

# Collecting and entering data

# 9

## syllabus reference

### Strand:

Statistics and probability

### Core topic:

Data collection and presentation

## In this chapter

- 9A Types of data
- 9B Collecting data
- 9C Organising and displaying data using column and sector graphs
- 9D Graphical methods of misrepresenting data
- 9E Histograms and frequency polygons
- 9F Stem-and-leaf plots
- 9G Five-number summaries and boxplots

## Introduction

In Australia, we are fortunate to be able to drink tap water. This is not the case in many countries throughout the world where the residents must consume bottled water. If that were the case here, we would be concerned with the price of bottled water. Since we would require ongoing supplies, it would be wise to shop around for brands and sizes that suit our particular needs.

In this chapter we investigate:

1. methods we could use to collect data
2. how we could prepare and organise the data
3. ways in which we could display the data
4. drawing conclusions from the data collected.

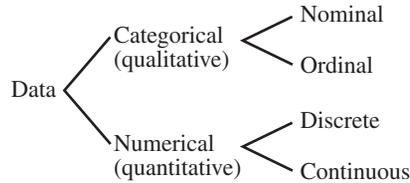


## SKILLS CHECK

- 1 Distinguish between the terms *qualitative* and *quantitative*.
- 2 Consider the following temperatures (in degrees Celsius).  
1, 3, 8, 6.5, -2, 25, 0, -1, 12
  - a Arrange the data in ascending order.
  - b Give the smallest, the largest and the middle temperatures.
- 3 At Central High School there are 117 students in Year 8 and 62 students in Year 12. In Years 9 and 10, there are 102 and 92 students respectively, while there are 77 students in Year 11.
  - a Arrange this data set in a logical table format.
  - b Draw a pie/sector graph to display the data.
  - c Represent the data as a column/bar graph.
- 4 Express the following in the units indicated.
  - a 500 g for 98c (express in c/g).
  - b \$1.10 for 1.25 L (in \$/L).
  - c 1.25 kg for \$9.50 (in c/g).
- 5 Which is the better buy:  
a 375-mL can of drink for \$1, or a 1-L bottle of drink for \$2.50?

# Types of data

Before we collect any **data**, we need to understand the various types of data we could gather. The diagram below distinguishes the types.



**Categorical data** are observations that fit some **qualitative** category. Data of this type do not involve numbers or measurements. If there is no order associated with the categories formed by the data, then they are termed **nominal data** (for example, answers to questions about a student's hair colour or method of transport used to travel to school). When the data categories are aligned to some qualitative scale, they are termed **ordinal data**. A response to a question on a scale (for example, *strongly disagree* to *strongly agree*) would constitute ordinal data (that is, some order is implied).

**Numerical data**, on the other hand, involve **quantitative** amounts. **Discrete data** are responses, observations or records that can take only certain, set values. Examples of discrete data would include, for example, the number of children in a family and the number of marks obtained in a test (even though you can be awarded half marks). **Continuous data** may take any value within the range of the data. Here, we often find data arising as the result of taking measurements (for example, a person's height, the daily temperature).

## WORKED Example 1

State whether the following pieces of data are categorical or numerical.

- a** The value of sales recorded at each branch of a fast-food outlet
- b** The breeds of dog that appear at a dog show

### THINK

- a** The value of sales at each branch can be measured.
- b** The breeds of dog at a show cannot be measured.

### WRITE

- a** The value of sales are numerical data.
- b** The breeds of dog are categorical data.



**WORKED Example 2**

State whether each of the following records of numerical data is discrete or continuous.

- a The number of people in each car that passes through a tollgate
- b The mass of a baby at birth

**THINK**

- a
  - 1 The number of people in the car must be a whole number.
  - 2 Give a written answer.
- b
  - 1 A baby's mass can be measured to various degrees of accuracy.
  - 2 Give a written answer.

**WRITE**

- a  
The data are numerical and discrete.
- b  
The data are numerical and continuous.

**remember**

1. Data can be classified as either:
  - (a) categorical — the data are in categories, or
  - (b) numerical — the data can be either measured or counted.
2. Numerical data can be either:
  - (a) discrete — the data can take only certain values, usually whole numbers, or
  - (b) continuous — the data can take any value depending on the degree of accuracy.

**EXERCISE 9A****Types of data****WORKED Example**

7

- 1 State whether the data collected in each of the following situations would be categorical or numerical.
  - a The number of matches in each box is counted for a large sample of boxes.
  - b The sex of respondents to a questionnaire is recorded as either M or F.
  - c A fisheries inspector records the lengths of 40 cod.
  - d The occurrence of hot, warm, mild and cool weather for each day in January is recorded.
  - e The actual temperature for each day in January is recorded.
  - f Cinema critics are asked to judge a film by awarding it a rating from one to five stars.

**WORKED  
Example****2**

2 State whether the numerical data gathered in each of the following situations are discrete or continuous.

- a The heights of 60 tomato plants at a plant nursery
- b The number of jelly beans in each of 50 packets
- c The time taken for each student in a class of six-year-olds to tie his or her shoelaces
- d The petrol consumption rate of a large sample of cars
- e The IQ (intelligence quotient) of each student in a class

3 For each of the following, state if the data are categorical or numerical. If numerical, state if the data are discrete or continuous.

- a The number of students in each class at your school
- b The teams people support at a football match
- c The brands of peanut butter sold at a supermarket
- d The heights of people in your class
- e The interest rate charged by each bank
- f A person's pulse rate

4 An opinion poll was conducted. A thousand people were given the statement 'Euthanasia should be legalised'. Each person was offered five responses: strongly agree, agree, unsure, disagree and strongly disagree. Describe the data type in this example.

5 A teacher marks her students' work with a grade A, B, C, D, or E. Describe the data type used.

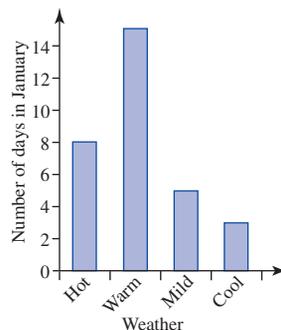
6 A teacher marks his students' work using a mark out of 100. Describe the data type used.

