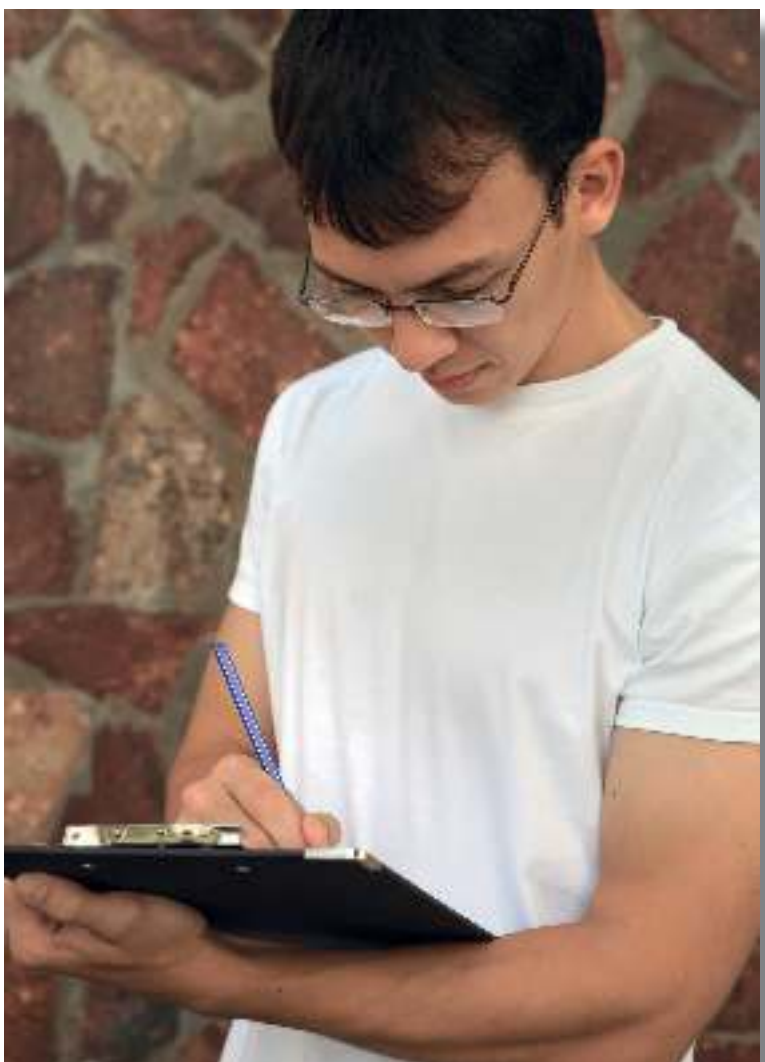
Chapter 4: Collecting, organising and displaying data

Key words

- Data
- Categorical data
- Qualitative
- Numerical data
- Quantitative
- Discrete
- Continuous
- Primary data
- Secondary data
- Frequency table
- Grouped
- Stem and leaf diagram
- Two-way table
- Pictogram
- Bar graph
- Pie chart
- Line graph

In this chapter you will learn how to:

- collect data and classify different types of data
- organise data using tally tables, frequency tables, stem and leaf diagrams and two-way tables
- draw pictograms, bar graphs, and pie charts to display data and answer questions about it.



This person is collecting information to find out whether people in his village know what government aid is available to them.

People collect information for many different reasons. We collect information to answer questions, make decisions, predict what will happen in the future, compare ourselves with others and understand how things affect our lives. A scientist might collect information from experiments or tests to find out how well a new drug is working. A businesswoman might collect data from business surveys to find out how well her business is performing. A teacher might collect test scores to see how well his students perform in an examination and an individual might collect data from magazines or the internet to decide which brand of shoes, jeans, make-up or car to buy. The branch of mathematics that deals with collecting data is called statistics. At this level, you will focus on asking questions and then collecting information and organising or displaying it so that you can answer questions.

RECAP

You should already be familiar with the following concepts from working with data:

Types of data and methods of collecting data

- Primary data – collected by the person doing the investigation.
- Secondary data – collected and stored by someone else (and accessed for an investigation).
- Data can be collected by experiment, measurement, observation or carrying out a survey.

Ways of organising and displaying data

Score	1	2	3	4	5	6
Frequency	3	4	3	5	2	3

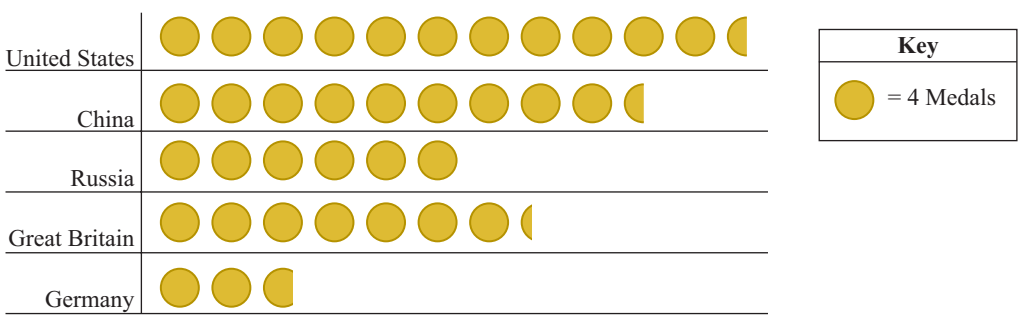
Frequency table (ungrouped data)

Amount spent ($)	Frequency
0 – 9.99	34
10 – 10.99	12
20 – 19.99	16
30 – 29.99	9

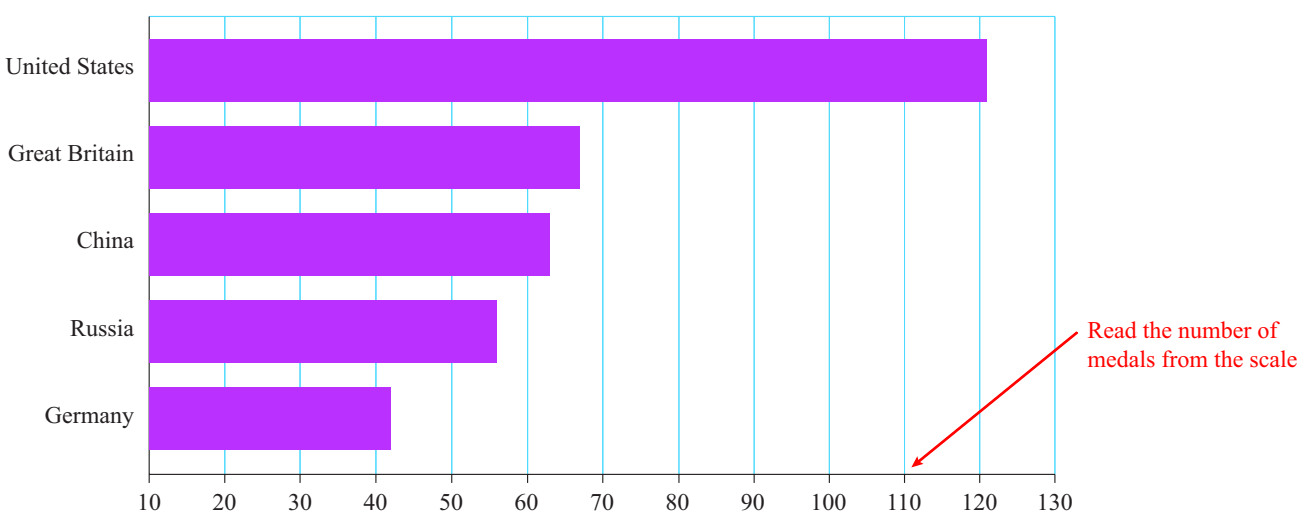
Number of data in that group, not individual values.

Class intervals are equal and should not overlap.

Frequency table (grouped data)

