10 How Quantities Combine: Understanding Data

0905CH10

10.1 Combining Things

In earlier grades, we looked at how the average conveys the 'centre' of given data. In this chapter, we shall extend the concept of average to a more general setting called 'weighted average'.

10.1.1 Average of Averages

Example 1: In a badminton academy, there is a group of 11 trainees — 8 seniors and 3 juniors. Their heights and average heights (in cm) are given in the table below. Find the average height of the whole group.

	Heights (in cm)	Average Height (in cm)
Seniors	165, 169, 164, 167, 170, 159, 164, 166	165.5
Juniors	146, 149, 153	149.33

Two students calculate the average height of the whole group in two ways:

Method 1

$$\frac{165.5 + 149.33}{2}$$

$$= \frac{314.83}{2}$$

$$= 157.415$$

Method 2

$$\frac{165 + 169 + 164 + 167 + 170 + 159 + 164 + 166 + 146 + 149 + 153}{11}$$

$$= \frac{1772}{11} = 161.09$$

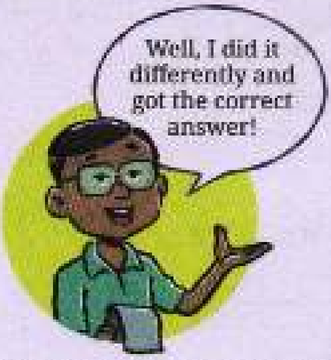
Whose calculation gives the correct average height of the whole group?

The average height of the whole group is the sum of the heights of all eleven members divided by 11, as in Method 2.

Think and Reflect

Why did Method 1 not work?

Method 1 does not work because it treats the average heights of seniors and of juniors equally — but there are many more seniors than juniors!



Shreyas calculated it differently as

$$\frac{(165.5 \times 8) + (149.33 \times 3)}{8 + 3}$$

$$= \frac{1324 + 448}{11} = \frac{1772}{11}$$

$$= 161.09.$$

Do you understand why this also works?

Can you see why 165.5 × 8 gives the sum of the heights of all the seniors and 149.33 × 3 gives the sum of the heights of all the juniors?

This phenomenon is explained below with a generalisation. Suppose we have two collections of data — Collection 1 and Collection 2.

Suppose the data in Collection 1, having n values, is $x_1, x_2, x_3, \ldots, x_n$.

The average of Collection 1 is

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = a.$$

Thus the sum of the data in Collection 1 is

$$x_1 + x_2 + x_3 + \cdots + x_n = an.$$

Suppose the data in Collection 2, having m values, is $y_1, y_2, y_3, \ldots, y_m$.

The average of Collection 2 is

$$\frac{y_1 + y_2 + y_3 + \cdots + y_m}{m} = b.$$

Thus the sum of the data in Collection 2 is

$$y_1 + y_2 + y_3 + \cdots + y_m = bm.$$

The average of the combined data is therefore

$$\frac{\textit{(Sum of values in Collection 1)} + \textit{(Sum of values in Collection 2)}}{\textit{(Number of values in Collection 1)} + \textit{(Number of values in Collection 2)}}$$

$$= \frac{an + bm}{n + m}.$$

Example 2: Jaspreet recently learnt cycling. She has explored different routes in her town. She tracked how much time she cycled on weekdays over the last 3 weeks. Find the mean time spent cycling per weekday over the last 3 weeks.

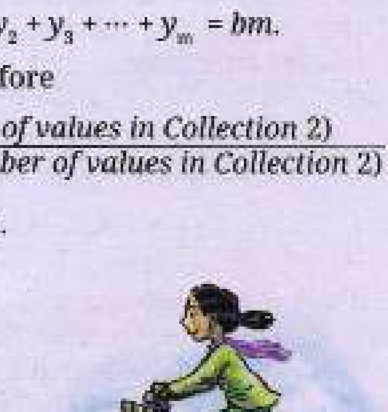


Table 10.1: Time spent cycling, in minutes, by Jaspreet

	Day 1	Day 2	Day 3	Day 4	Day 5	Weekly Average
Week 1	10	13	10	15	16	12.8
Week 2	14	10	20	17	18	15.8
Week 3	15	18	14	20	23	18

Two ways of calculating this are given here.

Method 1

$$\frac{(12.8 + 15.8 + 18)}{3}$$
$$= \frac{46.63}{3}$$
$$= 15.53$$

Method 2

$$\frac{(12.8 \times 5) + (15.8 \times 5) + (18 \times 5)}{5 + 5 + 5}$$
$$= \frac{(64 + 79 + 90)}{15}$$
$$= 15.53$$

The average time spent cycling by Jaspreet on weekdays is 15.53 minutes.

We saw earlier how Method 1 (badminton example) did not produce the correct value. Why does Method 1 give the correct answer in this case? When does Method 1 work and when does it not work? Let us find out.

In general, the average of data in three collections is given as

$$\frac{\textit{(Sum of values in Collection 1)} + \textit{(Sum of values in Collection 2)} + \textit{(Sum of values in Collection 3)}}{\textit{(No. of values in Collection 1)} + \textit{(No. of values in Collection 2)} + \textit{(No. of values in Collection 3)}}$$

$$= \frac{ap + bq + cr}{p + q + r},$$

where a, b, c are the averages of each collection, and p, q, r are the sizes of the respective collections.

In the cycling scenario, Jaspreet cycled 5 days every week for 3 weeks. This means that all three collections are of the same size, that is 5. Thus, in the general form, the average becomes,

$$\frac{ap + bp + cp}{p + p + p} \text{ (where } p \text{ is the size of the collections)}$$

$$= \frac{(a + b + c)p}{3p} = \frac{a + b + c}{3}.$$

This is the same as adding the 3 weekly averages and dividing the sum by 3. Here, each week has 5 days, so each average comes from the same number of days; that is why treating each week's average equally works. If the weeks had different numbers of cycling days, this method would no longer be correct.

We see that when the sizes of the collections are the same, to find the average of the combined collection, we can add the averages of the collections and divide the sum by the number of collections.

10.1.2 Mixtures

Let us consider what happens when things are mixed.

Example 3: Two glasses of equal quantities of lemonade are prepared. One glass has 10% jaggery and the other has 20% jaggery. If we mix the lemonade from both glasses, what is the concentration of jaggery in the mixture?

Since both the glasses contain equal amounts of lemonade, the mixture will have 15% jaggery. We can also arrive at this mathematically—

Suppose y is the quantity of lemonade in each glass.

The concentration of jaggery in the mixture is

$$\frac{\textit{quantity of jaggery in the mixture}}{\textit{quantity of the mixture}}$$

$$= \frac{\frac{10}{100}y + \frac{20}{100}y}{y+y}$$

$$= \frac{0.1y + 0.2y}{y+y} = \frac{0.3y}{2y} = 0.15.$$

Since the glasses had equal quantities of lemonade, the concentration of jaggery in the combined mixture is midway between the concentrations of jaggery in the individual glasses. This is similar to the cycling example, where all weeks had equal numbers of days.