**7 multiple choice**

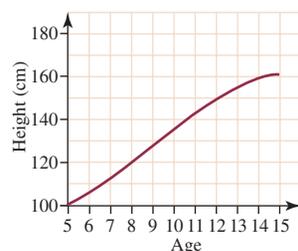
The number of people who are using a particular bus service are counted over a 2-week period. The data formed by this survey would best be described as:

- A categorical data
- B numerical and discrete data
- C numerical and continuous data
- D quantitative data

8 The graph at right shows the number of days of each weather type for the Gold Coast in January. Describe the data in this example.



9 The graph at right shows a girl's height each year for 10 years. Describe the data in this example.



## Collecting data

Data can be collected using a variety of techniques. Three common methods are:

1. observation
2. survey
3. experiment

### Observation

Let us return to consider collecting data on bottled water. We could obtain data by visiting outlets selling water and by observing the prices charged. Imagine that on a visit to three stores, a variety of brands, sizes and prices of bottled water were observed. The table below indicates the variation and costs of 1.5-L bottle of water at the three stores. The prices shown are shelf prices for one particular week.



Bottled water (1.5 L)	Corner store	Coles	Woolworths
Natural spring water	88c		
Mount Franklin Australian spring water	\$1.39	\$1.35	\$1.55
Frantelle Spring water	89c	\$1.17	99c
Rain Farm Pure Australian rainwater	\$1.14	\$1.09	\$1.19
Peats Ridge Springs Natural spring water	99c	\$1.20	
First Choice Spring water	\$1.09		
Evian Natural spring water	\$2.66	\$2.67	\$2.39
Brim-Brim Natural spring water	\$2.22		
Tourquay Natural spring water		\$1.19	
Savings Still spring water		90c	
Coles Natural spring water		\$1.29	
Bells Purified natural spring water		\$1.19	\$1.14
Schweppes Cool Ridge still spring water		\$1.70	

Farmland Natural spring water		89c	
Summit Natural spring water			\$1.19
Fraser Blue Natural Australian spring water			\$1.79
Aquaqueen Australian spring water			\$1.59
Cottonwood Valley Natural spring water			\$1.39
Snowy Mountain Natural spring water			\$1.03

Take note of:

1. the variety available at each venue
2. the variation in price between different brands (what could cause this?)
3. the brands that are common to the three stores
4. the variation in price for the same brand at the three venues.

Observations such as these would help you decide on a product suitable for your own personal needs. However, the 1.5-litre bottle may not be appropriate for you. We shall delve more deeply into this question with an investigation that also examines the variety of sizes available.



## Collecting data by observing bottled water prices

In this activity, we attempt to determine the amount of bottled water that you or your family needs.

- 1 Visit your local suppliers of bottled water (as many of them as possible). Make a record of the brands, sizes and costs of all types available at each outlet.
- 2 Organise your data into a table format (slight modifications to the table shown previously would be appropriate). These records would be entered as discrete numerical data. Check to ensure you have not made any obvious recording errors.
- 3 Compare the price of bulk buying with that of purchasing several smaller bottles of equivalent capacity.
- 4 Now look for a solution to your water requirements. (Consider your needs for a weekly supply.) Take into account the following points.
  - (a) Your bottled water supply must provide for all drinking and food preparation needs.
  - (b) The water has a long shelf life, so bulk purchases may be more economical (if you have storage room and funds available).

### Reporting your results

Write a formal report of this investigation, detailing:

1. Aim
2. Method of obtaining data
3. Table summarising data
4. Your own personal weekly needs
5. Purchases which would satisfy these needs — specify preferred brands, sizes and the total cost. Explain your choice of brand and size.



## More data collection by observation

It is possible to collect a variety of types of data by observation. The following are some suggestions you may wish to investigate.

- 1 Count the number of vehicles passing a particular point on a road during a given time period. (Road counters across roads typically record this information.)
- 2 Observe the number of people waiting at a department store for the doors to open the morning before a sale commences and on the morning on which it occurs.
- 3 Record the wildlife in a park.
- 4 Take note of the number of early morning joggers during the week compared with the numbers who jog during the weekend.
- 5 Record the variation in the price of petrol during the week.

There are many situations where data are obtained by observation. In some situations, such data provide evidence for development of a scheme, concept or physical resource. This form of data collection is commonly used for planning, marketing and the preparation of reports.

### Surveys

Collecting data by survey is the form that is most frequently used. The survey is administered with the aid of a **questionnaire**. The degree of success of obtaining meaningful and relevant data from a questionnaire depends largely on the care taken in designing the questions. Methods used to collect data include:

1. personal interviews, where the interviewer usually asks prepared questions, then records the respondents' replies
2. telephone interviews, where the interview is conducted over the telephone
3. self-administered questionnaires, which are usually mailed to individuals who complete the questionnaire, then return it in a pre-paid envelope or hold it for collection.



## Survey methods

This investigation is best undertaken in a small group.

There are advantages and disadvantages of each of the survey methods mentioned previously. As a group, discuss the various methods, then copy and complete the table below with as many advantages and disadvantages as you can.

Retain the table for consultation later in designing your own questionnaire.

Type of survey	Advantages	Disadvantages
Personal interview	• •	
Telephone interview	• •	
Self-administered questionnaire	• •	

### Questionnaires

As mentioned previously, it is important to take care in designing the questionnaire in order to obtain relevant and reliable data. In formulating the questions, we must also keep in mind that we require the data so collected to be in a form which is not difficult to analyse.

The questions posed can be either open or closed.

**Open questions** are those where the respondent has no guided boundaries within which to answer. Questions that belong to this class include:

‘Who is your favourite singer?’

‘What is your favourite food?’

The main difficulty with open questions is that they are often difficult to classify and analyse.

**Closed questions** are of the type where the respondent must answer within a category. The question about food (above) could be rephrased as:

‘Which of the following foods do you prefer most?’

