ACKNOWLEDGMENTS

The Ministry of Education and Learning Media would like to thank Pip Arnold (TEAM Solutions, Auckland) and Lisa Darragh (Kowhai Intermediate School, Auckland) for writing the Teachers' Notes and contributing to the Answers section and Mike Camden and Emma Mawby (Statistics New Zealand) for reviewing and contributing to all the material.

The photographs on the cover are by Mark Coote and Adrian Heke. The illustrations on the cover are by Ben Galbraith. All other photographs, diagrams, and figures (including graphs), except for the poster, the population table, and the image disc items noted below, are by Bunkhouse graphic design. These photographs and illustrations are copyright © Crown 2009.

The Data Detective poster on page 21 is copyright © CensusAtSchool New Zealand, www.censusatschool.org.nz. The Estimated Resident Population table on page 41 is from www.stats.govt.nz, copyright © Statistics New Zealand. In the side strips, the background on page 20 is from the *EtherealNet* image disc copyright © Eyewire 2000, the photograph on page 22 is from the *Business and Transportation* image disc copyright © PhotoDisc, Inc. 1994, and that on page 28 is from the *Medicine and Technology* photo disc copyright © PhotoDisc, Inc. 1999.

Editorial team: Susan Slaughter, June Hannah, Ian Reid
Designer: Bunkhouse graphic design

Published 2009 for the Ministry of Education by Learning Media Limited,
Box 3293, Wellington 6140, New Zealand.
www.learningmedia.co.nz

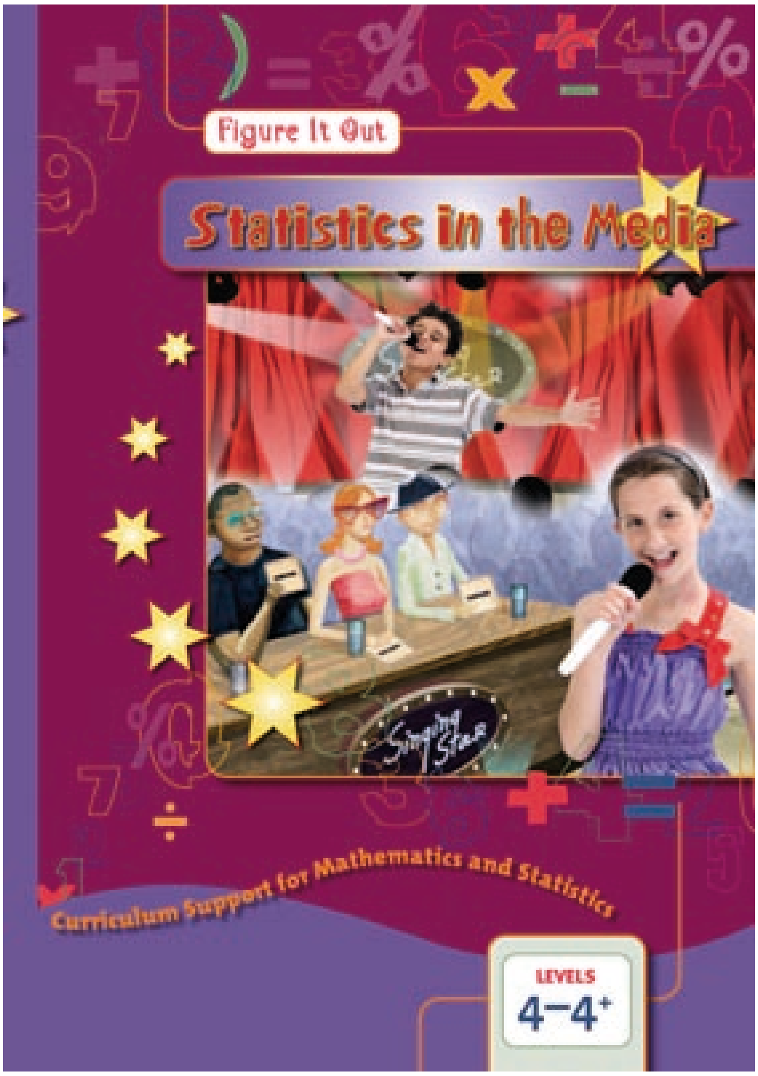
The *Answers and Teachers' Notes* are available online at
www.nzmaths.co.nz/node/1992

Dewey number 510
ISBN 978 0 7903 3402 8
Item number 33402
PDF ISBN 978 0 7903 2644 3
Students' book: item number 33409

Ministry of Education Customer Services: freephone 0800 660 662; freefax 0800 660 663;
email: orders@thechair.minedu.govt.nz; online: www.thechair.minedu.govt.nz
Please quote item number.

This resource has been developed to support
The New Zealand Curriculum

Answers and Teachers' Notes



CONTENTS

The books in the Figure It Out series are issued by the Ministry of Education to provide support material for use in New Zealand classrooms. In recent years, much of the Figure It Out student material has been aligned with Numeracy Development Project strategies, which are reflected in the Answers and in the Teachers' Notes where appropriate.

The mathematics and statistics learning area achievement objectives and the key competencies referred to in these *Answers and Teachers' Notes* are from *The New Zealand Curriculum*.

Students' books

The activities in the Figure It Out students' books are written for New Zealand students and are set in meaningful contexts, including real-life and imaginary scenarios. The contexts in the level 4–4+ *Statistics in the Media* reflect the ethnic and cultural diversity and the life experiences that are meaningful to students in years 7–8. However, you should use your judgment about whether to use the level 4–4+ book with older or younger students who are also working at this level.