Example 4: A bowl of 500 mL lemonade has 10% jaggery. A glass of 200 mL lemonade has 20% jaggery. If we mix the lemonade from the bowl and the glass, what is the concentration of jaggery in the mixture?

Can you estimate what part of the mixture is jaggery? Is it 15%, or is it more or less? Why do you think so?

Since the bowl has a larger quantity of lemonade, the concentration of jaggery in the mixture will be closer to 10% (less than midway).

The concentration of jaggery in the mixture is given by

$$\frac{\textit{quantity of jaggery in the mixture}}{\textit{quantity of the mixture}}$$

$$= \frac{500 \times 0.1 + 200 \times 0.2}{500 + 200}$$

$$= \frac{50 + 40}{700} = \frac{90}{700} \approx 0.13.$$

Notice that the final concentration is not the simple average of 10% and 20%. It is closer to 10% because the larger quantity of lemonade comes from the 10% mixture. In general, each concentration influences the final concentration according to the quantity of mixture it contributes.

Example 5: Brass is an alloy of copper and zinc. Batch A of brass weighs 200 kg of which 70% is copper. Batch B weighs 120 kg of which 50% is copper. Batch C is 45% copper. When all three batches are combined, we get an alloy that is 55% copper. What is the weight of Batch C?

The concentration of copper in the mixture is given by

$$\frac{\textit{quantity of copper in the mixture}}{\textit{quantity of the mixture}}.$$

Suppose y is the weight (in kg) of Batch C. From the given information, we know that

$$0.55 = \frac{\textit{(Quantity of copper in Batch A + Quantity of copper in Batch B + Quantity of copper in Batch C)}}{\textit{Total weight of brass in all batches}}$$

$$0.55 = \frac{200 \times 0.7 + 120 \times 0.5 + y \times 0.45}{200 + 120 + y}$$

$$0.55\,(320 + y) = 200 + 0.45y.$$

$$y = 240.$$

Batch C weighs 240 kg.

More generally, if n things with concentrations $x_1, x_2, x_3, \ldots, x_n$ and respective volumes/weights $w_1, w_2, w_3, \ldots, w_n$ are mixed, the concentration of the mixture is given by

$$x = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n}{w_1 + w_2 + w_3 + \cdots + w_n}.$$

Such an average is called the **weighted mean** or **weighted average**. Here, the prefix 'weighted' does not refer to actual weights but abstract weights; however, in some cases, these can be actual weights. That is, the weighted mean is an average where each value is weighted (multiplied) by how important or how big it is. We shall study other contexts where the weighted average is used.

The above formula for the weighted mean was written down by Brahmagupta in his work Brāhmasphuṭasiddhānta (c. 628 CE). He used the formula to define the mean depth x of an irregularly-shaped pit. If the pit has depth measurement x_1 for length w_1, depth measurement x_2 for length w_2, ... and depth measurement x_n for length w_n, then the mean depth x is given by the above formula.

The concept of weighted arithmetic mean was also used by Indian mathematicians like Śrīdharācārya (c. 750 CE) to estimate the purity of gold-alloys (essentially today's karat calculations).

The concept of the weighted mean eventually appeared in other parts of the world, e.g., Europe, where it gained prominence by the 1700s. The weighted mean now plays a crucial role across mathematics and statistics, allowing us to combine data of different sizes or importance into one number to give an overall accurate picture.

Exercise Set 10.1

1. The average score of students on a test in Section A is 72 and that of students in Section B is 76. What is the combined average of both the sections given that Section A has 30 students and Section B has 25 students?
2. A farmer mixes three equal quantities of fertilisers. The first one contains $\frac{1}{10}$ nitrogen, the second contains $\frac{9}{50}$ nitrogen, and the third contains $\frac{3}{60}$ nitrogen. What is the fraction of nitrogen in the mixture?
3. (Śrīdharācārya, Pāṭīgaṇita, c. 750 CE) In ancient India, Varṇa was the measure of gold purity. A purity of 16 varṇa meant pure gold; in general, a purity of k varṇa meant that the gold-alloy was k/16 gold and the rest impurities. (Now the term used is karat; 16 Varṇa = 24 karat.)

   Suppose a goldsmith melts together three pieces of gold: 9 units at 12 varṇa, 5 units at 10 varṇa, and 17 units at 11 varṇa. Find the purity in varṇa of the combined gold.
4. The average rainfall per day in the months of May, June, and July in a certain location are 3.5 mm, 10 mm and 8.7 mm respectively. Write an expression that gives their combined average.
5. Calculate the concentration of spice mix in these two scenarios.
   (i) A 100 mL *kashayam/kadha* with 5% spice mix, a 200 mL one with 10% spice mix, and a 300 mL one with 15% spice mix are combined.
   (ii) A 300 mL *kashayam/kadha* with 5% spice mix, a 200 mL one with 10% spice mix, and a 100 mL one with 15% spice mix are mixed.

10.1.3 Custom Weights

The weighted mean is also used whenever parts of data require different levels of importance or weightage. Let us look at an example.

Example 6: Rehmat's marks in Maths are as follows: 60% in internal tests, 64% in the project, 73% in the final exam. The annual percentage score is calculated by combining the internals, project, and final exam in the ratio 3 : 2 : 5. What is Rehmat's annual score in Maths?

Suppose all modes of assessment had an equal weightage. Then the annual score can be obtained by calculating the arithmetic mean of 60%, 64%, and 73%, which is $\frac{60 + 64 + 73}{3} = 65.6\%$.

The given ratio 3 : 2 : 5 indicates the relative weightage and importance that each form of evaluation gets. The data with these relative weights is equivalent to 60, 60, 60, 64, 64, 73, 73, 73, 73, 73. The arithmetic mean is given by $\frac{60 \times 3 + 64 \times 2 + 73 \times 5}{10} = 67.3\%$.

This can also be expressed as a weighted mean of 60%, 64%, 73% with weights 3, 2, 5 respectively:

$$\frac{\left(\frac{60}{100}\right) \times 3 + \left(\frac{64}{100}\right) \times 2 + \left(\frac{73}{100}\right) \times 5}{3 + 2 + 5} = \frac{\left(\frac{673}{100}\right)}{10} = 67.3\%.$$

The weights assigned to the values in any data indicate the relative importance/emphasis that each value gets with respect to the others.

Exercise Set 10.2

1. Savitri's marks in Kashmiri are as follows: 35 out of 50 in internal tests, 44 out of 60 in the project, and 80 out of 100 in the final exam. The annual percentage score is calculated by combining the internals, project, and final exam in the ratio 3 : 4 : 5. Which of the following expression(s) gives her annual score (as a percentage) in Kashmiri?