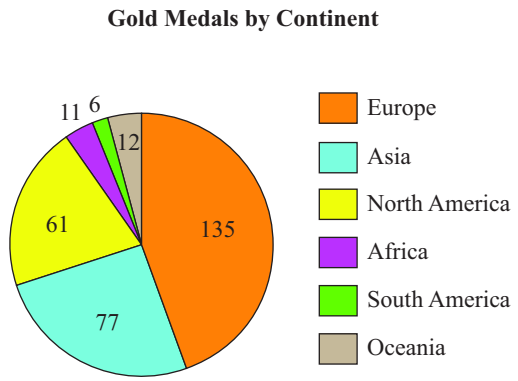


Pictogram – used mostly for visual appeal and effect

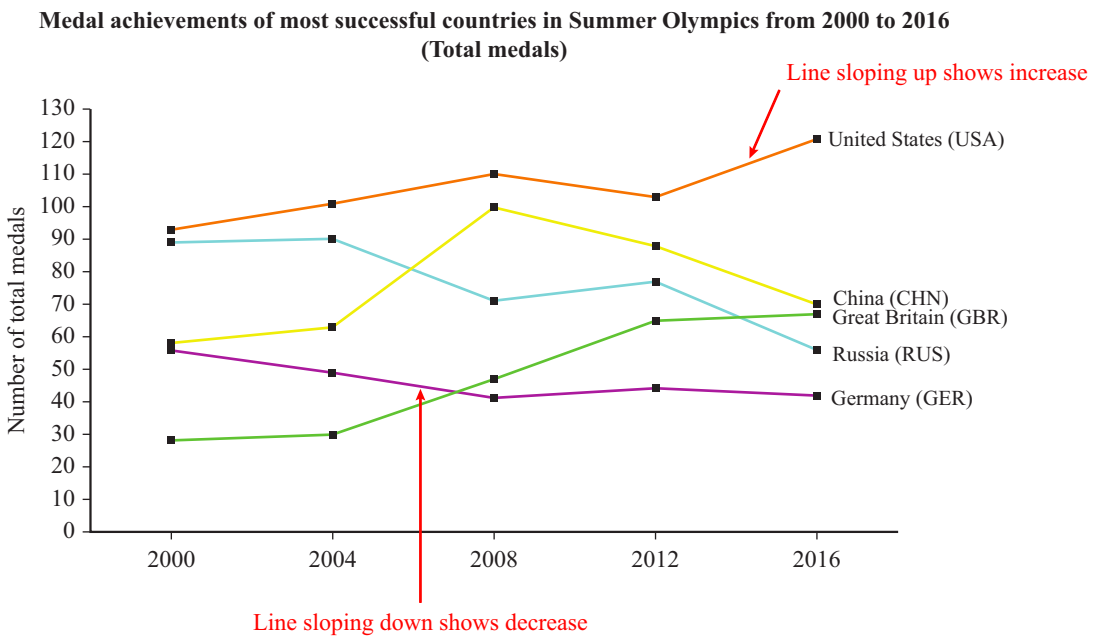


Bar charts – useful for discrete data in categories

RECAP



Pie charts – useful for comparing categories in the data set



Line graphs – useful for numerical data that shows changes over time

Graphs can be misleading. When you look at a graph think about:

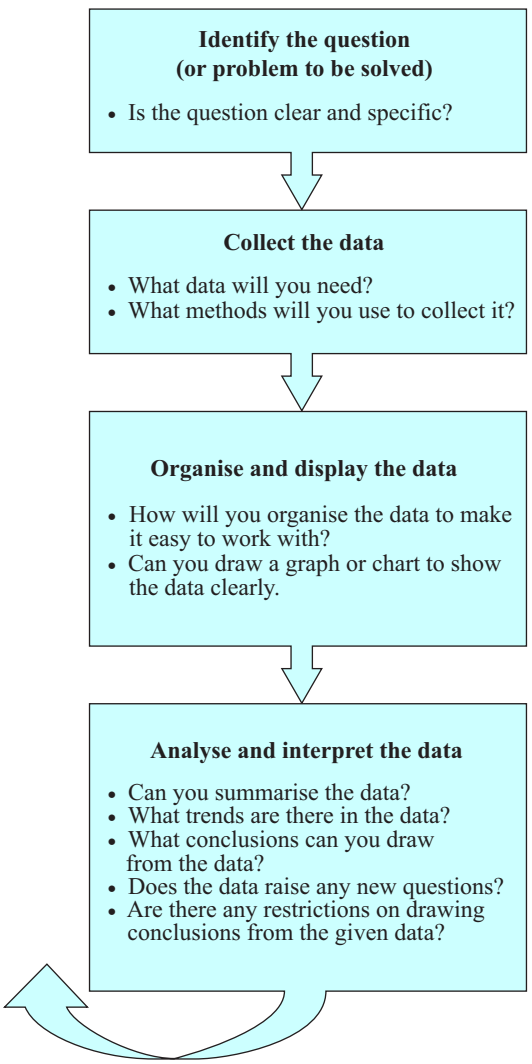
- The scale. The frequency axis should start at 0, it should not be exaggerated and it should be clearly labelled. Intervals between numbers should be the same.
- How it is drawn. Bars or sections of a pie chart that are 3-dimensional can make some parts look bigger than others and give the wrong impression of the data.

4.1 Collecting and classifying data

Data is actually the plural of the Latin word datum, but in modern English the word data is accepted and used as a singular form, so you can talk about a set of data, this data, two items of data or a lot of data.

Data is a set of facts, numbers or other information. Statistics involves a process of collecting data and using it to try and answer a question. The flow diagram shows the four main steps involved in this process of statistical investigation:

The information that is stored on a computer hard-drive or CD is also called data. In computer terms, data has nothing to do with statistics, it just means stored information.



LINK

All of this work is very important in biology and psychology, where scientists need to present data to inform their conclusions.

Different types of data

Answer these two questions:

- Who is your favourite singer?
- How many brothers and sisters do you have?

Your answer to the first question will be the name of a person. Your answer to the second question will be a number. Both the name and the number are types of data.

Categorical data is non-numerical data. It names or describes something without reference to number or size. Colours, names of people and places, yes and no answers, opinions and choices are all categorical. Categorical data is also called **qualitative** data.

Numerical data is data in number form. It can be an amount, a measurement, a time or a score. Numerical data is also called **quantitative** data (from the word quantity).

Numerical data can be further divided into two groups:

- **discrete data** – this is data that can only take certain values, for example, the number of children in a class, goals scored in a match or red cars passing a point. When you count things, you are collecting discrete data.
- **continuous data** – this is data that could take any value between two given values, for example, the height of a person who is between 1.5 m and 1.6 m tall could be 1.5 m, 1.57 m, 1.5793 m, 1.5793421 m or any other value between 1.5 m and 1.6 m depending on the degree of accuracy used. Heights, masses, distances and temperatures are all examples of continuous data. Continuous data is normally collected by measuring.

FAST FORWARD

You will need to fully understand continuous data when you study histograms in chapter 20. ▶

One way to decide if data is continuous is to ask whether it is possible for the values to be fractions or decimals. If the answer is yes the data is usually continuous. But be careful:

- age may seem to be discrete, because it is often given in full years, but it is actually continuous because we are getting older all the time
- shoe sizes are discrete, even though you can get shoes in half sizes, because you cannot get shoes in size $7\frac{1}{4}$ or $7\frac{3}{4}$ or $7\frac{8}{9}$.

Methods of collecting data

Data can be collected from primary sources by doing surveys or interviews, by asking people to complete questionnaires, by doing experiments or by counting and measuring. Data from primary sources is known as **primary data**.

Data can also be collected from secondary sources. This involves using existing data to find the information you need. For example, if you use data from an internet site or even from these pages to help answer a question, to you this is a secondary source. Data from secondary sources is known as **secondary data**.

Exercise 4.1

1 Copy this table into your book.

Categorical data	Numerical data
Hair colour	*Number of brothers and sisters*

a Add five examples of categorical data and five examples of numerical data that could be collected about each student in your class.
b Look at the numerical examples in your table. Circle the ones that will give discrete data.

LINK

In 2016, a leading financial magazine listed data scientist as the best paying and most satisfying job for the foreseeable future. The use of computers in data collection and processing has meant that data collection, display and analysis have become more and more important to business and other organisations.

2 State whether the following data would be discrete or continuous.

a Mass of each animal in a herd.
b Number of animals per household.
c Time taken to travel to school.
d Volume of water evaporating from a dam.
e Number of correct answers in a spelling test.
f Distance people travel to work.
g Foot length of each student in a class.
h Shoe size of each student in a class.
i Head circumference of newborn babies.
j Number of children per family.
k Number of TV programmes watched in the last month.
l Number of cars passing a zebra crossing per hour.

3 For each of the following questions state:

i one method you could use to collect the data
ii whether the source of the data is primary or secondary
iii whether the data is categorical or numerical
iv If the data is numerical, state whether it is discrete or continuous.

a How many times will you get a six if you throw a dice 100 times?
b Which is the most popular TV programme among your classmates?
c What are the lengths of the ten longest rivers in the world?

d What is the favourite sport of students in your school?
e How many books are taken out per week from the local library?
f Is it more expensive to drive to work than to use public transport?
g Is there a connection between shoe size and height?
h What is the most popular colour of car?
i What is the batting average of the national cricket team this season?
j How many pieces of fruit do you eat in a week?

4.2 Organising data

FAST FORWARD

You will use these methods and extend them in later chapters. Make sure that you understand them now. ▶

When you collect a large amount of data you need to organise it in some way so that it becomes easy to read and use. Tables (tally tables, frequency tables and two-way tables) are the most commonly used methods of organising data.

Tally tables

Tallies are little marks (////) that you use to keep a record of items you count. Each time you count five items you draw a line across the previous four tallies to make a group of five (𝍸). Grouping tallies in fives makes it much easier to count and get a total when you need one.

A tally table is used to keep a record when you are counting things.

Look at this tally table. A student used this to record how many cars of each colour there were in a parking lot. He made a tally mark in the second column each time he counted a car of a particular colour.

Colour	Number of cars
White	𝍸 𝍸 ///
Red	𝍸 𝍸 𝍸 𝍸 /
Black	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //
Blue	𝍸 𝍸 𝍸 𝍸 𝍸 //
Silver	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///
Green	𝍸 𝍸 𝍸 /

(The totals for each car are shown after Exercise 4.2 on page 78.)

Worked example 1

Anita wanted to find out what people thought about pop-up adverts on their social media feeds. She did a survey of 100 people. Each person chose an answer A, B C or D.

