inches)	90 cm	95 cm	100 cm	105 cm	110 cm	Total
Number of Shirts Sold	8	22	32	37	6	105

If he found the mean number of shirts sold, do you think that he would be able to decide which shirt sizes to keep in stock?

$$\text{Mean of total shirts sold} = \frac{\text{Total number of shirts sold}}{\text{Number of different sizes of shirts}} = \frac{105}{5} = 21$$

Should he obtain 21 shirts of each size? If he does so, will he be able to cater to the needs of the customers?

The shopkeeper, on looking at the record, decides to procure shirts of sizes 95 cm, 100 cm, 105 cm. He decided to postpone the procurement of the shirts of other sizes because of their small number of buyers.

Look at another example



The owner of a readymade dress shop says, "The most popular size of dress I sell is the size 90 cm.

Observe that here also, the owner is concerned about the number of shirts of different sizes sold. She is however looking at the shirt size that is sold the most. This is another representative value for the data. The highest occuring event is the sale of size 90 cm.This representative value is called the **mode** of the data.

The mode of a set of observations is the observation that occurs most often.

Example 4 Find the mode of the given set of numbers: 1, 1, 2, 4, 3, 2, 1, 2, 2, 4

Solution Arranging the numbers with same values together, we get

1, 1, 1, 2, 2, 2, 2, 3, 4, 4

Mode of this data is 2 because it occurs more frequently than other observations.

3.3.1 Mode of Large Data

Putting the same observations together and counting them is not easy if the number of observations is large. In such cases we tabulate the data. Tabulation can begin by putting tally marks and finding the frequency, as you did in your previous class.

Look at the following example:

Example 5 Following are the margins of victory in the football matches of a league.

1, 3, 2, 5, 1, 4, 6, 2, 5, 2, 2, 2, 4, 1, 2, 3, 1, 1, 2, 3, 2, 6, 4, 3, 2, 1, 1, 4, 2, 1, 5, 3, 3, 2, 3, 2, 4, 2, 1, 2

Find the mode of this data.

Solution Let us put the data in a tabular form:

Margins of Victory	Tally Bars	Number of Matches
1	𝍸 ||||	9
2	𝍸 𝍸 ||||	14
3	𝍸 ||	7
4	𝍸	5
5	|||	3
6	||	2
	Total	**40**

Try These

Find the mode of

(i) 2, 6, 5, 3, 0, 3, 4, 3, 2, 4, 5, 2, 4

(ii) 2, 14, 16, 12, 14, 14, 16, 14, 10, 14, 18, 14

Looking at the table, we can quickly say that 2 is the 'mode' since 2 has occured the highest number of times. Thus, most of the matches have been won with a victory margin of 2 goals.



Think, Discuss and Write

Can a set of numbers have more than one mode?

Example 6 Find the mode of the numbers: 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 8

Solution Here, 2 and 5 both occur three times. Therefore, they both are modes of the data.

Do This

1. Record the age in years of all your classmates. Tabulate the data and find the mode.
2. Record the heights in centimetres of your classmates and find the mode.



Try These

1. Find the mode of the following data:
12, 14, 12, 16, 15, 13, 14, 18, 19, 12, 14, 15, 16, 15, 16, 16, 15,
17, 13, 16, 16, 15, 15, 13, 15, 17, 15, 14, 15, 13, 15, 14

2. Heights (in cm) of 25 children are given below:
168, 165, 163, 160, 163, 161, 162, 164, 163, 162, 164, 163, 160, 163, 160, 165, 163, 162, 163, 164, 163, 160, 165, 163, 162
What is the mode of their heights? What do we understand by mode here?

Whereas mean gives us the average of all observations of the data, the mode gives that observation which occurs most frequently in the data.

Let us consider the following examples:

(a) You have to decide upon the number of chapattis needed for 25 people called for a feast.

(b) A shopkeeper selling shirts has decided to replenish her stock.

(c) We need to find the height of the door needed in our house.