   (i) $\frac{35 \times 3 + 44 \times 4 + 80 \times 5}{3 + 4 + 5}$

   (ii) $\frac{\left(\frac{35}{100}\right) \times 3 + \left(\frac{44}{100}\right) \times 4 + \left(\frac{80}{100}\right) \times 5}{3 + 4 + 5}$

   (iii) $\frac{\left(\frac{35}{50}\right) \times 3 + \left(\frac{44}{60}\right) \times 4 + \left(\frac{80}{100}\right) \times 5}{3 + 4 + 5} \times 100$

(iv) $$\frac{\left(\frac{35}{50} \times 100\right) \times 3 + \left(\frac{44}{60} \times 100\right) \times 4 + \left(\frac{80}{100} \times 100\right) \times 5}{3 + 4 + 5}$$

Here are a few more situations where the concept of weighted mean is relevant:

1. **Ratings of a hotel/eatery:** The experience of eating at a restaurant is influenced not only by the taste or quality of the food but also by the ambience and the service provided. Every customer is asked to rate each of these aspects. The customer's combined rating is determined by the relative weights assigned to food, service, and ambience. Suppose one of the customers has given a rating as shown in the picture. Suppose the weights assigned to food, service, and ambience are 4 : 3 : 2 respectively. The user's rating for the restaurant is $\frac{5 \times 4 + 3 \times 3 + 4 \times 2}{4 + 3 + 2} = 4.11$.

2. **Ratings of a product:** A store makes and sells portable furniture. The store owners make changes to their products regularly. To calculate the store's average rating, it makes sense for them to give their current products more weight than older products. For example, they may choose to assign twice the weightage to customer reviews in the last 3 months compared to older reviews.

We use weighted mean whenever we want the combined result to reflect how much each part contributes to the whole — by amount, time, or importance — rather than just counting each part equally.

Exercise Set 10.3

1. A stationery shop owner made ₹8000 selling books, of which 30% is the profit amount, and ₹1000 selling book covers, of which 50% is the profit amount. What is the percentage of profit on the total sales?
2. A white stork's migration is tracked. The average daily distance travelled, calculated over 20 days, is 44.5 km. On the 21st day, it flew 55 km. What is the average daily distance travelled over these 21 days? Make a guess before you calculate.

3. A 600 mL solution with 5% salt is mixed with a 300 mL solution with 8% sugar. What are the concentrations of salt and sugar in the mixture?

   (i) Salt: 5%, Sugar: 8%
   (ii) Salt: 13%, Sugar: 3%
   (iii) Salt: 6%, Sugar: 6%
   (iv) Salt: 5.55%, Sugar: 8.88%
   (v) Salt: 3.33%, Sugar: 2.67%
   (vi) Salt: 4.1%, Sugar: 7.08%

4. At a *panipuri* (*golgappa*) stall, the concentration of spice in the *pani* (spiced water) was 8%. Many customers complained that it was too spicy. What quantity of regular water should be mixed into the 10 litres of *pani* so that the spice level is reduced to $\left(\frac{3}{4}\right)^{\text{th}}$ of the original concentration?

5. A physical fitness evaluation is being undertaken. The final marks are calculated by combining the marks for strength, flexibility, and agility in the ratio 4 : 5 : 6. Rashi has scored 60, 65, and 70. Keerthi has scored 55, 65, and 75 respectively.
   (i) Find out whose total is more without calculating.
   (ii) What are their final marks?

6. A restaurant collected ratings from 10 customers on a scale of 1 to 5. The resulting data is shown in the table below.

	5★	4★	3★	2★	1★
Food	5	3	2	0	0
Ambience	0	4	5	1	0
Service	1	2	2	4	1

   What is the average rating if the metrics are combined with the weights food : ambience : service = 6 : 5 : 4?

7. The following table shows the weight data of langurs in an animal facility. Without doing any computations, can you tell

whether there are more male langurs or more female langurs? Or are they equal in number?

	Male langurs	Female langurs	All langurs
Average weight	16.5 kg	13.8 kg	14.925 kg
Number of langurs			60

(i) Which of the following expression(s) describes the given scenario?

(a) $\dfrac{16.5x + 13.8y}{x + y} = 14.925$

(b) $\dfrac{16.5x + 13.8y}{60} = -14.925$

(c) $\dfrac{16.5x + 13.8y}{2} = -14.925$

(d) $\dfrac{16.5x + 13.8y}{16.5 + 13.8} = -14.925$

(ii) Find out how many male langurs are present.

(iii) A female langur weighing 15.2 kg is admitted to the facility. What is the average weight of the female langurs after this?

(iv) Two male langurs weighing 16.9 kg and 16.1 kg are released from the facility. What is the average weight of the male langurs after this?

(v) Now, suppose one of the male langurs lost 1 kg of weight. What is the average weight of all the male langurs after this?

8. Dorjee has collected 1 litre of water from the Dead Sea! Using the information in the table, answer the following questions. A calculator can be used if necessary.

Water Source	Salinity
Dead Sea	≈ 34%
Other seas and oceans	≈ 3.5%
Ground water	≈ 0.01%
Purified drinking water	≈ 0.001%

(i) What is the salinity of the mixture if he mixes 1 litre of water from the Dead Sea with 2 litres of purified drinking water?

(ii) Is it possible to mix water from the Dead Sea and purified drinking water to get a mixture with the salinity of groundwater? Why/Why not? What quantity of purified drinking water should Dorjee mix with 1 litre of water from the Dead Sea to get a mixture having the salinity of groundwater?

(iii) Is it possible to mix water from the Dead Sea and groundwater to get a mixture with the salinity of purified drinking water? Why/Why not? What quantity of groundwater should he mix with 1 litre of water from the Dead Sea to get a mixture having the salinity of purified drinking water?

10.2 Visualising and Interpreting Data

10.2.1 Stacking Columns

The following table shows the average weekly expenditure, in rupees (₹), of three families across categories. What would a cluster-column chart for this data look like?

Table 10.2

	Housing	Food	Education	Transportation	Healthcare	Recreation
Family A	1940	1700	1280	1200	1770	535
Family B	1750	1546	1500	1280	1210	0
Family C	950	1700	1540	1400	1300	150

There are two possibilities depending on the choice of the cluster; the clusters can be categories or families:

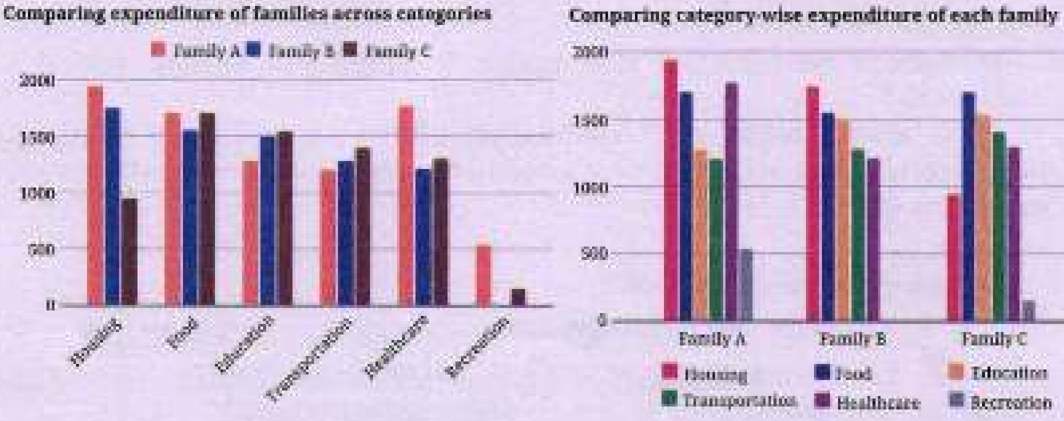


Fig. 10.1: Two cluster-column graphs showing different ways of visualising the data in Table 10.2

Answer the following questions based on the charts above. Try to answer each question by looking at Choice 1 first, and then Choice 2. Which was easier?