What do you think about this statement? Please choose one response.
Advertising should be strictly controlled on social media. Pop-up adverts should be banned from all social media feeds.
A I strongly agree
B I agree
C I disagree
D I strongly disagree

She recorded these results:

A	B	A	C	A	C	C	D	A	C
C	C	D	A	D	D	C	C	C	A
B	B	A	C	D	B	B	A	C	C
A	B	C	A	D	B	C	D	A	B
A	C	C	D	A	C	C	C	D	A
D	D	C	C	C	A	B	B	A	C
D	C	C	D	A	C	A	B	D	B
C	C	D	A	D	D	C	C	C	A
B	B	A	C	D	B	B	C	C	C
A	B	C	A	D	B	C	D	A	B

a Draw a tally table to organise the results.

b What do the results of her survey suggest people think about pop-up advertising on social media?

a

Response	Tally
A	𝍸 𝍸 𝍸 𝍸 ////
B	𝍸 𝍸 𝍸 ////
C	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //
D	𝍸 𝍸 𝍸 𝍸

Count each letter. Make a tally each time you count one.
It may help to cross the letters off the list as you count them.
Check that your tallies add up to 100 to make sure you have included all the scores. (You could work across the rows or down the columns, putting a tally into the correct row in your table, rather than just counting one letter at a time.)

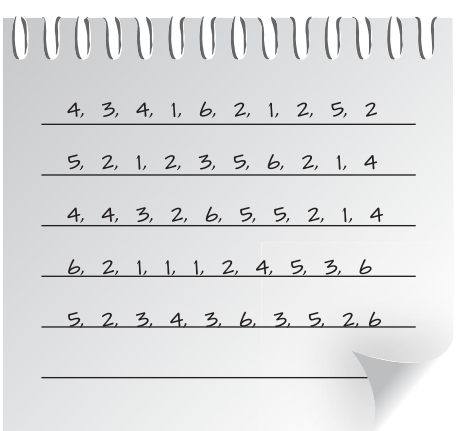
b The results suggest that people generally don't think advertising should be banned on social media. 57 people disagreed or strongly disagreed. Only 24 of the 100 people strongly agreed with Anita's statement.

LINK

By giving people a very definite statement and asking them to respond to it, Anita has shown her own bias and that could affect the results of her survey. It is quite possible that people feel some control is necessary, but not that adverts should be banned completely and they don't have that as an option when they answer. The composition of the sample could also affect the responses, so any conclusions from this survey would need to be considered carefully. You will deal with restrictions on drawing conclusions in more detail in Chapter 12.

Exercise 4.2

1 Balsem threw a dice 50 times. These are her scores. Draw a tally table to organise her data.



2 Do a quick survey among your class to find out how many hours each person usually spends doing his or her homework each day. Draw your own tally table to record and organise your data.

3 Faizel threw two dice together 250 times and recorded the score he got using a tally table. Look at the tally table and answer the questions about it.

Score	Tally
2	𝍸 //
3	𝍸 𝍸 𝍸
4	𝍸 𝍸 𝍸 𝍸 ///
5	𝍸 𝍸 𝍸 𝍸 𝍸 ////
6	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///
7	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 /
8	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸
9	𝍸 𝍸 𝍸 𝍸 𝍸 ///
10	𝍸 𝍸 𝍸 𝍸 /
11	𝍸 𝍸 //
12	𝍸 /

a Which score occurred most often?
b Which two scores occurred least often?
c Why do you think Faizel left out the score of one?
d Why do you think he scored six, seven and eight so many times?

Frequency tables

A **frequency table** shows the totals of the tally marks. Some frequency tables include the tallies.

This frequency table is the same as the tally table the student used to record car colours (page 76). It has another column added with the totals (frequencies) of the tallies.

Colour	Number of cars	Frequency
White	𝍸 𝍸 ///	13
Red	𝍸 𝍸 𝍸 𝍸 /	21
Black	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	37
Blue	𝍸 𝍸 𝍸 𝍸 𝍸 //	27
Silver	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///	43
Green	𝍸 𝍸 𝍸 /	16
	Total	157

Before you could draw any meaningful conclusions about what type of illness is most common at a clinic, you would need to know where this data was collected. The frequency of different diseases would be different in different parts of the world.

The frequency column tells you how often (how frequently) each result appeared in the data and the data is discrete.

The frequency table has space to write a total at the bottom of the frequency column. This helps you to know how many pieces of data were collected. In this example the student recorded the colours of 157 cars.

Most frequency tables will not include tally marks. Here is a frequency table without tallies. It was drawn up by the staff at a clinic to record how many people were treated for different diseases in one week.

Illness	Frequency
Diabetes	30
HIV/Aids	40
TB	60
Other	50
Total	180

Grouping data in class intervals

FAST FORWARD

You will soon use these tables to construct bar charts and other frequency diagrams. These diagrams give a clear, visual impression of the data. ▶

In this example, the test does not allow for fractions of a mark, so all test scores are integers and the data is discrete.

Sometimes numerical data needs to be recorded in different groups. For example, if you collected test results for 40 students you might find that students scored between 40 and 84 (out of 100). If you recorded each individual score (and they could all be different) you would get a very large frequency table that is difficult to manage. To simplify things, the collected data can be arranged in groups called class intervals. A frequency table with results arranged in class intervals is called a **grouped** frequency table. Look at the example below:

Points scored	Frequency
40–44	7
45–49	3
50–54	3
55–59	3
60–64	0
65–69	5
70–74	3
75–79	7
80–84	9
Total	40

The range of scores (40–84) has been divided into class intervals. Notice that the class intervals do not overlap so it is clear which data goes in what class.

Exercise 4.3

1 Sheldon did a survey to find out how many coins the students in his class had on them (in their pockets or purses). These are his results:

0	2	3	1	4	6	3	6	7	2
1	2	4	0	0	6	5	4	8	2
6	3	2	0	0	0	2	4	3	5

a Copy this frequency table and use it to organise Sheldon's data.

Number of coins	0	1	2	3	4	5	6	7	8
Frequency									

b What is the highest number of coins that any person had on them?

c How many people had only one coin on them?

d What is the most common number of coins that people had on them?
e How many people did Sheldon survey altogether? How could you show this on the frequency table?

2 Penny works as a waitress in a fast food restaurant. These are the amounts (in dollars) spent by 25 customers during her shift.

Note that currency (money) is discrete data because you cannot get a coin (or note) smaller than one cent.

43.55	4.45	17.60	25.95	3.75
12.35	55.00	12.90	35.95	16.25
25.05	2.50	29.35	12.90	8.70
12.50	13.95	6.50	39.40	22.55
20.45	4.50	5.30	15.95	10.50

a Copy and complete this grouped frequency table to organise the data.

Amount ($)	0–9.99	10–19.99	20–29.99	30–39.99	40–49.99	50–59.99
Frequency						

b How many people spent less than $20.00?
c How many people spent more than $50.00?
d What is the most common amount that people spent during Penny's shift?

3 Leonard records the length in minutes and whole seconds, of each phone call he makes during one day. These are his results:

3 min 29 s	4 min 12 s	4 min 15 s	1 min 29 s	2 min 45 s
1 min 32 s	1 min 09 s	2 min 50 s	3 min 15 s	4 min 03 s
3 min 04 s	5 min 12 s	5 min 45 s	3 min 29 s	2 min 09 s
1 min 12 s	4 min 15 s	3 min 45 s	3 min 59 s	5 min 01 s

Use a grouped frequency table to organise the data.

Stem and leaf diagrams

FAST FORWARD

You will work with stem and leaf diagrams again when you calculate averages and measures of spread in chapter 12. ▶

A **stem and leaf diagram** is a special type of table that allows you to organise and display grouped data using the actual data values. When you use a frequency table to organise grouped data you cannot see the actual data values, just the number of data items in each group. Stem and leaf diagrams are useful because when you keep the actual values, you can calculate the range and averages for the data.

In a stem and leaf diagram each data item is broken into two parts: a stem and a leaf. The final digit of each value is the leaf and the previous digits are the stem. The stems are written to the left of a vertical line and the leaves are written to the right of the vertical line. For example a score of 13 would be shown as:

Stem		Leaf
1	|	3

In this case, the tens digit is the stem and the units digit is the leaf.

A larger data value such as 259 would be shown as:

Stem		Leaf
25	|	9

In this case, the stem represents both the tens and the hundreds digits while the units digit is the leaf.

To be useful, a stem and leaf diagram should have at least 5 stems. If the number of stems is less than that, you can split the leaves into 2 (or sometimes even 5) classes. If you do this, each stem is listed twice and the leaves are grouped into a lower and higher class. For example, if the stem is tens and the leaves are units, you would make two classes like this:

Stem		Leaf
1	|	0 3 4 2 1
1	|	5 9 8 7 5 6

Values from 10 to 14 (leaves 0 to 4) are included in the first class, values from 15 to 19 (leaves 5 to 9) are included in the second class.

Stem and leaf diagrams are easier to work with if the leaves are ordered from smallest to greatest.

Worked example 2

This data set shows the ages of customers using an internet café.

34	23	40	35	25	28	18	32
37	29	19	17	32	55	36	42
33	20	25	34	48	39	36	30

Draw a stem and leaf diagram to display this data.