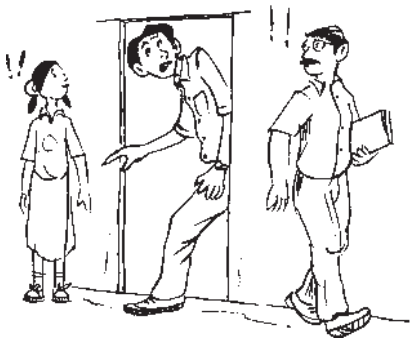
(d) When going on a picnic, if only one fruit can be bought for everyone, which is the fruit that we would get.

In which of these situations can we use the mode as a good estimate?

Consider the first statement. Suppose the number of chapattis needed by each person is 2, 3, 2, 3, 2, 1, 2, 3, 2, 2, 4, 2, 2, 3, 2, 4, 4, 2, 3, 2, 4, 2, 4, 3, 5

The mode of the data is 2 chapattis. If we use mode as the representative value for this data, then we need 50 chapattis only, 2 for each of the 25 persons. However the total number would clearly be inadequate. Would **mean** be an appropriate representative value?

For the third statement the height of the door is related to the height of the persons using that door. Suppose there are 5 children and 4 adults using the door and the height of each of 5 children is around 135 cm. The mode for the heights is 135 cm. Should we get a door that is 144 cm high? Would all the adults be able to go through that door? It is clear that mode is not the appropriate representative value for this data. Would **mean** be an appropriate representative value here?



Why not? Which representative value of height should be used to decide the doorheight?

Similarly analyse the rest of the statements and find the representative value useful for that issue.

Try These

Discuss with your friends and give

(a) Two situations where mean would be an appropriate representative value to use, and

(b) Two situations where mode would be an appropriate representative value to use.



3.4 Median



We have seen that in some situations, arithmetic mean is an appropriate measure of central tendency whereas in some other situations, mode is the appropriate measure of central tendency.

Let us now look at another example. Consider a group of 17 students with the following heights (in cm): 106, 110, 123, 125, 117, 120, 112, 115, 110, 120, 115, 102, 115, 115, 109, 115, 101.

The games teacher wants to divide the class into two groups so that each group has equal number of students, one group has students with height lesser than a particular height and the other group has students with heights greater than the particular height. How would she do that?

Let us see the various options she has:

(i) She can find the mean. The mean is

$$\frac{106+110+123+125+117+120+112+115+110+120+115+102+115+115+109+115+101}{17}$$

$$= \frac{1930}{17} = 113.5$$

So, if the teacher divides the students into two groups on the basis of this mean height, such that one group has students of height less than the mean height and the other group has students with height more than the mean height, then the groups would be of unequal size. They would have 7 and 10 members respectively.

(ii) The second option for her is to find mode. The observation with highest frequency is 115 cm, which would be taken as mode.

There are 7 children below the mode and 10 children at the mode and above the mode. Therefore, we cannot divide the group into equal parts.

Let us therefore think of an alternative representative value or measure of central tendency. For doing this we again look at the given heights (in cm) of students and arrange them in ascending order. We have the following observations:

101, 102, 106, 109, 110, 110, 112, 115, 115, 115, 115, 115, 117, 120, 120, 123, 125

The middle value in this data is 115 because this value divides the students into two equal groups of 8 students each. This value is called as **Median**. Median refers to the value which lies in the middle of the data (when arranged in an increasing or decreasing order) with half of the observations above it and the other half below it. The games teacher decides to keep the middle student as a refree in the game.

Here, we consider only those cases where number of observations is odd.

TRY THESE

Your friend found the median and the mode of a given data. Describe and correct your friends error if any:

35, 32, 35, 42, 38, 32, 34

Median = 42, Mode = 32

Thus, in a given data, arranged in ascending or descending order, the **median** gives us the middle observation.

Note that in general, we may not get the same value for median and mode.

Thus we realise that mean, mode and median are the numbers that are the representative values of a group of observations or data. They lie between the minimum and maximum values of the data. They are also called the measures of the central tendency.

EXAMPLE 7 Find the median of the data: 24, 36, 46, 17, 18, 25, 35

SOLUTION We arrange the data in ascending order, we get 17, 18, 24, 25, 35, 36, 46
Median is the middle observation. Therefore 25 is the median.



EXERCISE 3.2

1. The scores in mathematics test (out of 25) of 15 students is as follows:

19, 25, 23, 20, 9, 20, 15, 10, 5, 16, 25, 20, 24, 12, 20

Find the mode and median of this data. Are they same?