Figure It Out activities can be used as the focus for teacher-led lessons, for students working in groups, or for independent activities. You can also use the activities to fill knowledge gaps (hot spots), to reinforce knowledge that has just been taught, to help students develop mental strategies, or to provide further opportunities for students moving between strategy stages of the Number Framework.

Answers and Teachers' Notes

The Answers section of the *Answers and Teachers' Notes* for *Statistics in the Media* includes full answers and explanatory notes. Students can use this section for self-marking, or you can use it for teacher-directed marking. The teachers' notes for each activity, game, or investigation include comments on mathematical ideas, processes, and principles, and suggestions on teaching approaches. The *Answers and Teachers' Notes* can also be downloaded from nzmaths at www.nzmaths.co.nz/node/1992

Using Figure It Out in the classroom

Where applicable, each activity title page of the student book starts with a list of equipment that the students will need in order to do the activities. Encourage the students to be responsible for collecting the equipment they need and returning it at the end of the session.

Many of the activities suggest different ways of recording the solution to the problem. Encourage your students to write down as much as they can about how they did investigations or found solutions, including drawing diagrams and graphs. Discussion and oral presentation of answers is encouraged in many activities, and you may wish to ask the students to do this even where the suggested instruction is to write down the answer.

Students will have various ways of solving problems or presenting the process they have used and their solutions. You should acknowledge successful ways of solving questions or problems, and where more effective or efficient processes can be used, encourage the students to consider other ways of solving a particular problem.

Answers

Statistics in the Media

Pages 1–4 **Texting Olympics**

ACTIVITY ONE

1. Details to sort out will vary. For example, they could be based on these questions:
   - Can I have a practice first?
   - I've never had a cellphone. Can I get extra time?
   - How will the timer know the exact second I finish?
   - Can I set up my phone ready to text to someone?
   - Can we use predictive text? (Predictive text [when the cellphone predicts the word or words you want to use from some of the keys you press] can vary from phone to phone – for example, newer models have more advanced methods, so it would be fairer if no one used this feature.)
   - Most people our age use texting language, so can we use it in the marathon? (People invent various short cuts for texting [for example, "c u l8r" for "see you later"], and although they usually make sense, it would be very hard to judge because there are different forms of the same words and some would take less time than others. Also, what makes sense to the person doing the texting may not make sense to the reader.)

2.

Name	Sprint	Marathon	Hurdles
Rebecca	5.4	49.7	42.4
Aki	12.9	222.8	107.1
Aoife	18.8	175.6	106.7
Mere	5.0	60.0	49.6
Nina	6.0	63.0	53.4
Ariana	5.0	75.0	46.7
Thomas	4.7	61.0	41.0
Laurel	8.7	79.9	55.1
Abhay	11.1	104.8	66.6
Hua-Ling	12.2	80.4	81.8
Fiona	12.2	105.9	67.4
Ming	4.1	46.1	43.1
Rangi	6.7	72.1	43.5
Siri	8.2	51.3	28.4
Quaid	6.0	71.5	113.1
Andrew	5.5	108.8	74.8
Matthew	7.1	88.8	67.8
Yvette	4.3	108.5	88.7
Susan	5.5	90.0	41.7
Tariq	8.1	43.3	39.6
Chris	7.3	114.9	54.9
Toline	4.0	55.9	30.5
Taylah	7.9	99.9	60.7
Molly	5.0	61.4	45.0
Quinten	10.4	80.3	73.7
Koria	16.9	237.2	173.3
Mary	6.7	37.9	25.1
Aketu	7.4	88.5	40.4
Pania	4.8	70.0	53.7
Mike	5.5	63.6	51.6

3. a. A graph with the part-seconds in order would look like this:

Sprint Times

4	0 1 3 7 8
5	0 0 0 4 5 5 5
6	0 0 7 7
7	1 3 4 9
8	1 2 7
9	
10	4
11	1
12	2 2 9
13	
14	
15	
16	9
17	
18	8
Seconds	

   b. To win a sprint, you need to have the fastest time. The fastest time is 4.0 seconds, which was Toline's. There were 5 times below 5 seconds: 4.0, 4.1, 4.3, 4.7, and 4.8. This is $\frac{1}{6}$ of the class. Quinten wants to include the 3 students who were timed at exactly 5 seconds. But 8 out of 30 is more than $\frac{1}{4}$, and calling that many students "champions" makes the description less meaningful.

4. a. Possible observations are:
   - Very few students: took more than 14 seconds to do the sprint

     took between 10 and 12 seconds to do the sprint.

- Almost half the students: took between 4 and 6 seconds to do the sprint

  took between 6 and 12 seconds to do the sprint.

- No students: took less than 4 seconds to do the sprint

  took between 14 and 16 seconds to do the sprint.

- Most students: took more than 5 seconds to do the sprint

  took between 4 and 8 seconds to do the sprint.