Stem	Leaf
1	8 9 7
2	3 5 8 9 0 5
3	4 5 2 7 6 6 3 4 2 9 0
4	0 2 8
5	5

Key
1 | 7 = 17 years old

Group the ages in intervals of ten, 10 – 19; 20 – 29 and so on.

These are two-digit numbers, so the tens digit will be the stem.

List the stems in ascending order down the left of the diagram.

Work through the data in the order it is given, writing the units digits (the leaves) in a row next to the appropriate stem. Space the leaves to make it easier to read them.

If you need to work with the data, you can redraw the diagram, putting the leaves in ascending order.

Stem	Leaf
1	7 8 9
2	0 3 5 5 8 9
3	0 2 2 3 4 4 5 6 6 7 9
4	0 2 8
5	5

Key
1 | 7 = 17 years old

From this re-organised stem and leaf diagram you can quickly see that:

- the youngest person using the internet café was 17 years old (the first data item)
- the oldest person was 55 (the last data item)
- most users were in the age group 30 – 39 (the group with the largest number of leaves).

A back to back stem and leaf diagram is used to show two sets of data. The second set of data is plotted against the same stem, but the leaves are written to the left.

This stem and leaf plot compares the battery life of two different brands of mobile phone.

Brand X Leaf	Stem	Brand Y Leaf
9 4 8 7 2	0	5 8
7 8 7 2 3	1	4 7 8 2
8 4 6 2 7 9 8	2	8 9 7 1 5
7 2	3	7 2 1 0
	4	2
	5	1

Key
Brand X 8 | 2 = 28 hours Brand Y 4 | 2 = 42 hours

You read the data for Brand X from right to left. The stem is still the tens digit.

Exercise 4.4

1 The mass of some Grade 10 students was measured and recorded to the nearest kilogram. These are the results:

45 56 55 68 53 55 48 49 53 54
56 59 60 63 67 49 55 56 58 60

Construct a stem and leaf diagram to display the data.

2 The numbers of pairs of running shoes sold each day for a month at different branches of 'Runner's Up Shoe Store' are given below.

Branch A	175, 132, 180, 134, 179, 115, 140, 200, 198, 201, 189, 149, 188, 179, 186, 152, 180, 172, 169, 155, 164, 168, 166, 149, 188, 190, 199, 200
Branch B	188, 186, 187, 159, 160, 188, 200, 201, 204, 198, 190, 185, 142, 188 165, 187, 180, 190, 191, 169, 177, 200, 205, 196, 191, 193, 188, 200

a Draw a back to back stem and leaf diagram to display the data.
b Which branch had the most sales on one day during the month?
c Which branch appears to have sold the most pairs? Why?

3 A biologist wanted to investigate how pollution levels affect the growth of fish in a dam. In January, she caught a number of fish and measured their length before releasing them back into the water. The stem and leaf diagram shows the lengths of the fish to the nearest centimetre.

Length of fish (cm) January sample

1	2 4 4 6
2	0 1 3 3 4 5 8 9
3	3 5 6 6 6 7 8 9
4	0 2 5 7
5	2 7

Key
1 | 2 = 12 cm

a How many fish did she measure?
b What was the shortest length measured?
c How long was the longest fish measured?
d How many fish were 40 cm or longer?
e How do you think the diagram would change if she did the same survey in a year and:
  i the pollution levels had increased and stunted the growth of the fish
  ii the conditions in the water improved and the fish increased in length?

4 This stem and leaf diagram shows the pulse rate of a group of people measured before and after exercising on a treadmill.

Pulse rate

Before exercise Leaf	Stem	After exercise Leaf
0 1 3 6 8 7 2	6	
3 4 1 2	7	
7 3 2 7 8	8	7 6 4 3
0	9	0 2 4 1 3
1	10	3 1 7 8 9
	11	8 2
	12	7
	13	
	14	2

Key	
Before exercise	2 | 6 = 62 beats per minute
After exercise	8 | 7 = 87 beats per minute

a How many people had a resting pulse rate (before exercise) in the range of 60 to 70 beats per minute?
b What was the highest pulse rate measured before exercise?
c That person also had the highest pulse rate after exercise, what was it?
d What does the stem and leaf diagram tell you about pulse rates and exercise in this group? How?

Two-way tables

A **two-way table** shows the frequency of certain results for two or more sets of data. Here is a two way table showing how many men and woman drivers were wearing their seat belts when they passed a check point.

	Wearing a seat belt	Not wearing a seat belt
Men	10	4
Women	6	3

The headings at the top of the table give you information about wearing seat belts. The headings down the side of the table give you information about gender.

You can use the table to find out:

- how many men were wearing seat belts
- how many women were wearing seat belts
- how many men were not wearing seat belts
- how many women were not wearing seat belts.

You can also add the totals across and down to work out:

- how many men were surveyed
- how many women were surveyed
- how many people (men + women) were wearing seat belts or not wearing seat belts.

Here are two more examples of two-way tables:

Drinks and crisps sold at a school tuck shop during lunch break

	Sweet chilli	Plain	Cheese and onion
Cola	9	6	23
Fruit juice	10	15	12

How often male and female students use Facebook

	Never use it	Use it sometimes	Use it every day
Male	35	18	52
Female	42	26	47

Exercise 4.5

1 A teacher did a survey to see how many students in her class were left-handed. She drew up this two-way table to show the results.

	Left-handed	Right-handed
Girls	9	33
Boys	6	42

a How many left-handed girls are there in the class?
b How many of the girls are right-handed?
c Are the boys mostly left-handed or mostly right-handed?
d How many students are in the class?

2 Do a quick survey in your own class to find out whether girls and boys are left- or right-handed. Draw up a two-way table of your results.

3 Sima asked her friends whether they liked algebra or geometry best. Here are the responses.

Name	Algebra	Geometry
Sheldon		✓
Leonard	✓	
Raj	✓	
Penny		✓
Howard	✓	
Zarah		✓
Zohir		✓
Ahmed	✓	
Jenny	✓	

FAST FORWARD

Make sure you understand how to draw up and read a two-way table. You will use them again in chapter 8 when you deal with probability. ▶

Name	Algebra	Geometry
Priyanka		✓
Anne		✓
Ellen	✓	

a Draw a two-way table using these responses.
b Write a sentence to summarise what you can learn from the table.

Two-way tables in everyday life

Two-way tables are often used to summarise and present data in real life situations. You need to know how to read these tables so that you can answer questions about them.

Worked example 3

This table shows world population data for 2008 with estimated figures for 2025 and 2050.

Region	Population in 2008	Projected population 2025	Projected population 2050
World	6 705 000 000	8 000 000 000	9 352 000 000
Africa	967 000 000	1 358 000 000	1 932 000 000
North America	338 000 000	393 000 000	480 000 000
Latin America and the Caribbean	577 000 000	678 000 000	778 000 000
Asia	4 052 000 000	4 793 000 000	5 427 000 000
Europe	736 000 000	726 000 000	685 000 000
Oceania	35 000 000	42 000 000	49 000 000

(Data from Population Reference Bureau.)

a What was the total population of the world in 2008?
b By how much is the population of the world expected to grow by 2025?
c What percentage of the world's population lived in Asia in 2008? Give your answer to the closest whole per cent.
d Which region is likely to experience a decrease in population between 2008 and 2025?
  i What is the population of this region likely to be in 2025?
  ii By how much is the population expected to decrease by 2050?

a	6 705 000 000	Read this from the table.
b	8 000 000 000 − 6 705 000 000 = 1 295 000 000	Read the value for 2025 from the table and subtract the smaller figure from the larger.
c	$\frac{4\,052\,000\,000}{6\,705\,000\,000} \times 100 = 60.4325\% \approx 60\%$	Read the figures from the table and then calculate the percentage.
d	Europe **i** 726 000 000 **ii** 736 000 000 − 685 000 000 = 51 000 000	Look to see which numbers are decreasing across the row. Read this from the table. Read the values from the table and subtract the smaller figure from the larger.

Exercise 4.6 *Applying your skills*

This distance table shows the flying distance (in miles) between some major world airports.

	Mumbai	Hong Kong	London	Montreal	Singapore	Sydney
Dubai	1199	3695	3412	6793	3630	7580
Hong Kong	2673		8252	10 345	1605	4586
Istanbul	2992	7016	1554	5757	5379	11 772
Karachi	544	3596	5276	8888	2943	8269
Lagos	5140	8930	3098	6734	7428	11 898
London	4477	8252		3251	6754	10 564
Singapore	2432	1605	6754	9193		3912
Sydney	6308	4586	10 564	12 045	3916	

a Find the flying distance from Hong Kong to:

i Dubai **ii** London **iii** Sydney

b Which is the longer flight: Istanbul to Montreal or Mumbai to Lagos?

c What is the total flying distance for a return flight from London to Sydney and back?

d If the plane flies at an average speed of 400 miles per hour, how long will it take to fly the distance from Singapore to Hong Kong to the nearest hour?

e Why are there some blank blocks on this table?

4.3 Using charts to display data

FAST FORWARD

You also need to be able to draw and use frequency distributions and histograms. These are covered in chapter 20. ▶

Charts are useful for displaying data because you can see patterns and trends easily and quickly. You can also compare different sets of data easily. In this section you are going to revise what you already know about how to draw and make sense of pictograms, bar charts and pie charts.