- meat
- seafood
- poultry
- vegetables
- fruit

Analysis of these answers would be easier than trying to fit the open-ended answers into a category.

It must be noted that options such as:

- none of the above or
- don’t know

are to be avoided, if possible, as they provide the respondent with a non-compliance exit.

**WORKED Example 3**

Twenty people answered the open question:

‘What do you particularly *dislike* about the way the news is presented on TV?’.

They provided the following responses.

1. The ads interrupt too often.
2. Have to wait too long for the headline items.
3. The violent scenes are too graphic.
4. It was better when it was only half an hour instead of a full hour.
5. There’s too much violence.
6. The reporters are politically biased.
7. The accident scenes are not sensitively handled.
8. Some of the reports are too long.
9. It’s all about politics.
10. It mixes up local, interstate and overseas news.
11. I’d like it to be shorter.
12. There are more advertisements than news.
13. It seems to concentrate on murder and death.
14. I find it far too long.
15. The newsreaders don’t pronounce names correctly.
16. It’s too informal.
17. The newsreader is far too old.
18. I don’t like ads interrupting the news.
19. Some reports show too much blood and gore.
20. The interviews by the reporters are too long.

Classify the responses into categories to identify the main reasons given.

**THINK**

- 1 Look for 4 or 5 categories under which they could be classified.
- 2 Classify the responses under these headings.
- 3 Identify those not classified.

**WRITE**

Four main categories are apparent

1. Length
2. Violence
3. Advertisements
4. Newsreader

**Length**

Responses No. 4, 8, 11, 14, 20

**Violence**

Responses No. 3, 5, 7, 13, 19

**Advertisements**

Responses No. 1, 12, 18

**Newsreader**

Responses No. 15, 16, 17

Those not classified are:

- No. 2. Too long for headline items  
 No. 6. Reporters politically biased  
 No. 9. All about politics  
 No. 10. Mixes up news

**THINK**

- 4 Identify the main reasons.

**WRITE**

The main reasons for people disliking the TV news presentations seem to be centred around the length of the news and the violence portrayed.

*Note:* It is possible to classify these responses into different categories. No single way is correct. In practice, the responses from open-ended questions are generally classified several ways until the researcher is satisfied with the classifications.

**Preparing a questionnaire**

In preparing good questions, it is advisable to keep the following points in mind:

1. The questions should flow smoothly from one to the next.
2. Introductory remarks should be included outlining the aim and purpose of the questionnaire, along with any necessary instructions.
3. Jargon, slang and abbreviations should be avoided.
4. Do not ask questions which are vague or ambiguous.
5. Avoid bias and emotional language.
6. Avoid double-barrelled questions.
7. Do not pose leading questions (that is, those that lead to an expected response).
8. Make sure your questions are capable of being answered by your respondents.
9. Avoid questions with double negatives.
10. At the conclusion, thank the respondent for answering.

**Good and bad question writing**

As an aid to good questionnaire writing, it is helpful to practise writing questions, then do a pilot test to determine whether the questions are clear.

- 1 In small groups, discuss each of the points in the list above to gain a comprehensive understanding of good question writing.
- 2 Choose a topic that interests you.
- 3 Illustrate each of the points mentioned above by designing questions that do not conform to the guidelines of 'good question writing'.
- 4 Rewrite the questions in a clear format.
- 5 Conduct a pilot test of both sets of your questions on the members of your group.



## Survey of bottled water

With an understanding of good question writing, we can now put these techniques into practice by conducting a classroom survey.

Let us consider the bottled-water situation we encountered earlier in the chapter.

- The aim of this survey is to discover:
  - whether people prefer bottled water to tap water
  - if purchasing bottled water, their brand preference
  - how much bottled water they would drink in a week
  - what they really know about bottled water.
- To encourage your respondents to take care in completing your survey, it should have variety and be interesting. All your questions should not be of the same style. Be creative in your question construction. Questions such as: 'Draw a mouth on each face to show how much you like the types of water'



Tap water



Bottled water

create interest in your survey and encourage the respondents to answer with care. Also consider styles like 'circle the answer' and 'tick the box'.

- Design your questions to fulfil the aims of the survey. Decide on the variables to be recorded. Are they categorical, numerical, discrete or continuous?
- Make copies of your questionnaire to be answered by other members of your class; alternatively, draw up a table to record the answers if you wish to use the interview technique; then conduct your survey.
- Critically appraise your questionnaire's effectiveness after its administration. Did your questionnaire achieve the aims of the survey? Did your respondents find it interesting? Did you feel that the data obtained were reliable? Should you disregard any data?
- Retain these results for later when we will look at the variety of methods we can use to display data.

## Experiments

Collecting data by conducting experiments is commonly used in the sciences and related fields. In order to draw reliable conclusions from experiments, the data must be able to be reproduced. The conclusions from one experiment are frequently not sufficient evidence on which to base a report.

Let us return to consider the bottled-water observation discussed earlier. Much variation is evident among the prices of the 1.5-litre bottles of water. An assumption would be that the more expensive bottles contained better-quality water. An experiment could be devised to determine whether consumers can detect differences in the quality of the water samples with a relatively simple taste test.



## Bottled water taste test

The aim of this experiment is to discover whether people can distinguish, by taste,

1. between different samples of bottled water (Test 1)
2. tap water from bottled water (Test 2).

This experiment would be best performed as a whole class activity.

- 1 Prepare three jugs: one containing tap water, one with a cheap variety of bottled water and the third with an expensive variety of bottled water. Provide small cups for each jug.

Because the samples may look different (for example, tap water may be slightly coloured), this experiment would be best conducted with the participants blindfolded.

- 2 For Test 1, blindfolded students will each be given a sip of a cheap variety and a sip of an expensive variety of the bottled water. The test is to see whether the taster can identify which is the cheaper variety and which the dearer.
- 3 Test 2 is conducted similarly, using bottled water and tap water.
- 4 Prepare a result sheet (as below).