2. The runs scored in a cricket match by 11 players is as follows:

6, 15, 120, 50, 100, 80, 10, 15, 8, 10, 15

Find the mean, mode and median of this data. Are the three same?

3. The weights (in kg.) of 15 students of a class are:

38, 42, 35, 37, 45, 50, 32, 43, 43, 40, 36, 38, 43, 38, 47

(i) Find the mode and median of this data.

(ii) Is there more than one mode?

4. Find the mode and median of the data: 13, 16, 12, 14, 19, 12, 14, 13, 14

5. Tell whether the statement is true or false:

(i) The mode is always one of the numbers in a data.

(ii) The mean is one of the numbers in a data.

(iii) The median is always one of the numbers in a data.

(iv) The data 6, 4, 3, 8, 9, 12, 13, 9 has mean 9.



3.5 Use of Bar Graphs with a Different Purpose

We have seen last year how information collected could be first arranged in a frequency distribution table and then this information could be put as a visual representation in the form of pictographs or bar graphs. You can look at the bar graphs and make deductions about the data. You can also get information based on these bar graphs. For example, you can say that the mode is the longest bar if the bar represents the frequency.

3.5.1 Choosing a Scale

We know that a bar graph is a representation of numbers using bars of uniform width and the lengths of the bars depend upon the frequency and the scale you have chosen. For example, in a bar graph where numbers in units are to be shown, the graph represents one unit length for one observation and if it has to show numbers in tens or hundreds, one unit length can represent 10 or 100 observations. Consider the following examples:

Example 8 Two hundred students of 6^{th} and 7^{th} classes were asked to name their favourite colour so as to decide upon what should be the colour of their school building. The results are shown in the following table. Represent the given data on a bar graph.

Favourite Colour	Red	Green	Blue	Yellow	Orange
Number of Students	43	19	55	49	34

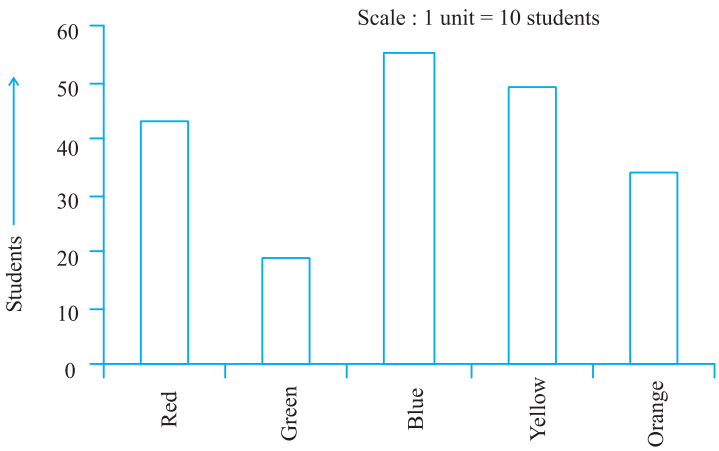
Answer the following questions with the help of the bar graph:

(i) Which is the most preferred colour and which is the least preferred?

(ii) How many colours are there in all? What are they?

Solution Choose a suitable scale as follows:

Start the scale at 0. The greatest value in the data is 55, so end the scale at a value greater than 55, such as 60. Use equal divisions along the axes, such as increments of 10. You



know that all the bars would lie between 0 and 60. We choose the scale such that the length between 0 and 60 is neither too long nor too small. Here we take 1 unit for 10 students.

We then draw and label the graph as shown.

From the bar graph we conclude that

(i) Blue is the most preferred colour (Because the bar representing Blue is the tallest).

(ii) Green is the least preferred colour. (Because the bar representing Green is the shortest).

(iii) There are five colours. They are Red, Green, Blue, Yellow and Orange. (These are observed on the horizontal line)

EXAMPLE 9 Following data gives total marks (out of 600) obtained by six children of a particular class. Represent the data on a bar graph.