1. Which family spent the least on housing?
2. Which family spent the most on healthcare?
3. Approximately how much did Family A spend on education?
4. Which category did Family B spend most on?
5. Did Family A spend more on food or on healthcare?

You may have found Choice 1 more suitable to answer Questions 1 and 2, and Choice 2 more suitable to answer Questions 4 and 5.

The choice of the cluster-column chart depends on what the primary focus is — comparing different families' expenditures in each category (Choice 1) or comparing category-wise expenditures within each family (Choice 2).

Find the answers to the following questions.

6. Which family spent the least amount overall?
7. Approximately how much did Family C spend overall?

8. Approximately what percentage of Family C's expenses were on housing?

Was it as straightforward to answer these questions as the earlier questions?

Is there a visualisation that makes it simpler to answer such questions?

Look at the following visualisation of the same data. Try to understand how the chart is organised and answer the three questions above.

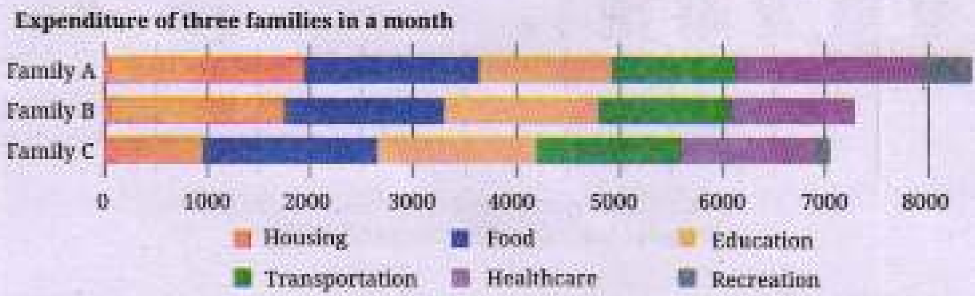


Fig. 10.2: A Stacked Bar Chart of the data in Table 10.2

Such a graph/chart is called a **stacked bar or column chart**. Bars within a cluster are joined one after the other in a particular order. **Stacked bar charts are used when we want to compare the totals but also want to keep track of the components that make up the totals.** Here, the total length of a full stacked (combined) bar represents the total expenditure of a family in that month. Also, the length of a smaller bar and the full stacked bar can be compared to get an estimate of what fraction or percentage the smaller bar makes up of the full stacked bar.

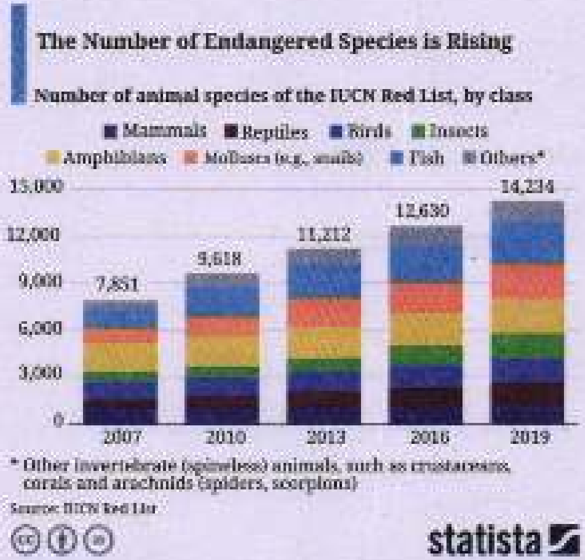
Now, use the stacked bar chart to answer the Questions 1–5 shown earlier.

Is it possible? Is it easy? If not, what made it harder than before?

Except for the first small bar, the rest of the stacked small bars do not begin from 0. This makes it a little harder to estimate the quantity corresponding to a bar. Also, bars across clusters need not have a common base or a starting point. So, comparing bars within clusters and across clusters may be harder. However, seeing the totals of all the clusters is easier.

Exercise Set 10.4

1. The following stacked column chart shows the number of animal species in the IUCN Red List, by class, over time.



The International Union for Conservation of Nature (IUCN) Red List is a large, regularly-updated database that tells us the extinction risk of species across the world. It classifies species into groups such as Least Concern, Vulnerable, Endangered, Critically Endangered, Extinct in the Wild and Extinct, based on scientific data about their numbers, trends and habitats.

(i) What does the number 14,234 on the top of the column 2019 represent?

(ii) Approximately how many reptile species were in the list in 2016?

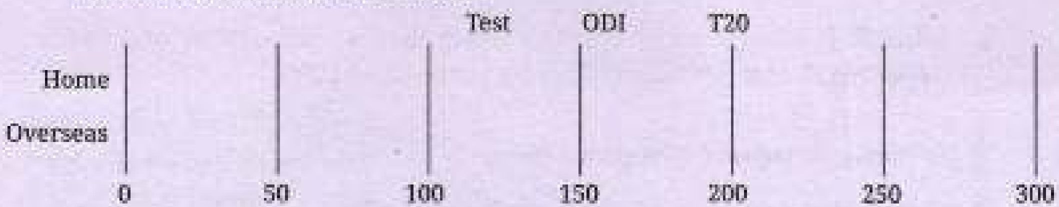
(iii) Which class(es) of species have seen a relatively small increase in count between 2007 and 2019?

2.

	Test	ODI	T20
Home	62	100	32
Overseas	172	49	71

The wickets taken by a bowler in International Cricket matches till 2025 are shown in the table. Complete the given stacked bar charts (the bar lengths can be approximate),

(i) comparing the total wickets taken at home vs. overseas:



(ii) The total wickets taken in each format:

10.2.2 An Alternative to Pie Chart

Try answering the following question based on the expenditure stacked bar chart you saw earlier:

Which family has the smallest share of their total expenditure towards healthcare?

Was it easy?

If not, what visualisation can help answer such questions easily?