Student	Test 1 (✓ or ✗)	Test 2 (✓ or ✗)

- a Record the taste tests.
  - b Are any conclusions obvious at this stage?
  - c Identify any problems with the tests.
- 5 Retain the results of this experiment for use later, when we will consider ways of displaying data.

We must not forget that the results of this taste test could not be published as conclusive evidence of the ability of consumers generally to be able (or unable) to determine the quality of water. The results obtained by your class would need to be confirmed by similar results from tests performed on many other groups.



## Gathering data from the World Wide Web (www)

The Web is an excellent resource for data collection. We will spend some time now investigating web sites that provide interesting and relevant data for use later on.

- 1 A **census** is conducted on the Australian population every five years. The Australian Bureau of Statistics (ABS) has published the results of the 1996 census with data available for viewing on their web site. It takes some time for the data from a census to be collated and analysed, so the results of the next census (2001) may not be available until some years afterwards. Browse the site, noting the data available. Take a note of the web address for future reference.
- 2 The Bureau of Meteorology publishes climatic data for numerous towns in Queensland. Locate its web site and investigate the range of data displayed. Take particular note of the data recorded for the city or town closest to where you live. Record the site address for future use.
- 3 What is a 'gallup poll'? This famous poll is named after its founder, the American statistician, George Gallup, who was born in 1901. Another well-known poll is the Morgan Gallup Poll which publishes statistics on a variety of topical issues throughout the world. Search the Web for information on gallup polls. Record your findings in the form of a poster and present the results of your search to the class.

### remember

1. Some important methods of collecting data are observations, surveys and experiments.
2. All data collection requires careful planning with regard to the variables recorded in order to maintain the quality of the data collected.
3. A survey may be conducted personally, via the telephone or it may be self-administered. Each of these methods has advantages and disadvantages.
4. Care must be exercised in constructing a questionnaire so that the responses collected are of good quality.
5. The questions may be of an open or closed format.
6. Open questions must eventually be placed in categories before proceeding to the stage of presentation.
7. The data obtained from experiments must be capable of being reproduced before generalisations can be made.
8. In all cases of data collection, the quality and reliability of the data must be examined before processing.

**EXERCISE 9B****Collecting data**

1 Explain what you understand by the terms ‘open’ and ‘closed’ questions. Give an example of an open question. Rewrite your question in a closed format.

**WORKED  
Example****3**

2 Twenty students were asked their opinions about the cause of congestion at the school’s front gate. Analyse their responses below, suggest categories into which they could be classified and identify the most commonly stated reasons for the congestion.

1. The cars shouldn’t come up the front driveway.
2. The front entrance is too small.
3. There should be another entrance.
4. The buses are the problem.
5. The students get in the way of the cars.
6. Bike riders should have a separate entrance.
7. The Senior school and the Junior school should start and finish at different times.
8. The cars block the gate.
9. The bike riders don’t know the road rules.
10. The buses should stop further down the road.
11. Too many students.
12. Parents don’t care where they park.
13. The front gates should be wider.
14. Bike riders should go out the back gate.
15. Kids block the cars.
16. Kids just sit around talking there.
17. There should be a traffic control officer there to direct the traffic.
18. Students should not just sit around there.
19. The buses all arrive at the same time.
20. The road is too narrow.

3 Thirty students were asked: ‘Identify one thing in your maths course which you particularly don’t like’. Classify the responses below into appropriate categories, then identify the main reasons.

1. It’s too hard.
2. There’s too much homework.
3. I can’t understand the teacher.
4. It’s boring.
5. The boys are too distracting.
6. The teacher doesn’t like girls.
7. It’s too much work.
8. We get homework every night.
9. I can’t understand it.
10. The boys show off.
11. The boys always get better marks.
12. I can’t concentrate for that length of time.
13. We do something new every lesson.
14. I don’t like working in groups.
15. The teacher expects too much.
16. Our teacher is too strict.

17. The teacher doesn't help us enough with our problems.
  18. There's too much work to cover.
  19. We're expected to do assignments.
  20. I don't like doing presentations to the class.
  21. Our class is too big.
  22. The work is not interesting.
  23. I just don't understand maths.
  24. I don't like the teacher.
  25. The teacher expects too much of us.
  26. There's too much work to cover.
  27. We're expected to remember too much.
  28. It won't help me later in life.
  29. The teacher picks on me because I don't understand the work.
  30. The course is not relevant.
- 4 Identify the areas of concern in the following questions, then rewrite each so that the meaning is clear and understandable.
- (a) How much do you earn?
  - (b) Do you exercise regularly?
  - (c) Is the GST in Australia less than the VAT in England?
  - (d) Do you generally support the causes of murderous terrorists who threaten the lives of peace-loving people?
  - (e) Do you support the Prime Minister's policy on wildlife preservation?
  - (f) What is your height in inches?
  - (g) Did you buy your sneakers for comfort and quality?
  - (h) You don't agree with charging more for skim milk (where they've taken out the cream) than for full cream milk, do you?
  - (i) Do you agree that we should do more for our 'diggers' who risked their lives during the war so that we could be free?
- 5 Write the following open questions in closed format.
- (a) What is your age?
  - (b) How much pocket money do you get each week?
  - (c) How do you travel to school?
  - (d) What type of destination do you prefer for a holiday?

## Data preparation

Having considered so far the collection of data by observation, survey and experimental methods, we must now reflect on the techniques at our disposal for treating these data. However, before we rush into doing calculations, compiling tables and drawing graphs, we must first carefully examine the data for anomalies.

We must make a decision with regard to **non-compliant responses** from a respondent. Sometimes a decision is made just to disregard those non-compliant responses and to count the remainder of the compliant responses of the survey from the respondent. At other times, the whole survey from that respondent is disregarded. Alternatively, we could cope with the dilemma (after the survey has been administered) by providing another category in the question to absorb all those responses which do not slot into the categories given.

Data must also be checked for **recording errors**, which commonly occur. Recording errors should not be included as part of the data as they will distort calculations. Where possible, when errors are detected, every effort should be made to correct the data (make new observations, contact the respondent etc.). If for any reason the errors cannot be rectified, that record should be excluded.