Pictograms

Pictograms are fairly simple charts. Small symbols (pictures) are used to represent quantities. The meaning of the symbol and the amount it represents (a 'key') must be provided for the graph to make sense.

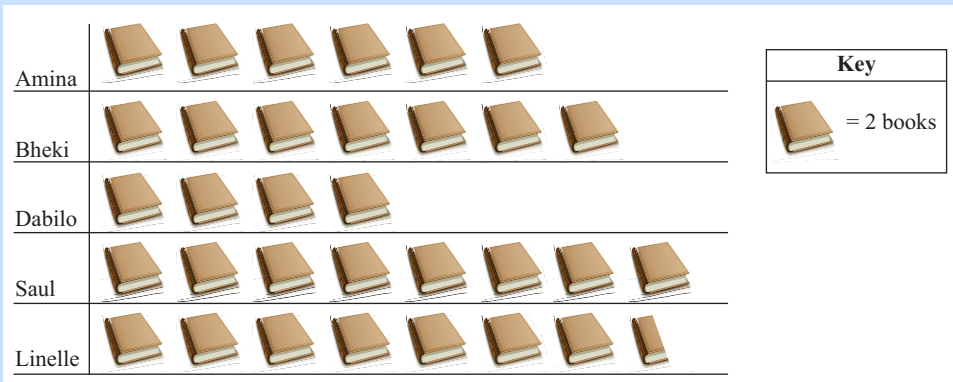
Worked example 4

The table shows how many books five different students have finished reading in the past year.

Student	Number of books read
Amina	12
Bheki	14
Dabilo	8
Saul	16
Linelle	15

Draw a pictogram to show this data.

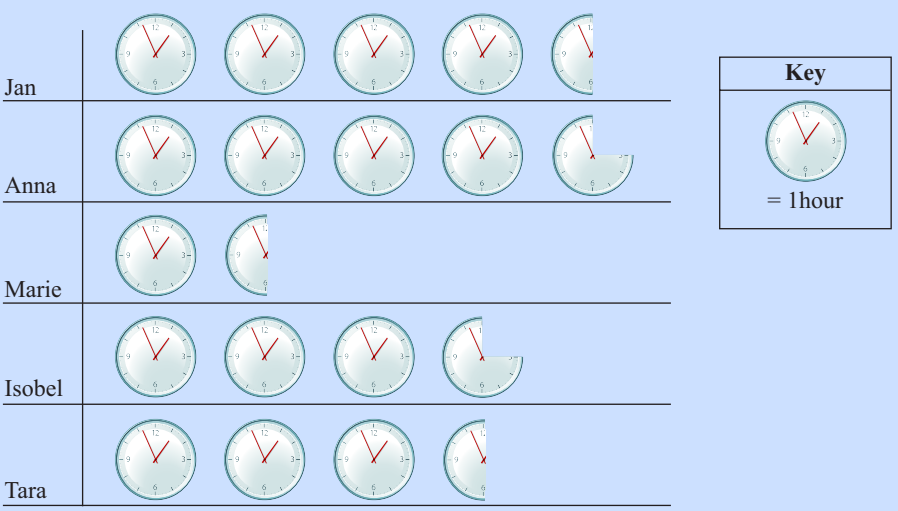
Number of books read



Worked example 5

This pictogram shows the amount of time that five friends spent talking on their phones during one week.

Times spent on the phone



a Who spent the most time on the phone that week?
b How much time did Isobel spend on the phone that week?
c Who spent $3\frac{1}{2}$ hours on the phone this week?
d Draw the symbols you would use to show $2\frac{1}{4}$ hours.

a	Anna	The person with the most clocks.
b	$3\frac{3}{4}$ hours	There are three whole clocks; the key shows us each one stands for 1 hour. The fourth clock is only three-quarters, so it must be $\frac{3}{4}$ of an hour.
c	Tara	She has three full clocks, each worth 1 hour, and one half clock.
d		Two full clocks to represent two hours, and a quarter of a clock to represent $\frac{1}{4}$ hours.

Exercise 4.7

1 A pictogram showing how many tourists visit the top five tourist destinations uses this symbol.

How many tourists are represented by each of these symbols?

a 

b 

c 

d 

The number of arrivals represented by the key should be an integer that is easily divided into the data; you may also need to round the data to a suitable degree of accuracy.

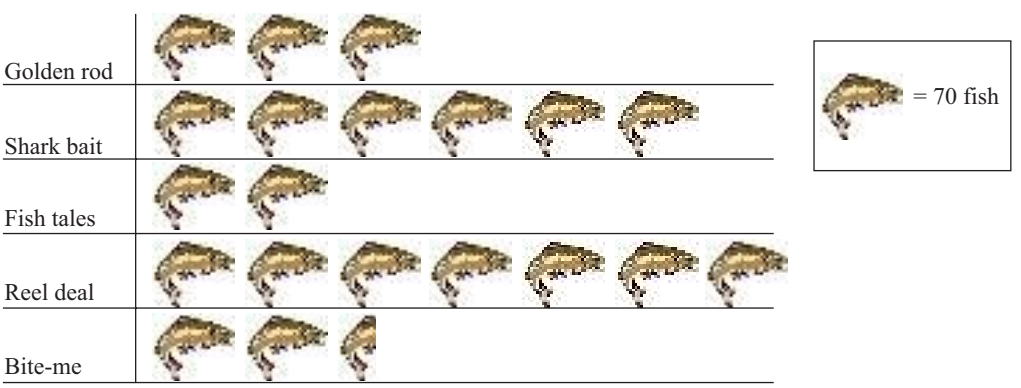
2 Here is a set of data for the five top tourist destination countries (2016). Use the symbol from question **1** with your own scale to draw a pictogram to show this data.

Most tourist arrivals

Country	France	USA	Spain	China	Italy
Number of tourists	84 500 000	77 500 000	68 200 000	56 900 000	50 700 000

3 This pictogram shows the number of fish caught by five fishing boats during one fishing trip.

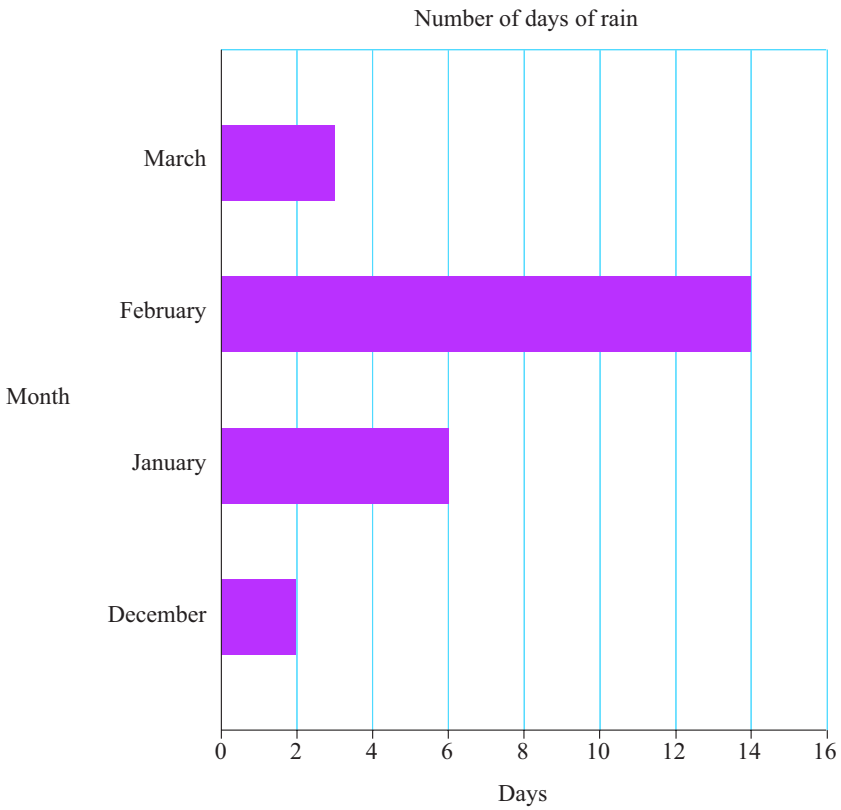
Number of fish caught per boat

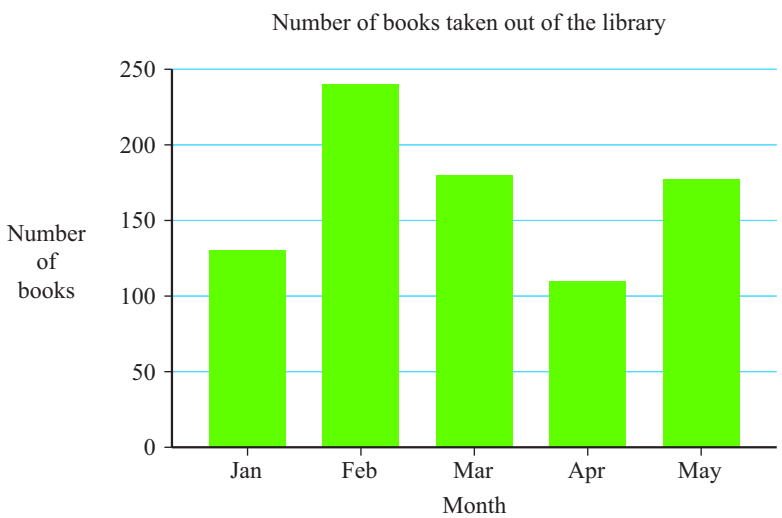


a Which boat caught the most fish?
b Which boat caught the least fish?
c How many fish did each boat catch?
d What is the total catch for the fleet on this trip?