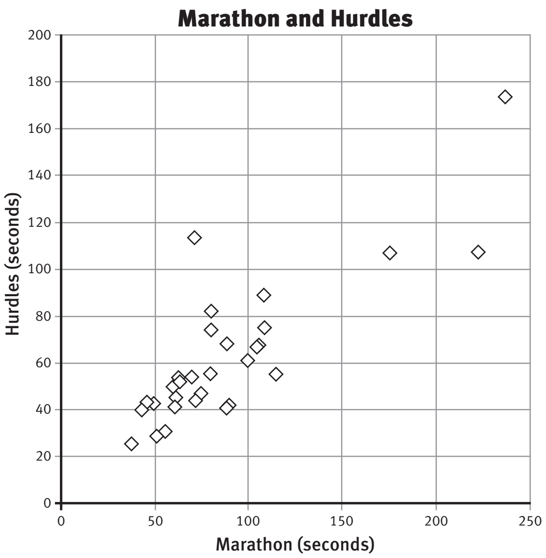
- More students took between 4 and 6 seconds to do the sprint than for any other time interval.

b. Opinions will vary. The stem-and-leaf graph gives more information than the histogram. The histogram summarises the data more, and individual values are no longer apparent. Gaps and peaks are probably more obvious in the histogram.

ACTIVITY TWO

1. a. Rebecca and Koria

b. A possible graph is:



2. a. Koria, Aki, Aoife

b. Possible choices are:

- Much better in marathon than hurdles: Quaid
- Much better in hurdles than marathon (although not necessarily very good at either!): Aki, Aoife, Abhay, Andrew, Susan, Chris, Koria, Aketu

c. Mary

d. Mary, Siri, Tariq, Toline, Ming, Rebecca

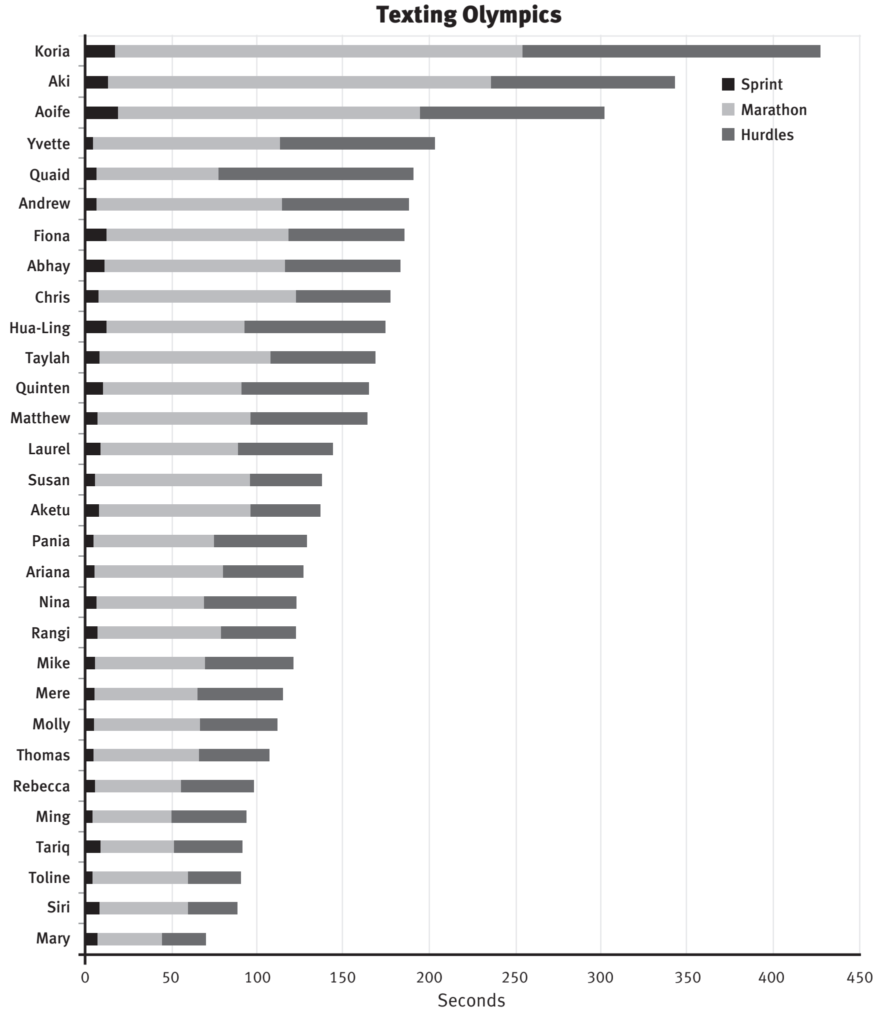
3. Generally, students with fast marathon times have fast hurdle times (see **2d**).

ACTIVITY THREE

1. a. A fourth column, sorted to show total time in order for all the students, would be:

Name	Total time
Mary	69.7
Siri	87.9
Toline	90.4
Tariq	91.0
Ming	93.3
Rebecca	97.5
Thomas	106.7
Molly	111.4
Mere	114.6
Mike	120.7
Rangi	122.3
Nina	122.4
Ariana	126.7
Pania	128.5
Aketu	136.3
Susan	137.2
Laurel	143.7
Matthew	163.7
Quinten	164.4
Taylah	168.5
Hua-Ling	174.4
Chris	177.1
Abhay	182.5
Fiona	185.5
Andrew	189.1
Quaid	190.6
Yvette	201.5
Aoife	301.1
Aki	342.8
Koria	427.4

A stacked bar graph showing the data (from longest to shortest total time) for all the students could look like this:



b. Mary, Siri, Toline, Tariq, Ming

2. a. Mary, Toline, Ming, Tariq, Siri

A points table in order for the people who scored points under Quinten's system would be:

Name	Sprint	Marathon	Hurdles	Total
Mary		5	5	10
Toline	5		3	8
Ming	4	3		7
Tariq		4	2	6
Siri		1	4	5
Yvette	3			3
Thomas	2			2
Rebecca		2		2
Pania	1			1
Aketu			1	1

For this class, the team of 5 remains the same, although the order changes.

b. Advantages and disadvantages of each system: Taylah's system generates the better all-round candidates and has more potential to get an overall winner in a gymnastics-style competition or "triathlon". Quinten's system highlights those who are very good in single events and has more potential to generate a gold medal in an individual event.

c. Decisions will vary.