Last year, we learnt how pie charts can capture proportions of quantities. Let us visualise this data through pie charts:

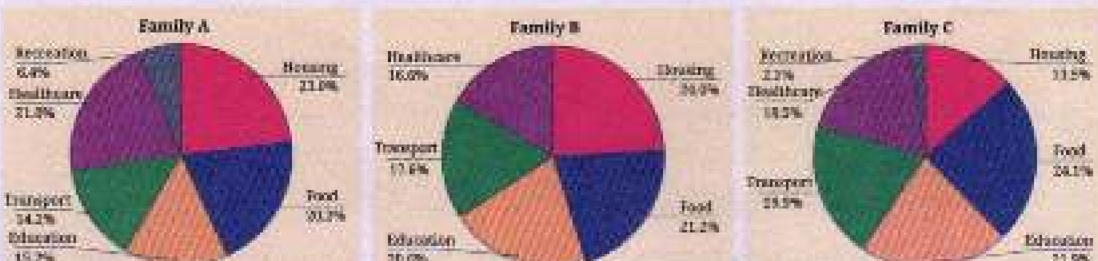


Fig. 10.3: Pie charts of data of each family in Table 10.2

Now, consider the following visualisation showing proportions for the same data:

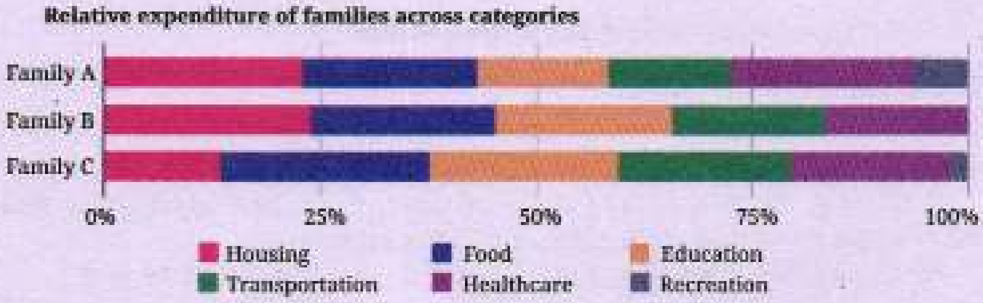


Fig. 10.4: A 100% stacked bar chart of data in Table 10.2

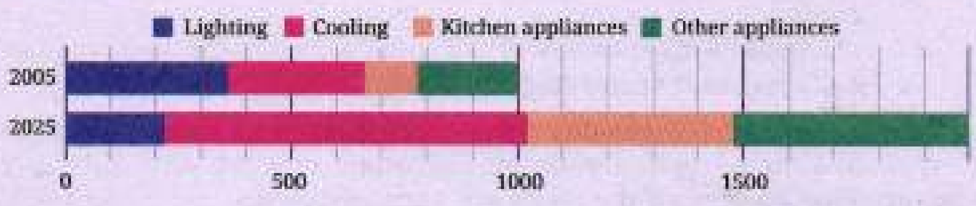
Here, the total expenditure of each of the three families is shown by a stacked bar of a fixed length. The lengths of the smaller bars in each stacked bar are proportional to the corresponding expenditures of that family. Such a chart is called a **100% stacked bar chart**. It is similar to a pie chart—the complete angle in a pie chart is split proportionately based on the quantities, whereas here, a fixed length is split proportionately.

In pie charts, comparing angles in different orientations may be difficult. It may be easier to compare lengths across stacked bars. However, even in 100% stacked bar charts, if the lengths of a particular bar are similar across the stacked bars but the location is different, comparison may be difficult, as in the chart above.

Powering Up!

Example 7: The average category-wise daily electricity consumption of a house, in kWh, in 2005 and 2025 is shown in the table below. The corresponding stacked bar chart is also shown.

	Lighting	Cooling	Kitchen appliances	Other appliances
2005	360	300	120	220
2025	220	800	460	520



Make the corresponding 100% stacked bar chart.

In 2005, the total unit consumption was 360 + 300 + 120 + 220 = 1000 units.

The percentage share of lighting in this was $\frac{360}{1000} \times 100 = 36\%$.

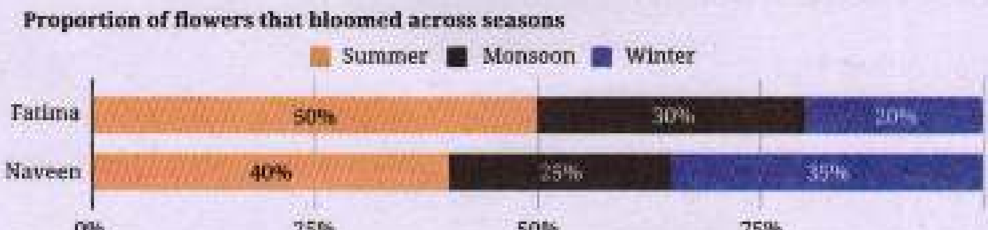
Similarly, the percentage share of cooling, kitchen appliances, and other appliances was 30%, 12%, and 22% respectively.

The calculation for the year 2025 can be done in the same way.

What do you observe? What do you find interesting in the data/graphs? Discuss with others.

What Can We Infer?

Example 8: The 100% stacked bar below compares the proportion of flowers blooming in Fatima's garden and Naveen's garden across seasons.



Which of the following statements can be inferred from this chart?

(i) In Fatima's garden, there were more blooms in the summer than in the monsoon.

(ii) In summer, Fatima's garden had more blooms than Naveen's garden.

(iii) In winter, Fatima's garden had fewer blooms than Naveen's garden.

(iv) The total number of flowers, across seasons, is the same in both gardens but it varies in each season.

(**Hint:** Find out the blooms per season in each garden if both the gardens had the same total number of blooms over the year. Similarly find out what happens if the totals are very different.)

What does the 100% stacked bar tell us?

In each garden, the total number of flowers for the whole year is shown as one bar and is taken as 100%. The colours show what fraction of that total bloomed in each season. So for Fatima's garden, the bar tells us: 50% of her flowers bloomed in summer, 30% in monsoon, and 20% in winter. For Naveen's garden, the bar tells us: 40% in summer, 35% in monsoon, and 25% in winter.

Statement (i): It compares seasons within the same garden (Fatima's). So, within Fatima's garden, the summer part is the largest fraction. This means she actually had more flowers in summer than in monsoon. So, (i) must be true.

Statements (ii) and (iii): These compare across gardens in the same season.

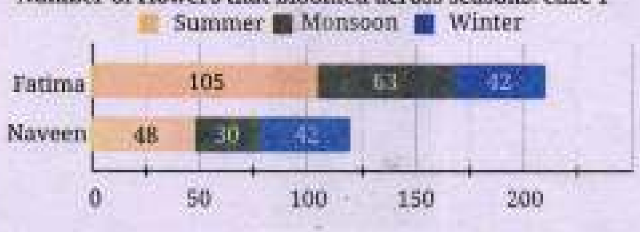
For example, in statement (ii), Fatima has a higher percentage in summer (50%) than Naveen (40%). But the chart does not tell us how many flowers they each had in total. Naveen's total could be much larger. Then even 40% of Naveen's total might still be more flowers than 50% of Fatima's total. That means we cannot be sure who has more flowers in that season. Hence, (ii) and (iii) cannot be inferred from this graph.

Statement (iv): A 100% stacked bar chart always shows each bar as '100%' of that garden's total. It does not tell us whether the totals for the two gardens are equal or not. Many different total numbers can give the same 100% bar. Therefore, we also cannot say anything about the total over the whole year from this graph.