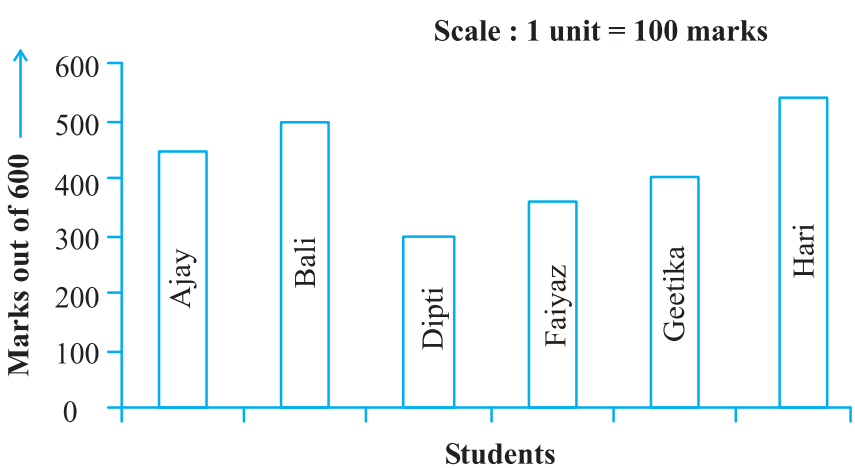
Students	Ajay	Bali	Dipti	Faiyaz	Geetika	Hari
Marks Obtained	450	500	300	360	400	540



SOLUTION

(i) To choose an appropriate scale we make equal divisions taking increments of 100. Thus 1 unit will represent 100 marks. (What would be the difficulty if we choose one unit to represent 10 marks?)

(ii) Now represent the data on the bar graph.



Drawing double bar graph

Consider the following two collections of data giving the average daily hours of sunshine in two cities Aberdeen and Margate for all the twelve months of the year. These cities are near the south pole and hence have only a few hours of sunshine each day.

In Margate												
	Jan.	**Feb.**	**Mar.**	**April**	**May**	**June**	**July**	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**
Average hours of Sunshine	2	$3\frac{1}{4}$	4	4	$7\frac{3}{4}$	8	$7\frac{1}{2}$	7	$6\frac{1}{4}$	6	4	2
In Aberdeen												
Average hours of Sunshine	$1\frac{1}{2}$	3	$3\frac{1}{2}$	6	$5\frac{1}{2}$	$6\frac{1}{2}$	$5\frac{1}{2}$	5	$4\frac{1}{2}$	4	3	$1\frac{3}{4}$

By drawing individual bar graphs you could answer questions like

(i) In which month does each city has maximum sunlight? or

(ii) In which months does each city has minimum sunlight?



However, to answer questions like "In a particular month, which city has more sunshine hours", we need to compare the average hours of sunshine of both the cities. To do this we will learn to draw what is called a double bar graph giving the information of both cities side-by-side.

This bar graph (Fig 3.1) shows the average sunshine of both the cities.

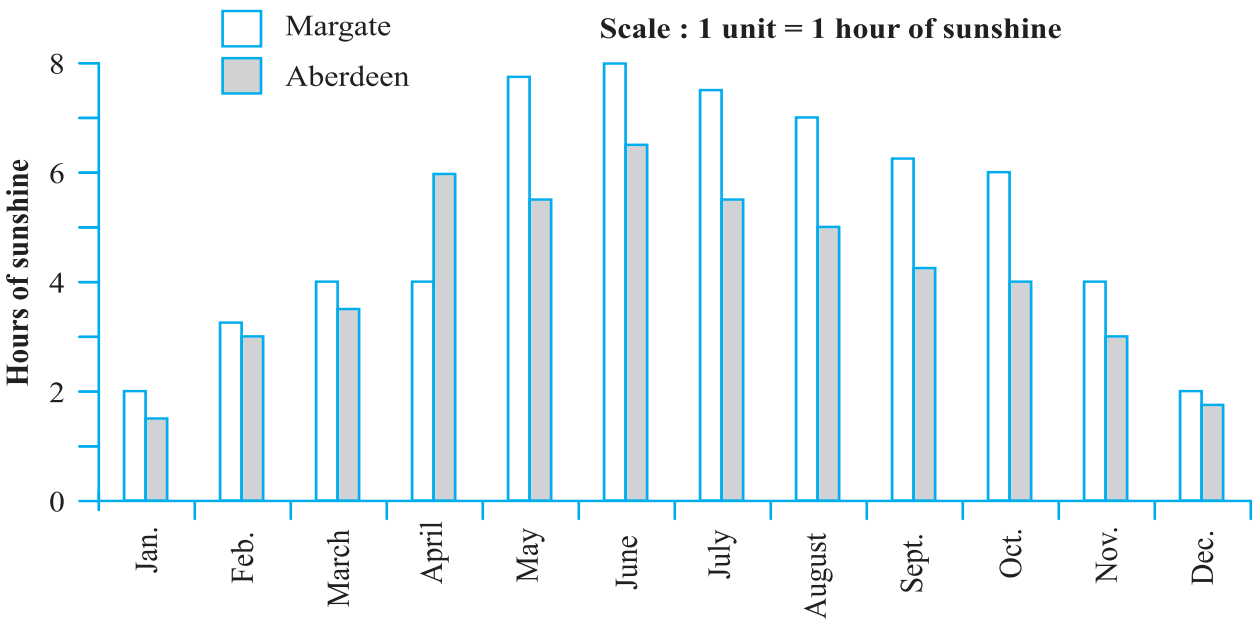
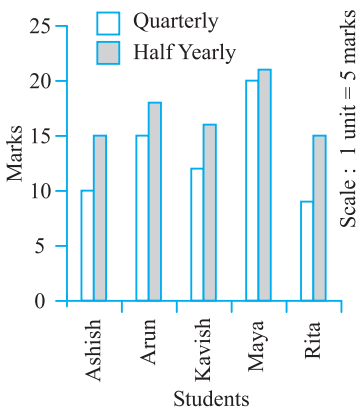