Under Quinten's system, the 6th person (the reserve) would be Yvette, who did very well (3rd) in the sprint but poorly in the other two events.

Under Taylah's system of total time, Rebecca (as 6th) would be the reserve, with a total time only 4.2 seconds less than Ming's.

INVESTIGATION

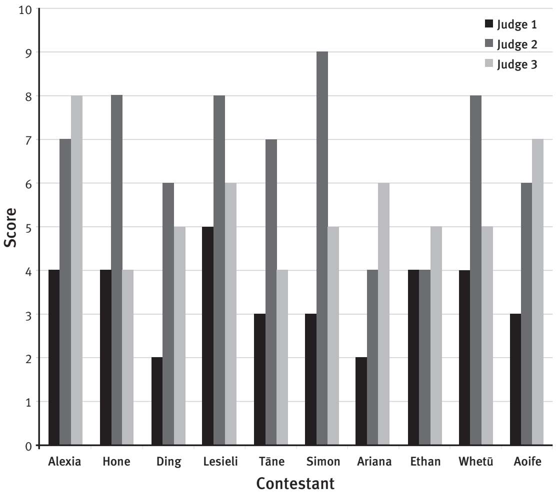
Investigations will vary.

Pages 5–7 Singing Star

ACTIVITY ONE

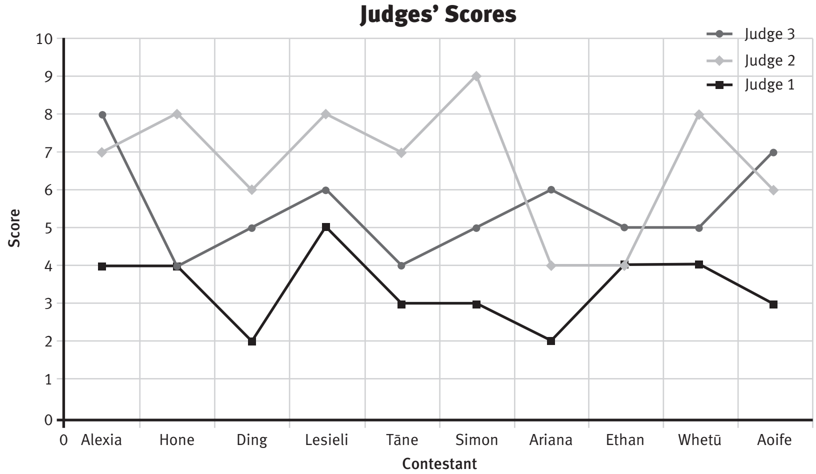
1. a. A suitable graph would be:

Judges' Scores



There are other possible graphs, including a multiple line graph like this:

Judges' Scores



b. Very little. (You can't tell how good the contestants really are because the judges' scores vary so much.)

c. The graph tells you quite a lot about the judges. For example:

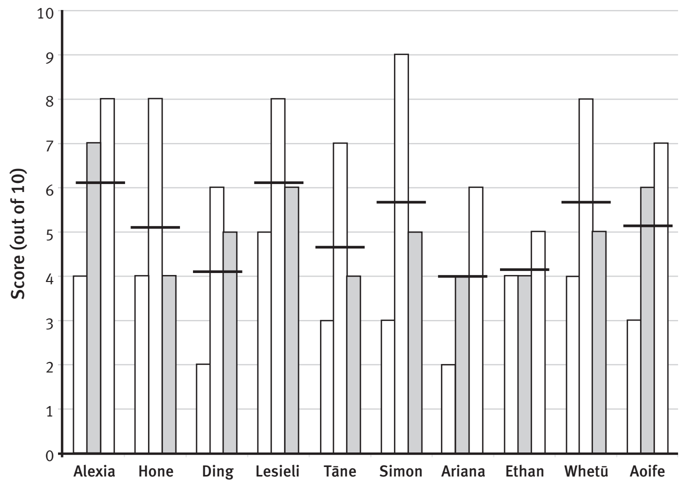
- Judge 2 generally gives higher scores than the other two.
- Judge 1 consistently gives the lowest scores.
- For 7 contestants, judge 2's score is at least double judge 1's score.
- Half of judge 1's scores are much lower than those of the other two judges, and the other half are equal to or 1 point less than judge 3's.
- Judge 2's scores vary most (6 are higher than those of the other judges, and 4 of these are 3–4 points higher than the next score).
- Ethan is the only contestant that all 3 judges agree on, with scores within 1 point.
- The judges' scores don't seem to be very consistent across the contestants.

2. a.

Order	Contestant	Judge 1	Judge 2	Judge 3	Total	Median	Mean (1 d.p.)
1	Alexia	4	7	8	19	7	6.3
2	Hone	4	8	4	16	4	5.3
3	Ding	2	6	5	13	5	4.3
4	Lesieli	5	8	6	19	6	6.3
5	Tāne	3	7	4	14	4	4.7
6	Simon	3	9	5	17	5	5.7
7	Ariana	2	4	6	12	4	4.0
8	Ethan	4	4	5	13	4	4.3
9	Whetū	4	8	5	17	5	5.7
10	Aoife	3	6	7	16	6	5.3

b. In the graph below, the highlighted bar shows the median and the line drawn across each set of three bars shows the mean.