A 100% stacked bar chart helps us compare fractions/percentages inside each garden, but it does not tell us the actual numbers of flowers.

Two possible stacked bar charts are shown below for the 100% stacked bar chart we just saw.

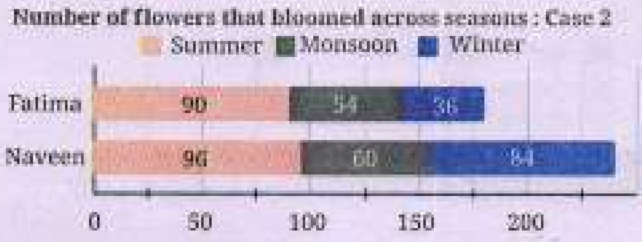
Number of flowers that bloomed across seasons: Case 1



$\frac{105}{210} \times 100 = 50\%$ $\frac{63}{210} \times 100 = 30\%$ $\frac{42}{210} \times 100 = 20\%$

$\frac{48}{120} \times 100 = 40\%$ $\frac{30}{120} \times 100 = 25\%$ $\frac{42}{120} \times 100 = 35\%$

Number of flowers that bloomed across seasons : Case 2



$\frac{90}{180} \times 100 = 50\%$ $\frac{54}{180} \times 100 = 30\%$ $\frac{36}{180} \times 100 = 20\%$

$\frac{96}{240} \times 100 = 40\%$ $\frac{60}{240} \times 100 = 25\%$ $\frac{84}{240} \times 100 = 35\%$

Notice that the percentage of flowers that bloomed in summer is more in Fatima's garden but this doesn't say anything about the actual number—Fatima's garden has more blooms than Naveen's in summer in Case 1 but her garden has fewer blooms than his in Case 2. Also, in Case 1 both Fatima's and Naveen's garden had 42 blooms but the respective percentages are different.

Think and Reflect

1. What does this say about the scope of the stacked bars and 100% stacked bars?
2. Given a stacked bar chart can we make a corresponding 100% stacked bar chart?
3. Given a 100% stacked bar chart can we make a corresponding stacked bar chart?
4. What kind of inferences or comparisons can be made from a stacked bar chart and in a 100% stacked bar chart?

A Typical Day

Example 9: The following chart shows the time spent on different activities in a day by the average Indian. What do you notice? What do you wonder about?

You may point out that you don't do much work and that your parents don't spend any time learning. Note that this strip is representative of people across ages, genders, regions, social/family backgrounds, etc., across India.

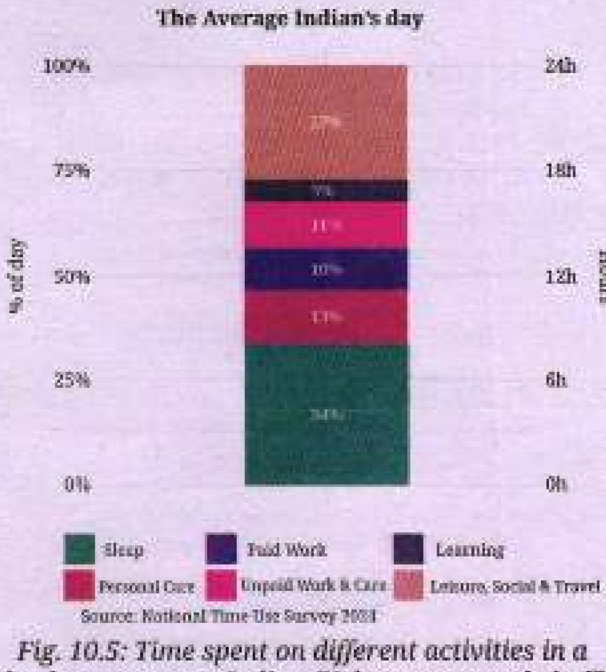


Fig. 10.5: Time spent on different activities in a day by an average Indian (Values are rounded off)

When we visualise the same data by grouping ages, we get the following stacked column chart:

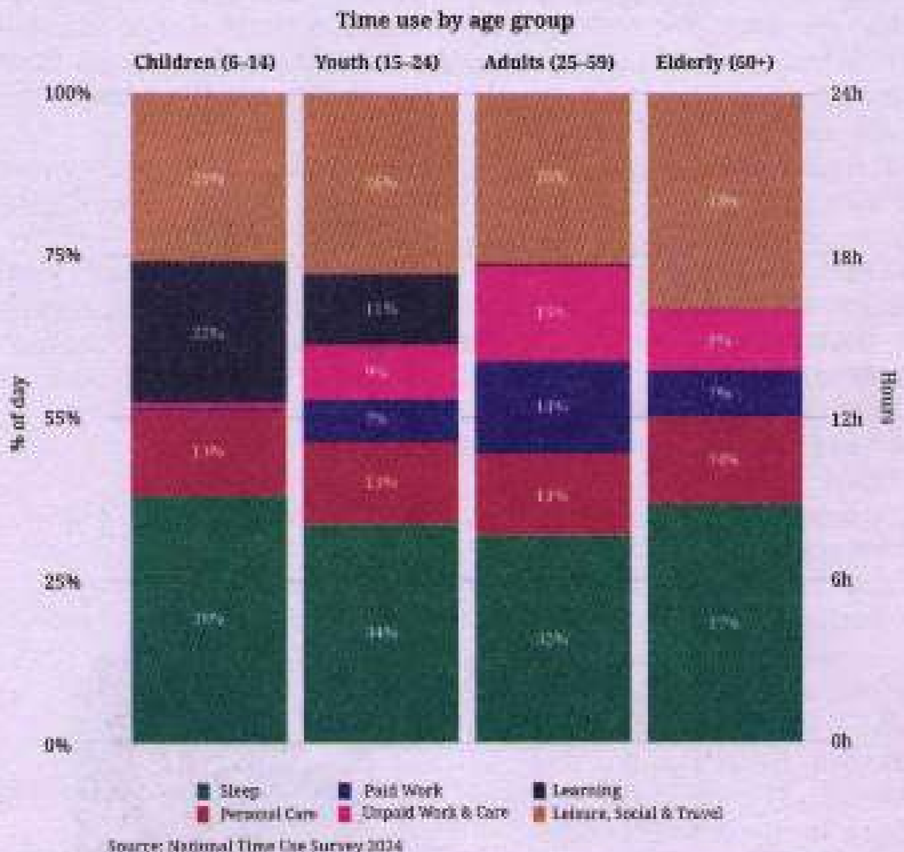


Fig. 10.6: Age-wise grouping of data (Values are rounded off)

Think and Reflect

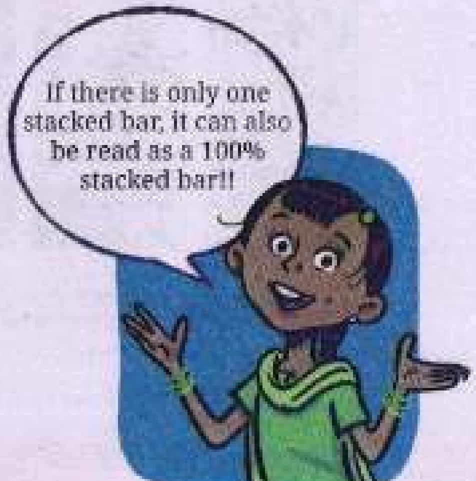
1. What do you find interesting in the chart above? What can you infer? Discuss.
2. Do you remember the sleep time over age trend that you studied last year? Does that trend align with this chart?
3. Can you explain why the learning time of people in the age group 15 – 24 has reduced significantly compared to that of the age group 6 – 14?
4. Do adults spend about an equal amount of time in paid and unpaid work? What do you think?