In the business world, **databases** are often created to allow responses for each question only within defined constraints (a database may not allow a number entry where a word was expected, for instance). This reduces the chance of recording invalid data, but would not stop you entering your age as 66 instead of 16, so careful scrutiny and checking of data are essential before the results are displayed and analysed.

## Organising and displaying data

### Organising data

Once data have been collected and checked for errors they need to be put into an organised form. This involves tallying the responses to a questionnaire, accurately recording your observations or tabulating the results of your research.

This task is made easier if the questionnaire is designed with ease of tabulation in mind. Usually the results are first organised into a table and the number of responses in each category recorded. This is often done with tally marks and using the gatepost method.

### WORKED Example 4

A survey is conducted among 24 students who were asked to name their favourite spectator sport. Their responses are recorded below.

AFL	Cricket	Cricket	Soccer	Rugby League
Cricket	Tennis	Cricket	AFL	Rugby League
AFL	AFL	Rugby Union	Soccer	Netball
Basketball	Basketball	Netball	AFL	Cricket
Cricket	AFL	Rugby League	Cricket	

Put these results into a table.

#### THINK

Draw a table and beside each sport put a tally mark for each response. Every fifth tally mark becomes a gatepost.

#### WRITE

Sport	Tally	Frequency
AFL	I	6
Basketball		2
Cricket		7
Netball		2
Rugby League		3
Rugby Union		1
Soccer		2
Tennis		1

For simplicity, numerical data may be tabulated in groups.

## WORKED Example 5

A Year 11 class was surveyed on their weekly income. The responses are shown below.

\$75    \$115    \$60    \$54    \$88    \$0    \$98    \$102  
 \$56    \$45    \$83    \$71    \$40    \$37    \$87    \$117  
 \$43    \$79    \$58    \$89    \$70    \$105    \$99    \$55

Complete the table below.

Income	Tally	Frequency
0–20		
21–40		
41–60		
61–80		
81–100		
101–120		



### THINK

Count the number of responses within each category and put a tally mark in the column.

### WRITE

Income	Tally	Frequency
0–20	I	1
21–40	II	2
41–60	IIII II	7
61–80	IIII	4
81–100	IIII I	6
101–120	IIII	4

## Displaying data

The most common way of displaying data is by using a graph. Different graphs have different purposes. We will now look briefly at column graphs and sector graphs, then look at histograms, stem plots and boxplots.

### Column graphs

A **column graph** (or **bar graph**) is used when we wish to show a quantity. Categories are written on the horizontal axis and frequencies on the vertical axis.

**WORKED Example 6**

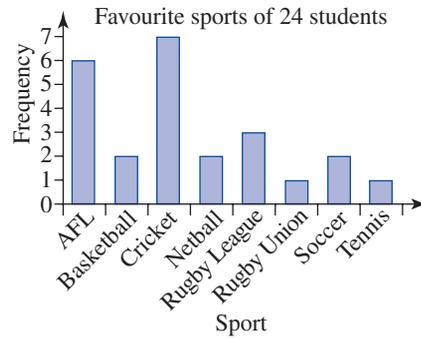
The table below shows the results of the survey on favourite sports.

Sport	Frequency
AFL	6
Basketball	2
Cricket	7
Netball	2
Rugby League	3
Rugby Union	1
Soccer	2
Tennis	1

Show this information in a column graph.

**THINK**

- 1 Draw the horizontal axis showing each sport.
- 2 Draw a vertical axis to show frequencies up to 7.
- 3 Draw the columns all the same width with gaps between.
- 4 Use a ruler.
- 5 Label the axes.
- 6 Give the graph a title.

**WRITE****Sector graphs**

A **sector graph** (circle graph, or pie graph) is used when we want the graph to display a comparison of quantities. An angle is drawn at the centre of the circle that is the same fraction of  $360^\circ$  as the fraction of people making each response.

**WORKED Example 7**

For the table in worked example 6, draw a sector graph.

**THINK**

- 1 Calculate each angle as a fraction of  $360^\circ$ .

- 2 Draw the graph.  
3 Label each sector or provide a legend.

**WRITE**

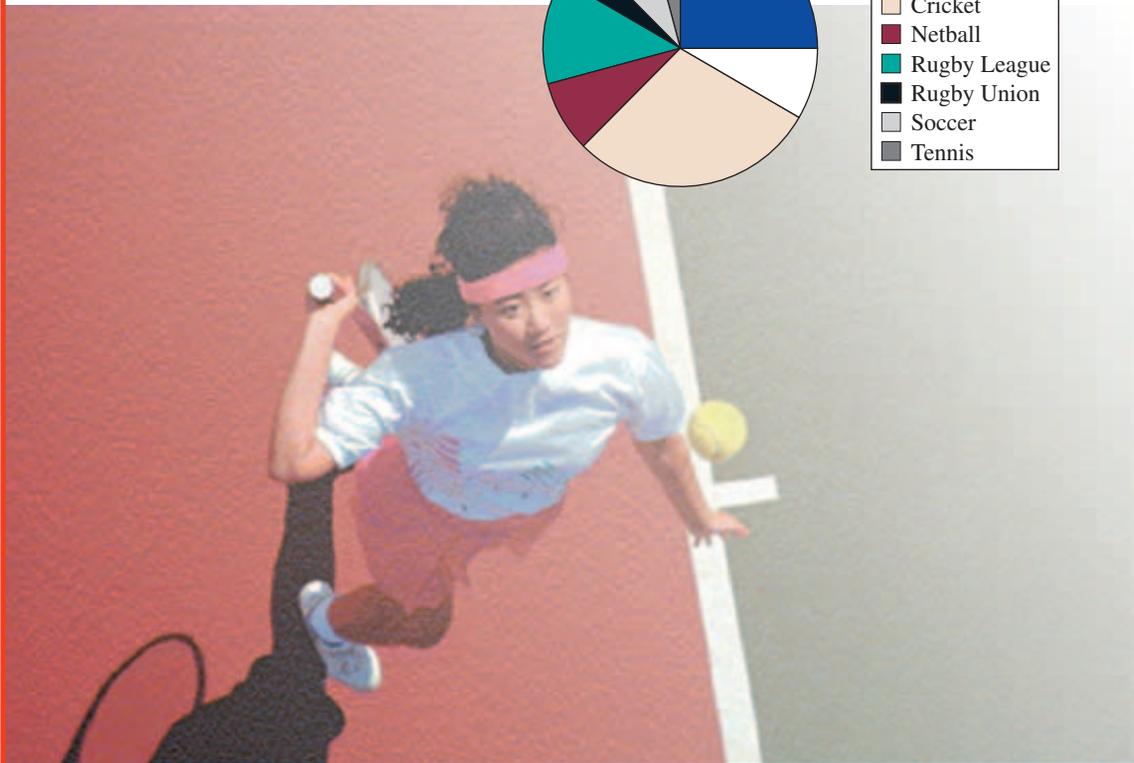
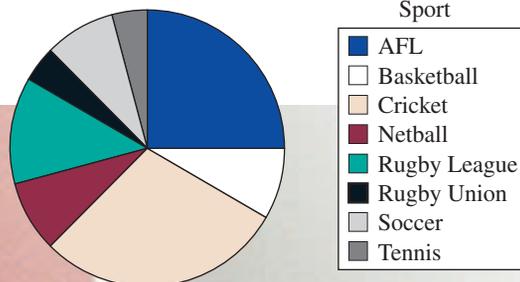
$$\text{AFL} = \frac{6}{24} \times 360^\circ = 90^\circ \quad \text{Basketball} = \frac{2}{24} \times 360^\circ = 30^\circ$$