Bar charts

Bar charts are normally used to display discrete data. The chart shows information as a series of bars plotted against a scale on the axis. The bars can be horizontal or vertical.





The bars should not touch for qualitative or discrete data.

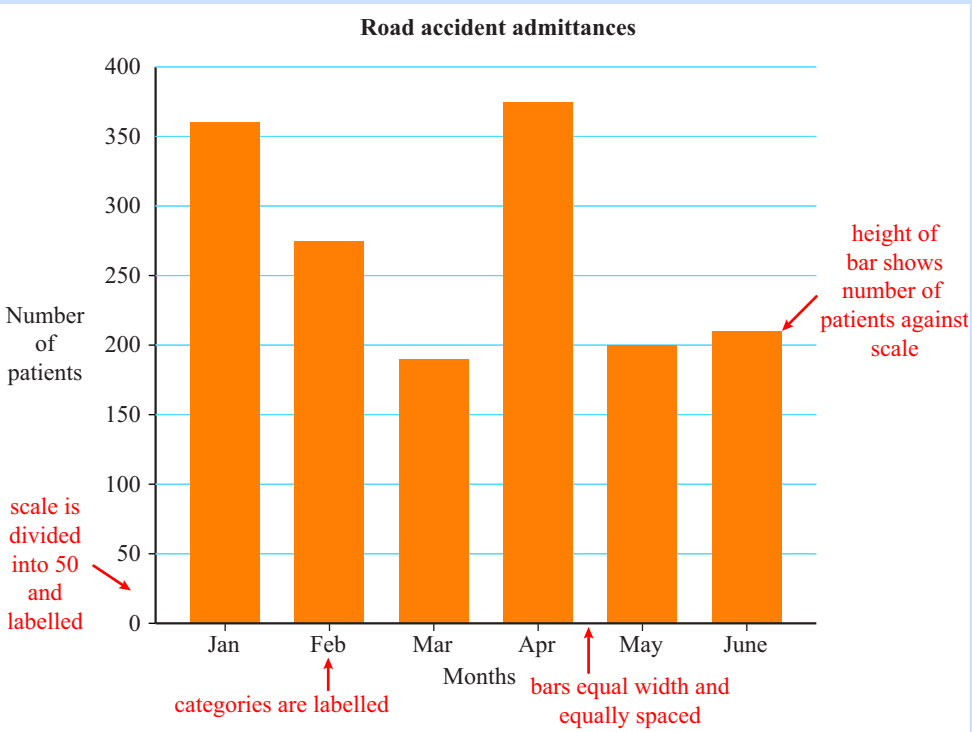
There are different methods of drawing bar charts, but all bar charts should have:

- a title that tells what data is being displayed
- a number scale or axis (so you can work out how many are in each class) and a label on the scale that tells you what the numbers stand for
- a scale or axis that lists the categories displayed
- bars that are equally wide and equally spaced.

Worked example 6

The frequency table shows the number of people who were treated for road accident injuries in the casualty department of a large hospital in the first six months of the year. Draw a bar chart to represent the data. Note that bar chart's frequency axes should start from zero.

Patients admitted as a result of road accidents	
Month	**Number of patients**
January	360
February	275
March	190
April	375
May	200
June	210

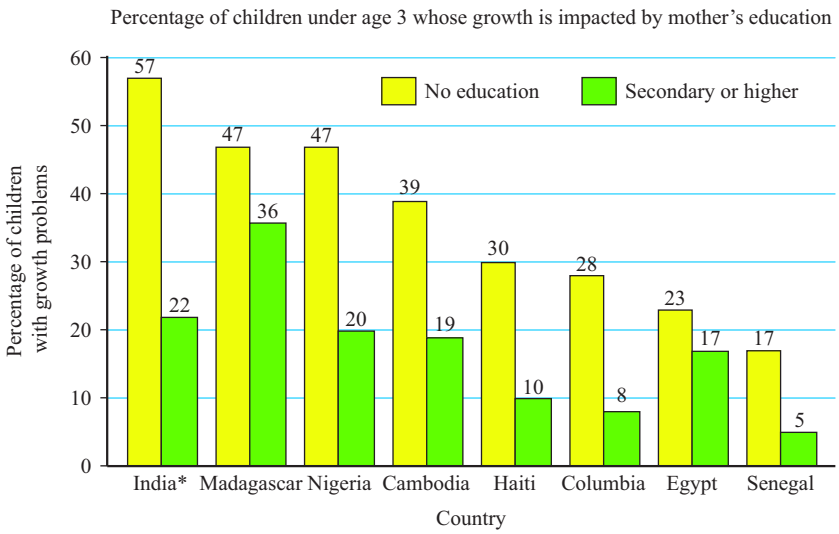


FAST FORWARD

A bar chart is not the same as a histogram. A histogram is normally used for continuous data. You will learn more about histograms in chapter 20. ▶

Compound bar charts

A compound bar chart displays two or more sets of data on the same set of axes to make it easy to compare the data. This chart compares the growth rates of children born to mothers with different education levels.



(Adapted from Nutrition Update 2010: www.dhsprogram.com)

You can see that children born to mothers with secondary education are less likely to experience growth problems because their bars are shorter than the bars for children whose mothers have only primary education. The aim of this graph is to show that countries should pay attention to the education of women if they want children to develop in healthy ways.

Exercise 4.8 *Applying your skills*

1 Draw a bar chart to show each of these sets of data.

a

Favourite take-away food	Burgers	Noodles	Fried chicken	Hot chips	Other
No. of people	40	30	84	20	29

b

African countries with the highest HIV/AIDS infection rates (2015 est)	
Country	**% of adults (aged 15 to 49) infected**
Swaziland	28.8
Botswana	22.2
Lesotho	22.7
Zimbabwe	14.7
South Africa	19.2
Namibia	13.3
Zambia	12.3
Malawi	9.1
Uganda	7.1
Mozambique	10.5

(Data taken from www.aidsinfo.unaids.org)

HIV is a massive global health issue. In 2017, the organisation Avert reported that 36.7 million people worldwide were living with HIV. The vast majority of these people live in low- and middle-income countries and almost 70% of them live in sub-Saharan Africa. The countries of East and Southern Africa are the most affected. Since 2010, there has been a 29% decrease in the rate of new infection in this region, largely due to awareness and education campaigns and the roll out of anti-retroviral medication on a large scale. (Source: www.Avert.org)

REWIND

Look at the earlier sections of this chapter to remind yourself about grouped frequency tables if you need to. ◀

In this example, the temperature groups/class intervals will be displayed as 'categories' with gaps between each bar. As temperature is continuous, a better way to deal with it is to use a histogram with equal class intervals; you will see these in chapter 20.

2 Here is a set of raw data showing the average summer temperature (in °C) for 20 cities in the Middle East during one year.

32	42	36	40	35	36	33	32	38	37
34	40	41	39	42	38	37	42	40	41

a Copy and complete this grouped frequency table to organise the data.

Temperature (°C)	32–34	35–37	38–40	41–43
Frequency				

b Draw a horizontal bar chart to represent this data.

3 The tourism organisation on a Caribbean island records how many tourists visit from the region and how many tourists visit from international destinations. Here is their data for the first six months of this year. Draw a compound bar chart to display this data.

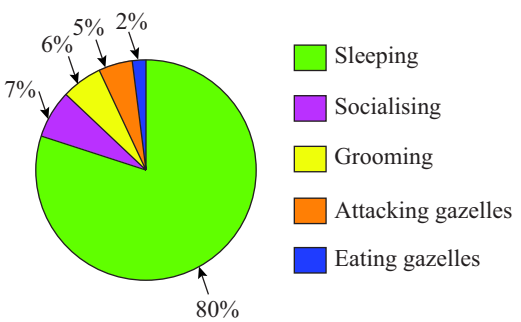
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**
Regional visitors	12 000	10 000	19 000	16 000	21 000	2 000
International visitors	40 000	39 000	15 000	12 000	19 000	25 000

Pie charts

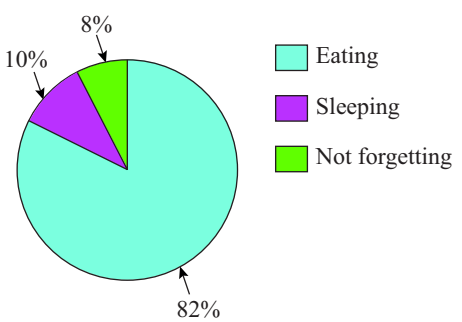
A **pie chart** is a circular chart which uses slices or sectors of the circle to show the data. The circle in a pie chart represents the 'whole' set of data. For example, if you surveyed the favourite sports played by everyone in a school then the total number of students would be represented by the circle. The sectors would represent groups of students who played each sport.