Judges' Scores



c. There are a number of comments you can now make about contestants, including:

- Based on the median, the top 3 contestants are Alexia, Lesieli, and Aoife.
- The median clearly identifies the lowest 4 contestants.
- The median does not help sort out the middle of the field because 3 contestants have a median score of 5.
- The mean puts Alexia and Lesieli in top-equal position and Whetū and Simon in third-equal position.
- Both median and mean identify Alexia as the top contestant.
- The median identifies Lesieli as second equal, and the mean identifies her as second.

3. a. The median looks at the middle judge's score and is not affected by the other two judges' scores. The mean considers all three judges' scores and can be strongly influenced by a particularly high or low score. Where the highest score is a lot higher than the other two scores (for example, Hone's), the mean score is higher than the median score. Where the lowest score is a lot lower than the other two scores (for example, Ding's), the mean score is lower than the median score. The median is always one of the judges' scores, but this may not be the case for the mean score.

b. Two possible rankings are:

Alexia	7
Lesieli	6
Aoife	6
Ding	5
Simon	5
Whetū	5
Hone	4
Tāne	4
Ariana	4
Ethan	4

Alexia	6.3
Lesieli	6.3
Simon	5.7
Whetū	5.7
Hone	5.3
Aoife	5.3
Tāne	4.7
Ding	4.3
Ethan	4.3
Ariana	4.0

The first table is sorted by the median, and the second is sorted by mean scores. (You might also sort by the total score. This would have the same effect as sorting by the mean, so Alexia would still be top.)

c. Ariana because she has one of the lowest median scores and the lowest mean score

4. a.

Simon	36 876
Alexia	34 787
Whetū	33 239
Lesieli	32 723
Hone	23 876
Ariana	22 454
Tāne	21 989
Aoife	11 872
Ding	11 750
Ethan	11 200

b. Comments could include:

- The public's top 4 contestants (Simon, Alexia, Whetū, and Lesieli) are also the top 4 contestants according to the means of the judges' scores, but in a different order.
- The public's top 4 contestants are all in the group of the top 6 contestants according to the medians of the judges' scores, but in a different order.
- Of the public's bottom 3 contestants, Ethan is equal bottom according to the judges' median rankings. He and Ding are second bottom according to the judges' mean ranking, but Aoife doesn't show in the bottom 4 in either ranking.
- Ariana is the lowest ranked by the judges' mean score and equal bottom according to the medians, but she is ranked 6th by the public.

5. Choices and reasons may vary. A good place to start would be the 3 lowest-scoring contestants in the public poll: Ethan, Ding, and Aoife. Of these, Aoife scored best with the judges (second-equal median and fifth-equal mean). That leaves Ethan and Ding. Ethan got the lowest public vote and a lower median judges' score than Ding, which suggests Ethan could fairly be the one to go. Although Ariana scored bottom with the judges, she came sixth in the public vote, so she should probably remain in the competition at this stage. Alternatively, you could start with the bottom 3 according to the judges and then compare them with the public vote. From both perspectives, Ethan should be eliminated.

ACTIVITY TWO

1. The graph shows that there is a peak when the fifth performer is on stage. The fifth performer is Tāne, who may be the contestant whom people love to hate (one member of the audience says he is "so cool", but that doesn't mean he can perform well!). Although he scores in the bottom 4 with both judges and public, he generates big audiences, so the producer may not even look at eliminating him at this stage.

2. Ideas may vary, but the most important improvement would be to replace the "performance order" numbers on the horizontal axis of the graph with the names of the performers. This would not only save you having to refer back to the table on page **6**, it would also mean that the graph showed all the information needed to tell its story.

Pages 8–9 Billboards

ACTIVITY ONE

a.–b. Possible answers include:

i. Target market: people with access to computers who like playing games on them

Location of sign: near a school (to catch the attention of students, associated adults, and passing traffic); on a city site with lots of passing traffic (vehicles and pedestrians); near a shopping complex or a shop that sells computer games

ii. Target market: travellers ready for a meal break; people on work breaks or those going to and from work who might decide to stop and eat on the way; shift workers on their way to or from work; people going into or out of the city for shopping or other business

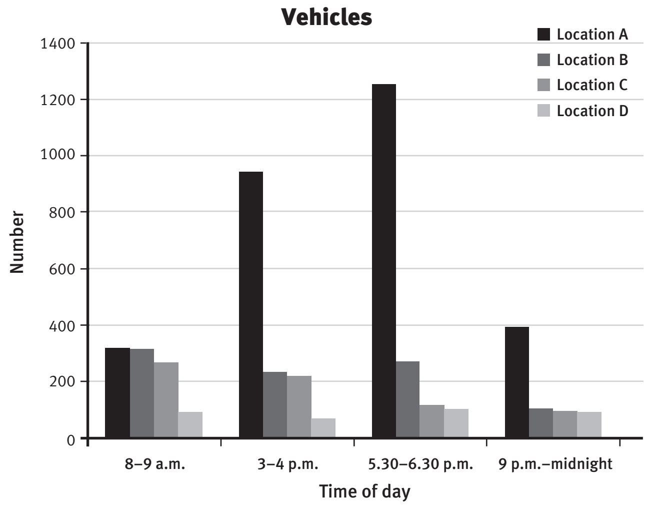
Location of sign: within easy reach of a convenient place to pull out of traffic on way in or out of the city; near some large businesses employing many people; near big employers of shift workers. (The distance on the sign would need to match the distance from the location chosen.)