$$\text{Cricket} = \frac{7}{24} \times 360^\circ = 105^\circ \quad \text{Netball} = \frac{2}{24} \times 360^\circ = 30^\circ$$

$$\text{Rugby League} = \frac{3}{24} \times 360^\circ = 45^\circ$$

$$\text{Rugby Union} = \frac{1}{24} \times 360^\circ = 15^\circ$$

$$\text{Soccer} = \frac{2}{24} \times 360^\circ = 30^\circ \quad \text{Tennis} = \frac{1}{24} \times 360^\circ = 15^\circ$$



These graphs can also be drawn using a spreadsheet and the charting tool.

In our next investigation we shall explore how to enter data into a spreadsheet and how to display the results in a variety of graphical forms.

## remember

1. When data are collected they are usually first organised into table form.
2. Data can be easily counted using a tally column and the gatepost method.
3. Sometimes numerical data are better organised into categories.
4. A column graph is drawn when we want to display quantities.
5. A sector graph is drawn when we want to compare quantities.

## EXERCISE 9C

### Organising and displaying data using column and sector graphs

**WORKED Example**

4

- 1 A class of students was asked to identify the make of car their family owned. Their responses are shown below.

Holden	Ford	Nissan	Mazda	Toyota	Holden
Ford	Holden	Ford	Mitsubishi	Toyota	Toyota
Nissan	Holden	Holden	Ford	Toyota	Mazda
Mazda	Toyota	Ford	Holden	Holden	Ford
Mitsubishi	Toyota	Holden	Ford	Ford	Toyota

Put these results into a table.

- 2 The results of a spelling test done by 30 students are shown below.

6	7	6	8	4	6	6	7	5	9
5	7	8	10	5	9	7	7	7	6
4	7	8	8	7	8	6	5	9	7

Put these results into a table.

**WORKED Example**

5

- 3 The marks scored on a Maths exam, out of 100, by 25 Year 11 students are shown below.

87	44	95	66	78	69	66	92	78
54	60	66	69	66	77	79	66	71
71	83	74	81	69	70	57		

Copy and complete the table below.

Mark	Tally	Frequency
40–49		
50–59		
60–69		
70–79		
80–89		
90–99		

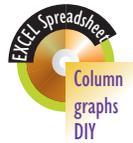


- 4 The data below show the number of customers that entered a shop each day in a certain month.

114	195	175	163	180	120	204	199
178	216	200	147	168	173	102	150
169	185	173	164	130	119	158	163
141	155	132	143	190	179	200	

Choose suitable groupings to tabulate these data.

- 5 Draw a column graph to display the data from question 1.
- 6 Draw a sector graph to display the data from question 1.
- 7 Draw a column graph to display the data from question 2.
- 8 Draw a column graph to display the data from question 3.
- 9 Draw a column graph to display the data from question 4.
- 10 Draw a sector graph to compare the number of people in each category from question 3.



**WORKED Example 6**

**WORKED Example 7**

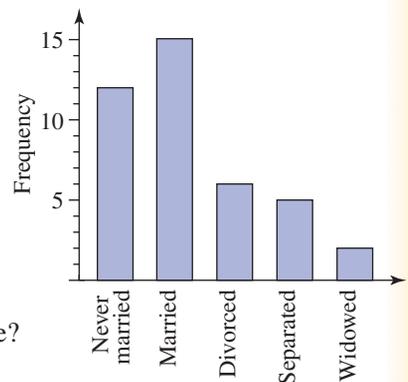
## 10 QUICK QUESTIONS 1

For questions 1–4, state if the data are quantitative or qualitative. If they are quantitative, also state whether they are continuous or discrete.

- Customers in a video shop vote for their favourite movie.
- Customers in a video shop have records kept on the number of movies they hire each year.
- The video shop keeps records of the number of times each movie has been hired.
- The video shop keeps records of the length of each movie.

The bar chart at right shows the marital status of respondents to a survey.