Like other charts, pie charts should have a heading and a key. Here are some fun examples of pie charts:

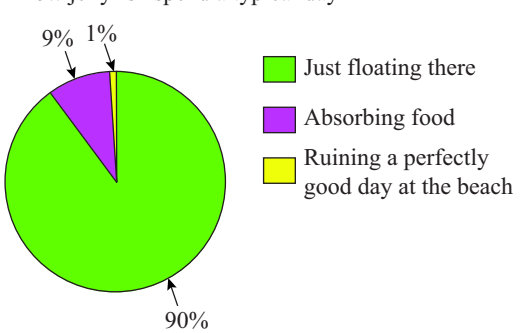
How lions spend a typical day



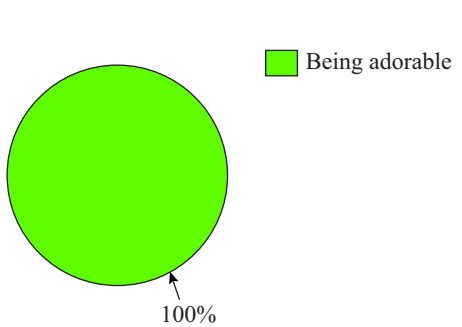
How elephants spend a typical day



How jellyfish spend a typical day



How pandas spend a typical day



Worked example 7

The table shows how a student spent her day.

Activity	School	Sleeping	Eating	Online	On the phone	Complaining about stuff
Number of hours	7	8	1.5	3	2.5	2

Draw a pie chart to show this data.

$7 + 8 + 1.5 + 3 + 2.5 + 2 = 24$ First work out the total number of hours.

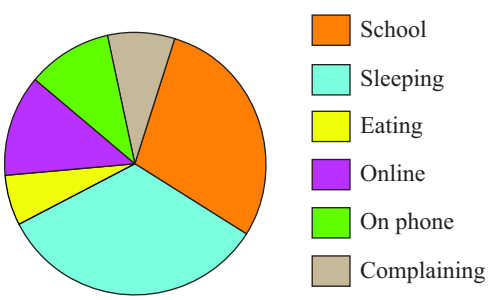
Then work out each category as a fraction of the whole and convert the fraction to degrees:

	(as a fraction of 24)	(convert to degrees)
School	$= \frac{7}{24}$	$= \frac{7}{24} \times 360 = 105°$
Sleeping	$= \frac{8}{24}$	$= \frac{8}{24} \times 360 = 120°$
Eating	$= \frac{1.5}{24} = \frac{15}{240}$	$= \frac{15}{240} \times 360 = 22.5°$
Online	$= \frac{3}{24}$	$= \frac{3}{24} \times 360 = 45°$
On the phone	$= \frac{2.5}{24} = \frac{25}{240}$	$= \frac{25}{240} \times 360 = 37.5°$
Complaining	$= \frac{2}{24}$	$= \frac{2}{24} \times 360 = 30°$

It is possible that your angles, once rounded, don't quite add up to 360°. If this happens, you can add or subtract a degree to or from the largest sector (the one with the highest frequency).

Activity	School	Sleeping	Eating	Online	On the phone	Complaining about stuff
Number of hours	7	8	1.5	3	2.5	2
Angle	105°	120°	22.5°	45°	37.5°	30°

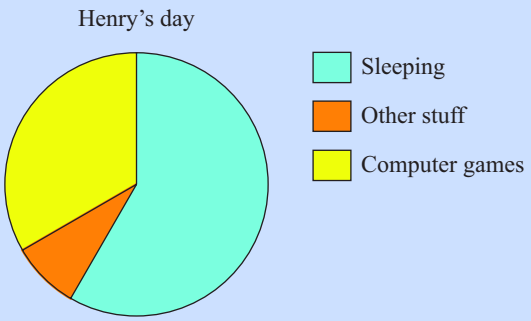
A student's day



- Draw a circle to represent the whole day.
- Use a ruler and a protractor to measure each sector.
- Label the chart and give it a title.

Worked example 8

This pie chart shows how Henry spent one day of his school holidays.



a What fraction of his day did he spend playing computer games?
b How much time did Henry spend sleeping?
c What do you think 'other stuff' involved?

a $\frac{120}{360} = \frac{1}{3}$ — Measure the angle and convert it to a fraction. The yellow sector has an angle of 120°. Convert to a fraction by writing it over 360 and simplify.

b $\frac{210}{360} \times 24 = 14$ hours — Measure the angle, convert it to hours.

c Things he didn't bother to list. Possibly eating, showering, getting dressed.

Exercise 4.9

1 The table shows the results of a survey carried out on a university campus to find out about the use of online support services among students. Draw a pie chart to illustrate this data.

Category	Number of students
Never used online support	180
Used online support in the past	120
Use online support presently	100

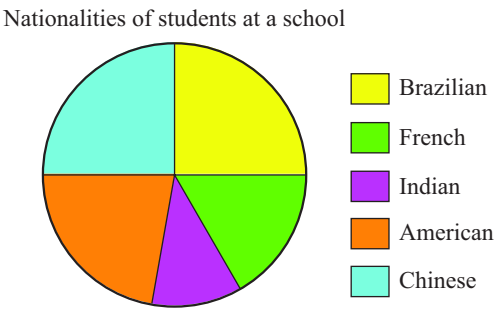
2 The table shows the home language of a number of people passing through an international airport. Display this data as a pie chart.

Language	Frequency
English	130
Spanish	144
Chinese	98
Italian	104
French	24
German	176
Japanese	22

3 The amount of land used to grow different vegetables on a farm is shown below. Draw a pie chart to show the data.

Vegetable	Squashes	Pumpkins	Cabbages	Sweet potatoes
Area of land (km²)	1.4	1.25	1.15	1.2

4 The nationalities of students in an international school is shown on this pie chart.



a What fraction of the students are Chinese?
b What percentage of the students are Indian?
c Write the ratio of Brazilian students : total students as a decimal.
d If there are 900 students at the school, how many of them are:
i Chinese? **ii** Indian? **iii** American? **iv** French?

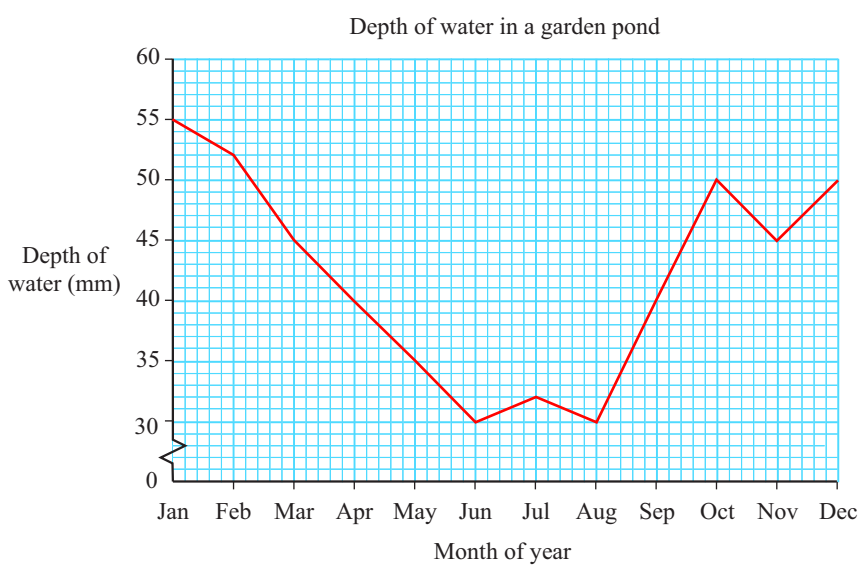
FAST FORWARD

Graphs that can be used for converting currencies or systems of units will be covered in chapter 13. Graphs dealing with time, distance and speed are covered in chapter 21. ▶

Line graphs

Some data that you collect changes with time. Examples are the average temperature each month of the year, the number of cars each hour in a supermarket car park or the amount of money in your bank account each week.

The following **line graph** shows how the depth of water in a garden pond varies over a year. The graph shows that the water level is at its lowest between June and August.



When time is one of your variables it is always plotted on the horizontal axis.

Choosing the most appropriate chart

Tip
You may be asked to give reasons for choosing a particular type of chart. Be sure to have learned the advantages and disadvantages in the table.

You cannot always say that one type of chart is better than another – it depends very much on the data and what you want to show. However, the following guidelines are useful to remember:

- Use pie charts or bar charts (single bars) if you want to compare different parts of a whole, if there is no time involved, and there are not too many pieces of data.
- Use bar charts for discrete data that does not change over time.
- Use compound bar charts if you want to compare two or more sets of discrete data.
- Use line graphs for numerical data when you want to show how something changes over time.

The table summarises the features, advantages and disadvantages of each different types of chart/graph. You can use this information to help you decide which type to use.

Tip
Before you draw a chart decide:
- how big you want the chart to be
- what scales you will use and how you will divide these up
- what title you will give the chart
- whether you need a key or not.