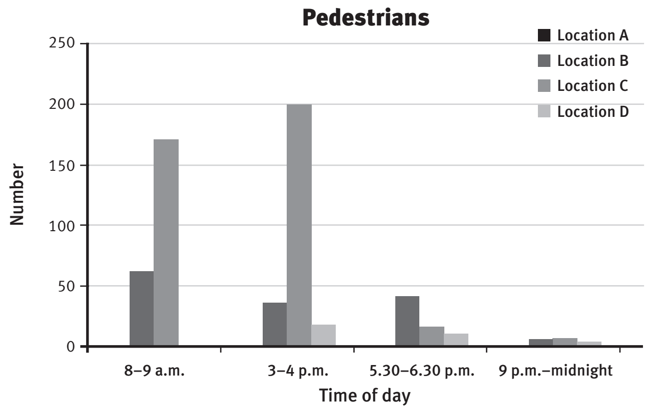
iii. Target market: people travelling in or out of the city; someone prompted by the sign to realise their fuel is low; busy people who don't want to queue at an inner-city service station

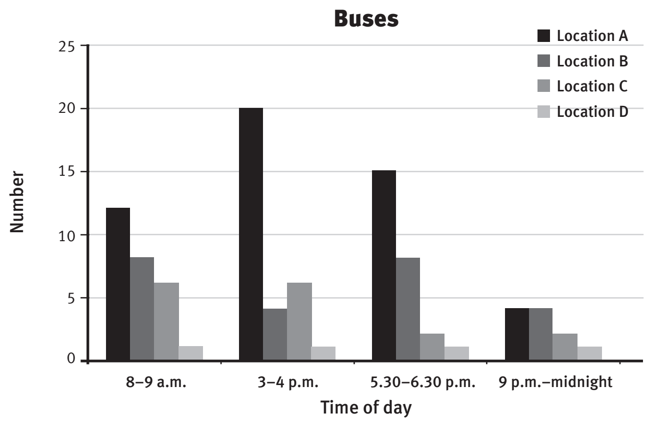
Location of sign: close to the service station but far enough away to give drivers time to look at their gauges and make a safe exit off the road; at a convenient place to pull out of traffic on way in or out of the city.

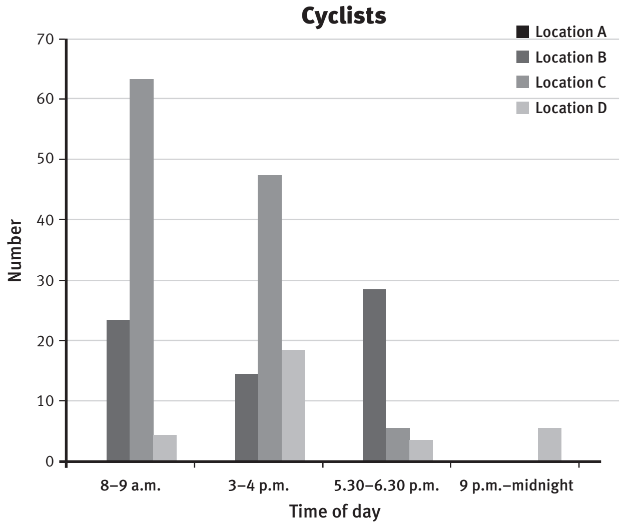
ACTIVITY TWO

1. a.









b. i. Statements will vary. For example:

Vehicles graph:

- Location A's peak flow is between 5.30 and 6.30 p.m. (presumably because of people going home from work), and it has a higher flow than other locations at other times. Location A has by far the highest overall vehicle flow.
- Location B has a reasonable flow of vehicles all day up to 6.30 p.m., which fits with most city working/shopping hours.
- Location C has its best vehicle flow between 8 and 9 a.m. and between 3 and 4 p.m., which relate to school hours.
- Location D has a low flow at all times.

Pedestrians graph:

- Location A has no pedestrian traffic.
- Location B has a steady flow all day but little at night.
- Location C has by far the highest overall pedestrian flow of all the locations, although this is greatest between 8 and 9 a.m. and between 3 and 4 p.m. (which relate to school hours).
- Location D has very low pedestrian traffic at all times.

Buses graph:

- Location A has by far the best flows between 3 p.m. and 6.30 p.m. and the highest bus flow overall.
- Location B has more buses between 8 and 9 a.m. and between 5.30 and 6.30 p.m. than at other times.
- Location C has more buses at the start and finish of school than at other times.
- Location D has few buses passing at any time.

Cyclists graph:

- Location C has by far the best flow at the start and finish of school but no flow at all between 9 p.m. and midnight.
- Location D is the only location that has cyclist traffic at all times, but the numbers are low.
- Some locations have no cyclist traffic at certain times. One location has no cyclists at all.

b. ii. Discussion will vary.

2. Use of data and graphs will vary. Possible recommendations include:

Computer game ad: position C, which has reasonable vehicles and buses' traffic flow and high pedestrian and cycle numbers between 8 and 9 a.m. and 3 and 4 p.m. If the ad is partly aimed at schoolchildren or associated adults, the location near a school is suitable.