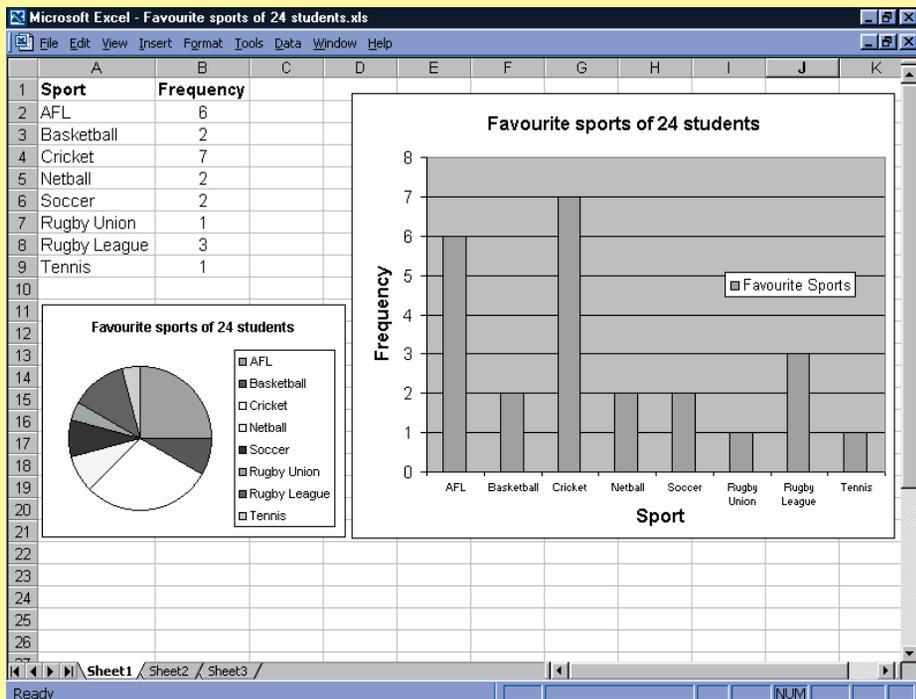
- How many people responded to the survey?
- What was the most common marital status?
- How many people were married?
- How many respondents were either divorced or separated?
- How many people had been married at some time?
- Draw a pie/sector graph of the data.





## Spreadsheets — Displaying numerical data

Worked examples 6 and 7 explained the calculations needed to display manually the same numerical data as a column graph and as a sector (pie) graph. In this activity, we look at entering the data into a spreadsheet, then exploring some of the graphical tools available. The instructions provided refer to the Excel spreadsheet. If you are using a different spreadsheet, your teacher will give you the equivalent commands.



- 1 Enter the heading 'Sport' and the categories of sport in Column A as shown above.
- 2 Enter the heading 'Frequency' and the relative frequencies in Column B as shown.
  - Highlight the sporting classifications and the frequencies, enter the **Chart Wizard** and select the **Column** graph option. Follow the instructions, remembering to label the axes and title the graph, to produce the column graph shown above.
  - Select the data again and follow through the **Chart Wizard** to produce the pie graph shown.
  - Print out a copy of your table and graphs.



## Spreadsheets — Displaying categorical coded data

The previous investigation dealt with entering and graphing numerical data. Frequently, data are collected in categories that are each given a code. Examine the shopping survey shown below. It represents a section of the 'Australian Shoppers Survey' distributed by PMP Data Based Marketing to households in Australia.

### 2. SHOPPING

**1. How many litres of soy milk do you consume per week?**

1–2 litres 1  3–4 litres 2  4 litres + 3

**2. Which of the following beverages would you consider having home delivered?**

Juice 1  Water 2  Soft Drink 3

**3. How often do you have your hair coloured at a salon?**

More than once per 6 weeks 1  Every 3 months 3  Once a year 5   
Once every 6 weeks 2  Every 6 months 4  Never 6

**4. How often do you purchase shampoo?**

Once a week 1  Once a month 3   
Once a fortnight 2  Once every 2 mths 4

**5. What type of washing machine do you own/use?**

Top Loader 1  Front Loader 3

**6. What brands of laundry detergent do you regularly use?**

Once a week 01  Dynamo 05  Radiant 09   
Cold Power 02  Dynamomatic 06  Surf 10   
Cold Powermatic 03  Omo 07  Other 11   
Drive 04  Omomatic

**7. Which of the following is your main toy store?**

Big W 01  Mr Toys 05  Toys 'R' Us 09   
dstore 02  Myer 06  tospot.com 10   
Grace Bros. 03  Target 07  Toyworld 11   
K-mart 04  Toy Kingdom

**8. How many dogs do you have in your home?**

None 1  One 2  Two 3  3 or more 4

**9. How many cats do you have in your home?**

None 1  One 2  Two 3  3 or more 4

**10. Do you treat your pets for any of the following?**

Fleas/Ticks 1  Worms 2

**11. Do you or your partner smoke?** Yes 1  No 2

**12. Please sign that you are over 18 and a smoker:**

Your signature \_\_\_\_\_

Partner's signature \_\_\_\_\_

**13. Have you purchased or would you consider purchasing any of the following by mail or telephone?**

	Have Bought		Considering	
	You	Ptnr	You	Ptnr
Books	01 <input type="checkbox"/>	23 <input type="checkbox"/>	12 <input type="checkbox"/>	34 <input type="checkbox"/>
Computer Equipment	02 <input type="checkbox"/>	24 <input type="checkbox"/>	13 <input type="checkbox"/>	35 <input type="checkbox"/>
Computer Software	03 <input type="checkbox"/>	25 <input type="checkbox"/>	14 <input type="checkbox"/>	36 <input type="checkbox"/>
Cosmetics	04 <input type="checkbox"/>	26 <input type="checkbox"/>	15 <input type="checkbox"/>	37 <input type="checkbox"/>
Craft Products	05 <input type="checkbox"/>	27 <input type="checkbox"/>	16 <input type="checkbox"/>	38 <input type="checkbox"/>
Fashion Wear	06 <input type="checkbox"/>	28 <input type="checkbox"/>	17 <input type="checkbox"/>	39 <input type="checkbox"/>
Kitchen Products	07 <input type="checkbox"/>	29 <input type="checkbox"/>	18 <input type="checkbox"/>	40 <input type="checkbox"/>
Music (Tapes, CDs, Records)	08 <input type="checkbox"/>	30 <input type="checkbox"/>	19 <input type="checkbox"/>	41 <input type="checkbox"/>
Videos	09 <input type="checkbox"/>	31 <input type="checkbox"/>	20 <input type="checkbox"/>	42 <input type="checkbox"/>
Vitamins/Health Supplements	10 <input type="checkbox"/>	32 <input type="checkbox"/>	21 <input type="checkbox"/>	43 <input type="checkbox"/>
Wine	11 <input type="checkbox"/>	33 <input type="checkbox"/>	22 <input type="checkbox"/>	44 <input type="checkbox"/>