There might be a very small fraction of people whose day is similar/close to that of the average Indian. As we saw here, looking at only the average of a diverse collection often hides the differences and nuances within the data. We observed age-specific activity trends using this chart, which gave us a better idea. We could break these down further with respect to gender to examine any gender-specific trends within age groups.

You may wonder if there are differences across regions (e.g., rural and urban), or income, etc. You may be curious to look into the details—such as, what is included under personal care for different groups, or how does time spent on travel vary across cities and towns/villages? What we spend time on also depends on the times we live in. Clearly, a few decades ago, people didn't have access to mobile phones or televisions. Only time will tell what such charts will look like after a few decades. We may also ask—how do people in other countries spend their time? How similar or different are the trends?

What type of chart is this—a stacked bar chart or a 100% stacked bar chart?

It is both! It is a stacked bar because each small bar represents the amount of time spent on a certain activity. It is also a 100% stacked bar as each stacked bar's total is the same, i.e., 24 hours, and the lengths of the smaller bars within and across the stacks also represent the relative proportions of time spent.



Think and Reflect

Do you remember the 'What Can A Strip Say?' activity from last year? In each strip, if we club the tiny strips belonging to each activity together, will we get a 100% stacked bar chart like the one in Fig. 10.6?

Exercise Set 10.5

1. The smart watch data of 5 people was tracked from Monday to Friday. The recorded data tells us how many hours each person spent lying down, sitting, and standing or moving around, etc., as shown in the following table.

Table 10.3: Average number of hours spent per day (Monday – Friday) lying down, sitting, and standing or moving around, etc., of 5 people

	Lying down	Sitting	Standing or Moving around etc.
Pavani (Patient)	20	3	1
Raghu (Nurse)	7	4	13
Zakir (Teacher)	8	6	10
Sahana (Student)			
Julie (________)	8	10	6

(i) How much time does Sahana spend per day in each of these body-states? Make a reasonable guess and fill her row in the table.

(ii) Guess what activity or work Julie could be engaged in based on the data. Find out what your classmates' guesses are.

(iii) Complete the chart below based on the tabular data.

Think and Reflect

1. Will the bars look similar if a different teacher's data is considered instead of Zakir's?
2. Will the bars look different if the data for all 7 days of the week is considered?
3. Some bars in the chart may also match people doing other kinds of work/activities. Can you think of any? Discuss.

End-of-Chapter Exercises

1. In cricket, the **run rate** is the average number of runs scored per over. In a T20 match, a team scored 6 runs in the first over making the run rate 6.
   (i) In the second over they scored 12 runs. What is the run rate now?
   (ii) After Over 19, their run rate was 6. What is the run rate after 20 overs given the team made 12 runs in the last over?

2. Five friends who play badminton surveyed the city and collected the price of shuttlecocks across brands and types (Nylon and Feather). The following is the list of the prices, in rupees (₹), that each one compiled.

Nylon shuttle (N) Feather shuttle (F)

   Yusuf: 40(N), 105(N), 383(F), 108(N), 165(F), 116(F)
   Srikanth: 194(N), 85(N), 93(N), 121(N)
   Kashvi: 49(N), 297(F), 105(N), 275(F), 40(N)
   Prasanna: 124(N), 333(F), 182(N), 258(F)
   Gracy: 220(F), 458(F), 129(F), 183(N)

   Can you describe a way they can work together to find the average price of a shuttlecock across types?
   (i) Each one calculated the average of the prices they had gathered. Suppose these are a_y, a_s, a_k, a_p, a_g respectively. Write an expression that gives the combined average.
   (ii) Suppose a_n, a_f are the average prices of the nylon shuttlecocks and the feather shuttlecocks, respectively. Write an expression that gives the average price of a

shuttlecock. Will this be equal to the answer we got using the method in the previous part?

3. Shreyas holds 25 shares of a company at an average price of ₹150 and Vaishnavi holds 5 shares of the same company at an average price of ₹150.
   (i) Shreyas buys 10 shares of this company at a price of ₹30 each. What is the average price per share for him after the purchase?
   (ii) Vaishnavi buys some shares of this company at a price of ₹30 each and the average price per share for her after the purchase is ₹70. How many shares did she buy?

4. (Pṛthūdakasvāmi, commentary on Brahmagupta's Brāhmasphuṭasiddhānta, c. 864 CE) A "hasta" (meaning "forearm") refers to a unit of length measuring around 18 inches.
   A pool 30 hastas long is dug to different depths along its length. It is divided into 5 sections having lengths 4, 5, 6, 7, and 8 hastas, and is dug to depths of 9, 7, 7, 3, and 2 hastas, respectively. Find the mean depth of the pool.

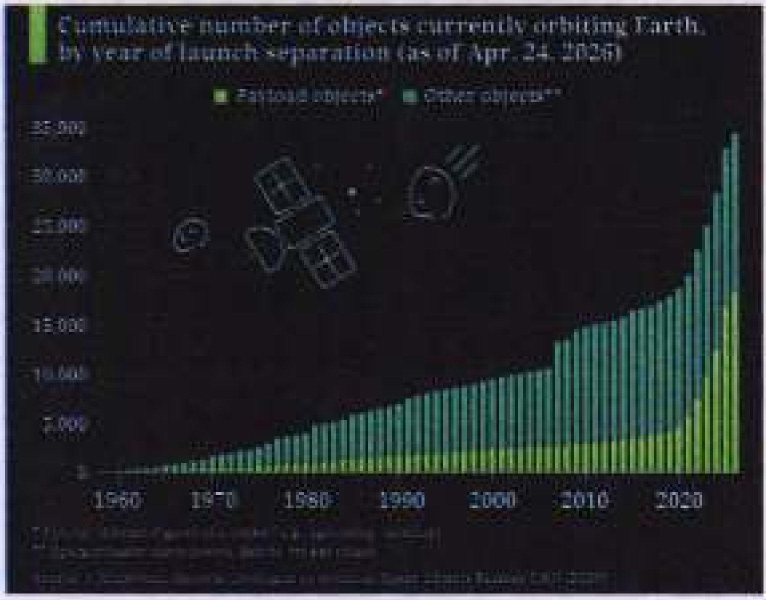
5. Suvarna had purchased 1g gold at ₹15k (short for ₹15,000). This is shown by point O denoting the average price of gold that she possesses. Six different scenarios are given below for her next transaction. For each scenario estimate and mark the average price of gold she will have after the transaction.