Food ad: position B. It has reasonable numbers of vehicles, cyclists, and pedestrians between 5.30 and 6.30 p.m., which is near evening meal time; the traffic will probably be slower at that busier time because of the traffic lights, and it may be on people's way home from work.

Petrol station: position A. It's a main road in or out of the city and has the highest vehicle traffic of all the locations.

ACTIVITY THREE

Discussion will vary. It isn't wise to base a decision on data collected on 1 day only because data collected on different days of the week or at other times of the day may lead to different recommendations.

The data doesn't tell us about factors such as the speed of traffic, the level of concentration the drivers need at those locations, how well lit the billboards are at night, or other possible distractions for those who pass by.

Pages 10–11 Cheap Reads

ACTIVITY ONE

1. **a.** *Starry-eyed*

   b. *Woman's Time*

2. There are significantly more casual readers than primary readers. The magazines with more primary readers tend to have more casual readers as well.

3. **a.** Comments will vary. Primary readership is based on sales and can therefore be quite accurate. Casual readership is more likely to be based on data from media research companies, who do surveys to make estimates. (An Internet search on magazine readership will give more information on this.)

   b. Magazine publishers and advertising agencies would be interested in this information. They would obtain information from media research companies (who would get some of their information from distributors of the magazines).

   c. Many people hand their magazines on to places such as hospitals, hair salons, dentists, or doctors for people to read while they are filling in time.

ACTIVITY TWO

1. **a.** *C-Zone* and *Cartooner*

   b. *2 Much* and *Cartooner*

   c. Answers will vary, depending on the type of ad, but probably *C-Zone* or *Cartooner* because the greatest number of people (from this survey) purchased and/or read them.

2. **a.** The data is not very reliable for the wider population of that age group but is probably quite representative of students of that age in that school (depending on how the 20 students were selected).

   b. It depends on whether the focus group is restricted in terms of age, gender, location, income, and so on.

3. Sarah might conclude that a wide range of magazines are typically read by young people of her age. Some of the more popular reads are *Cartooner* and *C-Zone*, with 90–95% of those surveyed reading them. *Frenz* and *2 Much* are also popular, with 75% of those surveyed reading them. The least popular were *Polly* and *Wrestle*, with only 45% readership.

INVESTIGATION

Investigations and conclusions will vary.

Pages 12–13 Blockbusters

ACTIVITY ONE

1. **a.** Most of the movies were between 100 and 130 minutes; just under one-third were 130 minutes or over, but these were widely spread between 130 and 200 minutes; the most common length was 120–130 (14 movies), closely followed by 110–120 (12 movies).

   b. Given that there is an even number of movies (50), there is no middle: the median actually falls between the 25th and 26th movie. Both these movies are to be found somewhere in the 120–130 interval. So the "mid-length" or median movie is somewhere between 120 and 130 minutes.

2. **a.** The typical best picture movie length in the 1980s was between 110 and 130 minutes. Just over half of the data falls into this interval, but if you also include the 8 movies between 100 and 110 minutes, nearly 70% of movies were between 100 and 130 minutes.

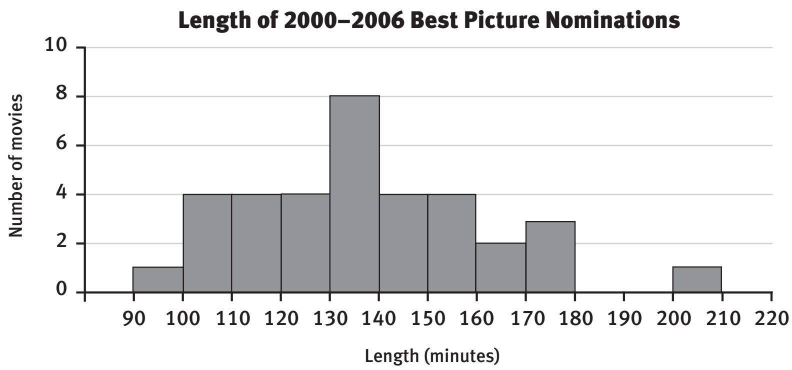
   b. 180–200 minutes

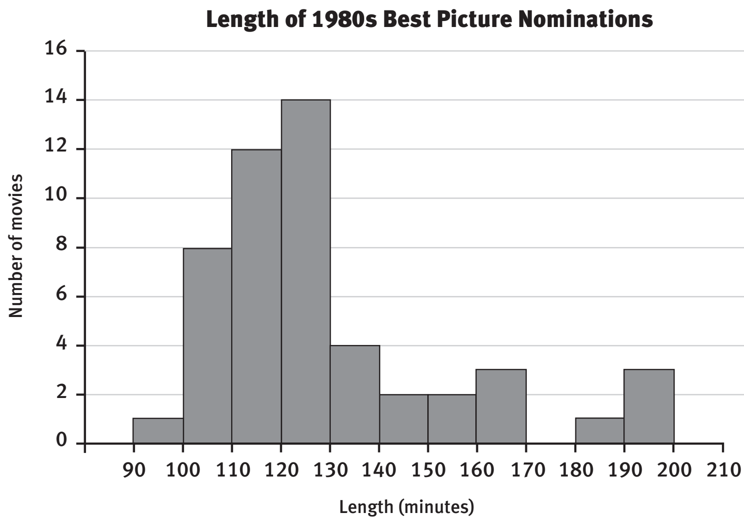
3. You could use a dot plot to show the length of each film. To do so, you would have to use a horizontal axis that extended from 90 to 200 minutes. If you wanted to use a histogram or bar graph, you would have to group the data as shown in the example in the student book.