**14. Have you/your partner bought or searched for goods to buy over the internet?**

You 1  Your partner 2

Thinking about buying beauty care products which are new to the market, please indicate whether you agree or disagree with the following statements:

**15. In general, I am among the first to buy new beauty care products when they appear on the market.**

Strongly Agree 1  Somewhat Disagree 4   
Somewhat Agree 2  Strongly Disagree 5   
Neither Agree or Disagree 3

**16. I enjoy taking chances in buying new beauty care products.**

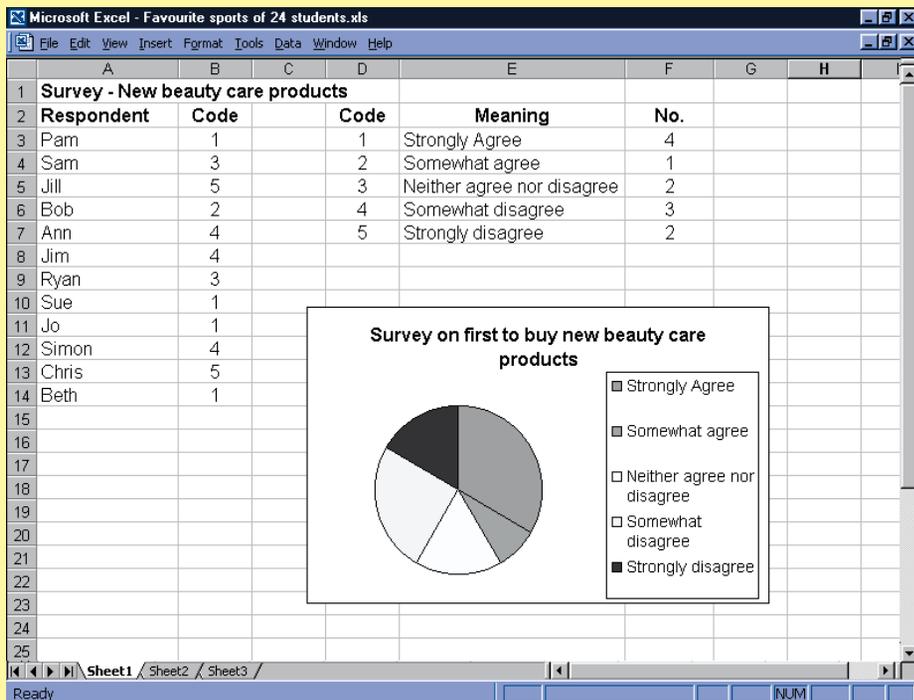
Strongly Agree 1  Somewhat Disagree 4   
Somewhat Agree 2  Strongly Disagree 5   
Neither Agree or Disagree 3

Notice that:

1. all the questions are of the closed type
2. the response categories for the questions each have a code associated with them (this enables them to be entered into a computer database or spreadsheet for analysis)
3. the ordinal data in questions 15 and 16 are coded.

Let us take a small number of responses to question 15 and see how the data might be treated.

- 1 Open a spreadsheet, head Column A with 'Respondent' and enter the respondents' names as shown below.



- 2 Head Column B as 'Code', then enter the coded responses shown beneath the heading.
- 3 Columns A and B represent the raw data. We will collate these records on the spreadsheet to the right of these columns.
- 4 Head Column D with the word 'Code', then enter the coded categories 1 to 5 beneath.
- 5 In Column E, enter the meanings of the coded categories beside their respective code (this just makes the spreadsheet and graphs more meaningful).
- 6 Head Column F with the word 'Number'. In this column we are going to count the number of 1's, 2's etc. which occur in all the responses in Column B. The formula that enables us to do this is the **COUNTIF** command. Its format is **=COUNTIF(range,criteria)**.
- 7 To count the number of 1's in the range B3 to B14, the formula would be **=COUNTIF(B3:B14,1)**. Enter this formula in Cell F3.

- 8 To count the number of 2's, in Cell F4 enter the formula **=COUNTIF(B3:B14,2)**.
- 9 Complete Cells F5, F6 and F7 with similar formulas to count the number of responses coded 3, 4 and 5 in Column B.
- 10 Use the **Chart Wizard** with Columns E and F to display graphically the responses.
- 11 Print out a copy of your spreadsheet and graph/s.



## Spreadsheets — Displaying data from the Web

In previous web searches you have noted web sites with interesting and relevant data. Select one of these sites and obtain the necessary data.

- 1 Enter the data into a spreadsheet. Remember to include meaningful headings to enhance the quality of your display.
- 2 Present your data in both table and graphical form (don't forget to label axes and include a title).
- 3 Print out a copy of your presentation and arrange it in a poster format.
- 4 Examine your graph/s and write a summary of the data.
- 5 Present your findings to the class.

## Methods of misrepresenting data

Many people have reasons for misrepresenting data: politicians may wish to magnify the progress achieved during their term, or business people may wish to accentuate their reported profits. There are numerous ways of misrepresenting data. In this section, only graphical methods of misrepresentation are considered.

### Vertical axis and horizontal axis

It is a truism that the steeper the graph the better the growth appears. A 'rule of thumb' for statisticians is that for the sake of appearances, the vertical axis should be two-thirds to three-quarters the length of the horizontal axis. This rule was established in order to have some comparability between graphs.

The following figure illustrates how distorted the graph appears when the vertical axis is disproportionately large.