   (i) Purchase 1g gold at ₹30k
   (ii) Purchase 2g gold at ₹30k
   (iii) Purchase 1g gold at ₹10k
   (iv) Purchase 10g gold at ₹10k
   (v) Purchase 0.5g gold at ₹30k
   (vi) Purchase 0.5g gold at ₹15k

6. Given some data with corresponding weights, how would the weighted average change if all the weights are doubled? If required, experiment with some data. What do you observe? Justify your answer using algebra.

7. Answer the following questions based on the graph.
   (i) In 2003, approximately how many total objects were found orbiting Earth in space? In what year did this number double?
   (ii) Find the approximate number and share of payload objects and other objects in the year 2025.
   (iii) What can you say about the number of payload objects over time, and the number of other objects over time? What can

you say about the share of payload objects over time, and the share of other objects over time?



8. Observe the following infographic.

(i) Identify the correct inference(s) from the statements given.
  (a) More schools have a playground in Haryana compared to Uttarakhand.
  (b) Approximately every 2 out of 3 schools in Arunachal Pradesh have a playground.
  (c) Punjab has the highest number of schools with a playground.
  (d) Suppose it is given that Maharashtra has more schools than Telangana. Then the number of schools having a playground is more in Maharashtra.

(ii) Using the information given, can we find the nation-wide percentage of schools with a playground? If not, what additional information is needed?

9. Decision Dilemma:
  (i) Which of the plays would you choose to watch based on the following chart?

  (ii) The corresponding stacked bar chart is shown below. Would you change your decision after looking at this chart? Why/ Why not? Discuss.

10. A triathlon is an endurance race consisting of swimming 3.8 km, cycling 180 km, and running 42.2 km. Athletes compete for the fastest overall time, completing each segment in that order. The following table shows the finish time of 3 athletes in each segment.

Table 10.4: Finish time of three athletes in each segment of a triathlon race (values in hh : mm)

	Swimming	Cycling	Running
Athlete 1	01:08	05:00	03:15
Athlete 2	01:05	05:10	03:35
Athlete 3	01:22	05:55	03:50

(i) What is a suitable representation of this data — a stacked bar chart or a 100% stacked bar chart? Why do you think so?

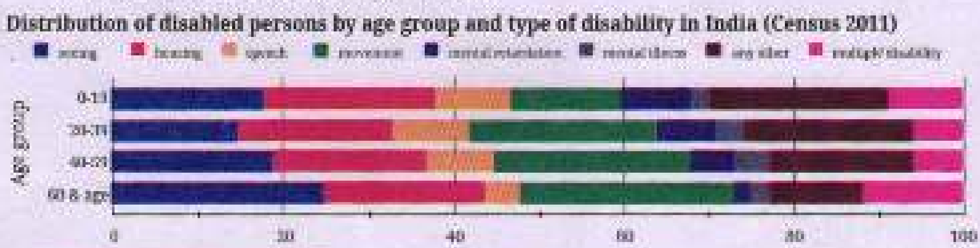
(ii) Suppose a 100% stacked bar chart is drawn. Which of the following question(s) can be answered by looking at just that chart?

(a) Who finished the race first?

(b) Approximately what fraction of their race time did Athlete 1 spend cycling?

(c) Which athlete took the longest for running?

11. Look at the following graph. What do you notice? What do you wonder? Write your inferences.



12. **Individual project:** Do at least one of the following.

(i) Recall the previous day and fill in the approximate time spent lying down, sitting, and standing/moving around, etc. Ask at least 2 of your family members about their day and fill it in. Alternatively, you can choose to track just your own body-state for any one of the weekdays, Saturday, and Sunday. Visualise it using a stacked bar chart. Discuss with your class how you arrived at the estimated time spent in each state.

(ii) Visualise your family's monthly expenditure using a 100% stacked bar chart.

(a) First, identify the expense categories in your family budget.

(b) Discuss with your family members to decide how you will collect and organise the data.

(c) Collect expense data for at least 3 months.

(d) Represent the data in a 100% stacked bar chart, with each bar showing how the total monthly expenditure is divided across the different categories.

(e) Finally, write your observations and inferences based on the chart.

13. **Small-group project:** Make a group of 3–4 students. Choose one scenario to design a custom rating scheme by assigning appropriate weights: (a) shopping experience at a cloth store, (b) travel experience in a bus, (c) tourism experience of a nearby tourist spot, (d) clinic/hospital experience

    (i) Discuss and arrive at 4–6 aspects to rate. For each aspect chosen, justify why it matters for an overall experience.

    (ii) Discuss and decide the relative weights and justify.

    (iii) Collect or imagine ratings from at least 10 people on a scale of 1–5 (5 being very good).

    (iv) Compute each individual rating. Compute the overall average across individuals.

    (v) Make a 100% stacked bar chart using your data.

    (vi) Write a short note on the observations and inferences, what your rating system captures well and what it misses.

14. **Whole class project:** Each student shares the average age of their family and the number of family members. Discuss among the class and come up with a way to find out the average age of all the families of the class.

*15. Given some data with corresponding weights, what would happen to the weighted average if all the weights are increased by a constant value, say 1? If required, experiment with some data. What do you observe? Justify your answer using algebra.

*16. A farm has some cows, sheep, and chickens. Last year the cows made up 60%, the sheep 25%, and the chickens 15%. There was a decrease in the number of all three animals' population over the year. Choose the possibilities for the change in their respective shares of the population —

(i) % of cows decreased, % of sheep decreased, % of chickens decreased

(ii) % of cows increased, % of sheep increased, % of chickens increased

(iii) % of cows remained the same, % of sheep remained the same, % of chickens remained the same

(iv) % of cows decreased, % of sheep increased, % of chickens remained the same

(v) % of cows increased, % of sheep increased, % of chickens decreased.

Chapter Summary

- The **weighted average** or **weighted mean** of $x_1, x_2, x_3, \ldots, x_n$ with weights $w_1, w_2, w_3, \ldots, w_n$ respectively is given by
$$x = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_n x_n}{w_1 + w_2 + w_3 + \cdots + w_n}.$$
- An average of averages can be computed using the weighted mean. The weighted mean can also be used to calculate the mean when certain values are added to or removed from data.
- The weighted average is also useful in calculating the concentration of mixtures.
- Whenever different parts of data have different number of values or levels of importance, the weighted mean can be used by assigning appropriate weights to each part of the data. Examples include rating systems and combining marks of different assessments to get a final score.
- **Stacked bar charts** may be used in place of cluster bar charts if the total of each cluster is also a matter of interest.
- **100% stacked bar charts**, like pie charts, compare proportions and not absolute values.
- Given a stacked bar chart, a corresponding 100% stacked bar chart can be made but not vice versa.