Fig 3.1

For each month we have two bars, the heights of which give the average hours of sunshine in each city. From this we can infer that except for the month of April, there is always more sunshine in Margate than in Aberdeen. You could put together a similiar bar graph for your area or for your city.

Let us look at another example more related to us.

EXAMPLE 10 A mathematics teacher wants to see, whether the new technique of teaching she applied after quarterly test was effective or not. She takes the scores of the 5 weakest children in the quarterly test (out of 25) and in the half yearly test (out of 25):



Students	Ashish	Arun	Kavish	Maya	Rita
Quarterly	10	15	12	20	9
Half yearly	15	18	16	21	15

SOLUTION She draws the adjoining double bar graph and finds a marked improvement in most of the students, the teacher decides that she should continue to use the new technique of teaching.

Can you think of a few more situations where you could use double bar graphs?

TRY THESE



1. The bar graph (Fig 3.2) shows the result of a survey to test water resistant watches made by different companies.

Each of these companies claimed that their watches were water resistant. After a test the above results were revealed.

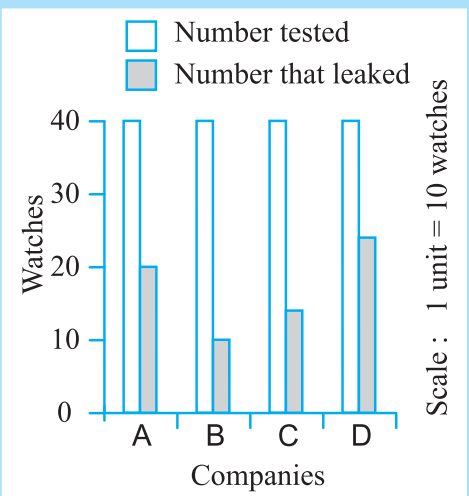


Fig 3.2

(a) Can you work out a fraction of the number of watches that leaked to the number tested for each company?

(b) Could you tell on this basis which company has better watches?

2. Sale of English and Hindi books in the years 1995, 1996, 1997 and 1998 are given below:

Years	1995	1996	1997	1998
English	350	400	450	620
Hindi	500	525	600	650

Draw a double bar graph and answer the following questions:

(a) In which year was the difference in the sale of the two language books least?.

(b) Can you say that the demand for English books rose faster? Justify.

EXERCISE 3.3

1. Use the bar graph (Fig 3.3) to answer the following questions.

(a) Which is the most popular pet? (b) How many students have dog as a pet?