Chart/graph and their features	Advantages	Disadvantages
Pictogram Data is shown using symbols or pictures to represent quantities. The amount represented by each symbol is shown on a key.	Attractive and appealing, can be tailored to the subject. Easy to understand. Size of categories can be easily compared.	Symbols have to be broken up to represent 'in-between values' and may not be clear. Can be misleading as it does not give detailed information.
Bar chart Data is shown in columns measured against a scale on the axis. Double bars can be used for two sets of data. Data can be in any order. Bars should be labelled and the measurement axis should have a scale and label.	Clear to look at. Easy to compare categories and data sets. Scales are given, so you can work out values.	Chart categories can be reordered to emphasise certain effects. Useful only with clear sets of numerical data.
Pie charts Data is displayed as a fraction, percentage or decimal fraction of the whole. Each section should be labelled. A key and totals for the data should be given.	Looks nice and is easy to understand. Easy to compare categories. No scale needed. Can shows percentage of total for each category.	No exact numerical data. Hard to compare two data sets. 'Other' category can be a problem. Total is unknown unless specified. Best for three to seven categories.
Line graph Values are plotted against 'number lines' on the vertical and horizontal axes, which should be clearly marked and labelled.	Shows more detail of information than other graphs. Shows patterns and trends clearly. Other 'in-between' information can be read from the graph. Has many different formats and can be used in many different ways (for example conversion graphs, curved lines).	Useful only with numerical data. Scales can be manipulated to make data look more impressive.

FAST FORWARD
You will work with line graphs when you deal with frequency distributions in chapter 20. ▶

Exercise 4.10

1 Which type of graph would you use to show the following information? Give a reason for your choice.

a The number of people in your country looking for jobs each month this year.
b The favourite TV shows of you and nine of your friends.
c The number of people using a gym at different times during a day.
d The favourite subjects of students in a school.
e The reasons people give for not donating to a charity.
f The different languages spoken by people in your school.
g The distance you can travel on a tank of petrol in cars with different sized engines.

Applying your skills

2 Collect ten different charts from newspapers, magazines or other sources.

Stick the charts into your book.
For each graph:

a write the type of chart it is
b write a short paragraph explaining what each chart shows
c identify any trends or patterns you can see in the data.
d Is there any information missing that makes it difficult to interpret the chart? If so what is missing?
e Why do you think the particular type and style of chart was used in each case?
f Would you have chosen the same type and style of chart in each case? Why?

Summary

Do you know the following?

- In statistics, data is a set of information collected to answer a particular question.
- Categorical (qualitative) data is non-numerical. Colours, names, places and other descriptive terms are all categorical.
- Numerical (quantitative) data is collected in the form of numbers. Numerical data can be discrete or continuous. Discrete data takes a certain value; continuous data can take any value in a given range.
- Primary data is data you collect yourself from a primary source. Secondary data is data you collect from other sources (previously collected by someone else).
- Unsorted data is called raw data. Raw data can be organised using tally tables, frequency tables, stem and leaf diagrams and two-way tables to make it easier to work with.
- Data in tables can be displayed as graphs to show patterns and trends at a glance.
- Pictograms are simple graphs that use symbols to represent quantities.
- Bar charts have rows of horizontal bars or columns of vertical bars of different lengths. The bar length (or height) represents an amount. The actual amount can be read from a scale.
- Compound bar charts are used to display two or more sets of data on the same set of axes.
- Pie charts are circular charts divided into sectors to show categories of data.
- The type of graph you draw depends on the data and what you wish to show.

Are you able to … ?

- collect data to answer a statistical question
- classify different types of data
- use tallies to count and record data
- draw up a frequency table to organise data
- use class intervals to group data and draw up a grouped frequency table
- construct single and back-to-back stem and leaf diagrams to organise and display sets of data
- draw up and use two-way tables to organise two or more sets of data
- construct and interpret pictograms
- construct and interpret bar charts and compound bar charts
- construct and interpret pie charts.

Examination practice

Exam-style questions

1 Salma is a quality control inspector. She randomly selects 40 packets of biscuits at a large factory. She opens each packet and counts the number of broken biscuits it contains. Her results are as follows:

0	0	2	1	3	0	0	2	3	1
1	1	2	3	0	1	2	3	4	2
0	0	0	0	1	0	0	1	2	3
3	2	2	2	1	0	1	2	1	2

a Is this primary or secondary data to Salma? Why?
b Is the data discrete or continuous? Give a reason why.
c Copy and complete this frequency table to organise the data.

No. of broken biscuits	Tally	Frequency
0		
1		
2		
3		
4		

d What type of graph should Salma draw to display this data? Why?

2 The number of aircraft movements in and out of five main London airports during April 2017 is summarised in the table.

Airport	Gatwick	Heathrow	London City	Luton	Stansted
Total flights	23 696	39 660	6380	10 697	15 397

a Which airport handled most aircraft movement?
b How many aircraft moved in and out of Stansted Airport?
c Round each figure to the nearest thousand.
d Use the rounded figures to draw a pictogram to show this data.

3 This table shows the percentage of people who own a laptop and a mobile phone in four different districts in a large city.

District	Own a laptop	Own a mobile phone
A	45	83
B	32	72
C	61	85
D	22	68

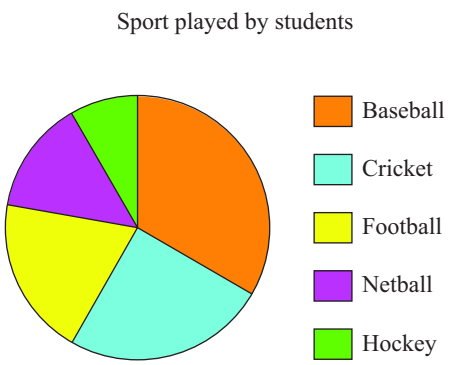
a What kind of table is this?
b If there are 6000 people in District A, how many of them own a mobile phone?
c One district is home to a University of Technology and several computer software manufacturers. Which district do you think this is? Why?
d Draw a compound bar chart to display this data.

4 This table shows how a sample of people in Hong Kong travel to work.

Mode of transport	Percentage
Metro	36
Bus	31
Motor vehicle	19
Cycle	14

Represent this data as a pie chart.

5 Study this pie chart and answer the questions that follow.



The data was collected from a sample of 200 students.

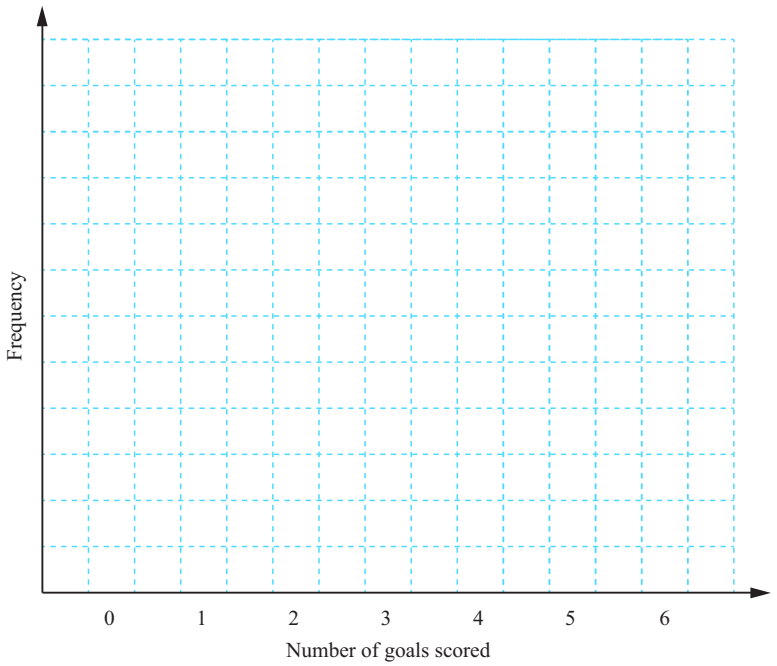
a What data does this graph show?
b How many different categories of data are there?
c Which was the most popular sport?
d What fraction of the students play cricket?
e How many students play netball?
f How many students play baseball or hockey?

Past paper questions

1 The table shows the number of goals scored in each match by Mathsletico Rangers.

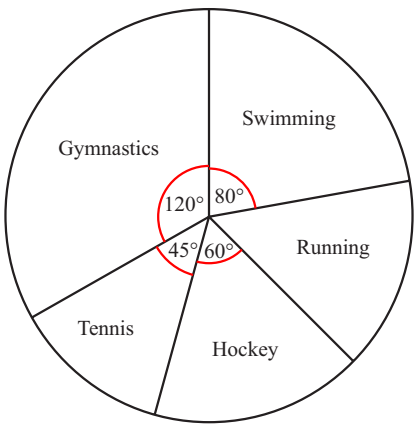
Number of goals scored	Number of matches
0	4
1	11
2	6
3	3
4	2
5	1
6	2

Draw a bar chart to show this information.
Complete the scale on the frequency axis. [3]



[Cambridge IGCSE Mathematics 0580 Paper 33 Q1 d(i) October/November 2012]

2 Some children are asked what their favourite sport is.
The results are shown in the pie chart.



i Complete the statements about the pie chart.
The sector angle for running is ..... degrees.
The least popular sport is .....
$\frac{1}{6}$ of the children chose .....
Twice as many children chose ..... as ..... [4]

ii Five more children chose swimming than hockey.
Use this information to work out the number of children who chose gymnastics. [3]

[Cambridge IGCSE Mathematics 0580 Paper 32 Q5a) October/November 2015]