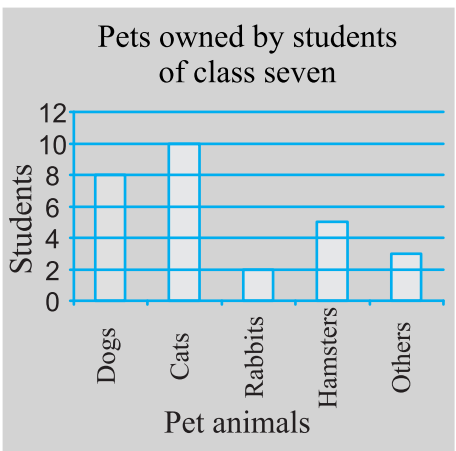


Fig 3.3

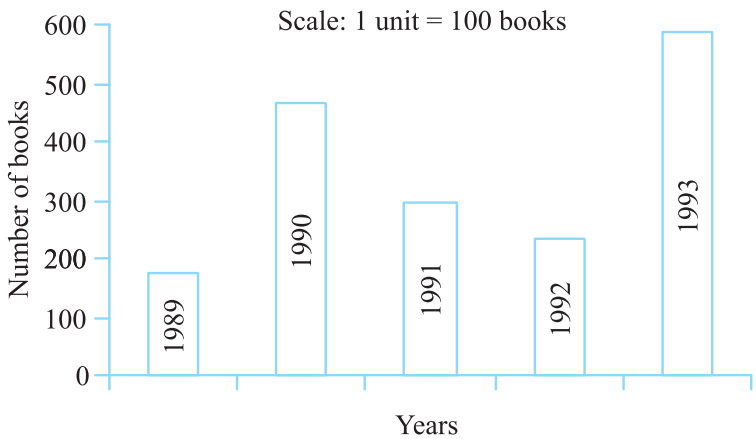


Fig 3.4

2. Read the bar graph (Fig 3.4) which shows the number of books sold by a bookstore during five consecutive years and answer the following questions:

(i) About how many books were sold in 1989? 1990? 1992?

(ii) In which year were about 475 books sold? About 225 books sold?

(iii) In which years were fewer than 250 books sold?

(iv) Can you explain how you would estimate the number of books sold in 1989?

3. Number of children in six different classes are given below. Represent the data on a bar graph.

Class	**Fifth**	**Sixth**	**Seventh**	**Eighth**	**Ninth**	**Tenth**
Number of Children	135	120	95	100	90	80

(a) How would you choose a scale?

(b) Answer the following questions:

(i) Which class has the maximum number of children? And the minimum?

(ii) Find the ratio of students of class sixth to the students of class eight.

4. The performance of a student in 1st Term and 2nd Term is given. Draw a double bar graph choosing appropriate scale and answer the following:

Subject	**English**	**Hindi**	**Maths**	**Science**	**S. Science**
1st Term (M.M. 100)	67	72	88	81	73
2nd Term (M.M. 100)	70	65	95	85	75

(i) In which subject, has the child improved his performance the most?

(ii) In which subject is the improvement the least?

(iii) Has the performance gone down in any subject?

5. Consider this data collected from a survey of a colony.

Favourite Sport	Cricket	Basket Ball	Swimming	Hockey	Athletics
Watching	1240	470	510	430	250
Participating	620	320	320	250	105



(i) Draw a double bar graph choosing an appropriate scale.
What do you infer from the bar graph?

(ii) Which sport is most popular?

(iii) Which is more preferred, watching or participating in sports?

6. Take the data giving the minimum and the maximum temperature of various cities given in the beginning of this Chapter (Table 3.1). Plot a double bar graph using the data and answer the following:

(i) Which city has the largest difference in the minimum and maximum temperature on the given date?

(ii) Which is the hottest city and which is the coldest city?

(iii) Name two cities where maximum temperature of one was less than the minimum temperature of the other.

(iv) Name the city which has the least difference between its minimum and the maximum temperature.

WHAT HAVE WE DISCUSSED?

1. Average is a number that represents or shows the central tendency of a group of observations or data.

2. Arithmetic mean is one of the representative values of data.

3. Mode is another form of central tendency or representative value. The mode of a set of observations is the observation that occurs most often.

4. Median is also a form of representative value. It refers to the value which lies in the middle of the data with half of the observations above it and the other half below it.

5. A bar graph is a representation of numbers using bars of uniform widths.

6. Double bar graphs help to compare two collections of data at a glance.