ACTIVITY TWO

1.

Length (in minutes) of 2000–2006 Best Picture Nominations												
90–99	100–109	110–119	120–129	130–139	140–149	150–159	160–169	170–179	180–189	190–199	200–209	210–219
/	////	////	////	~~////~~ ///	////	////	//	///			/	
1	4	4	4	8	4	4	2	3	0	0	1	0



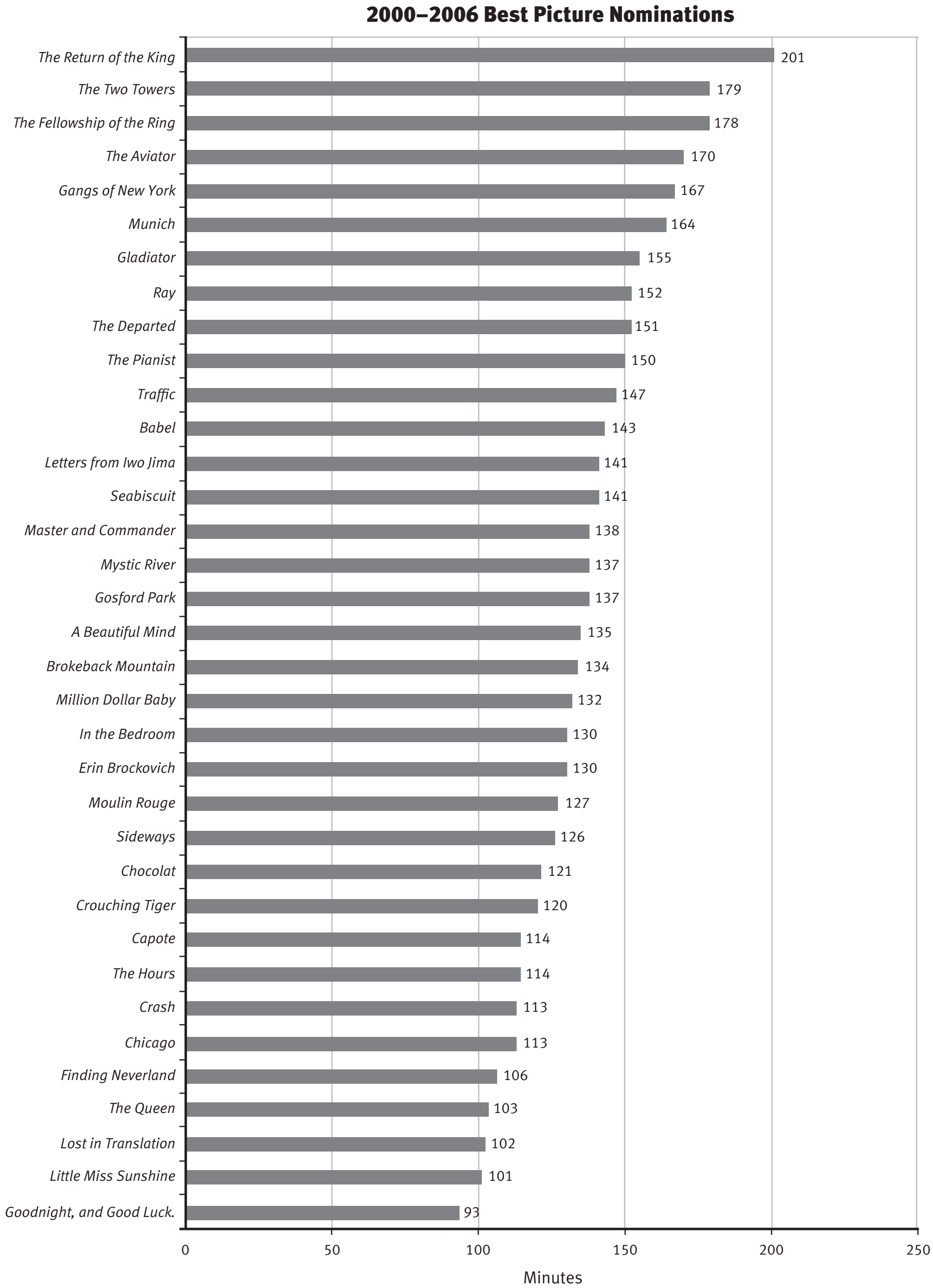


The histogram from page **12** in the students' book is repeated here to make it easier to compare the features of the two histograms. (You could also graph this information on dot plots.)

2. The second set of data is for 7 years only, so you cannot compare actual numbers. However, you can make comments such as: Compared with the 1980s graph, the 2000–2006 data is more evenly spread; only one bar stands out, and 5 of the bars are the same height. The movies are generally longer – just over 60% were 130 minutes or longer, compared with 30% of the 1980s movies. Reasons may include:
   - Using the latest digital technology, it may cost less to make longer movies than it did in the 1980s.
   - It may be that long, blockbuster-type movies are more in demand.
   - It may be that more of the top movies happen to be based on lengthy novels.
   - It may be that people have come to expect at least 2 hours of movie time to get value for their money.
   - Directors may be going to more trouble to develop their characters, and this takes time.
   - Perhaps movies only get taken seriously (and nominated for Best Picture!) if they are long.

3. **a.** The actual lengths of individual movies, the names of the movies, the year of each movie

   b. Different graphs are possible, for example, dot plots and stacked bar graphs. For example, on the following page is a bar graph that shows names and lengths.

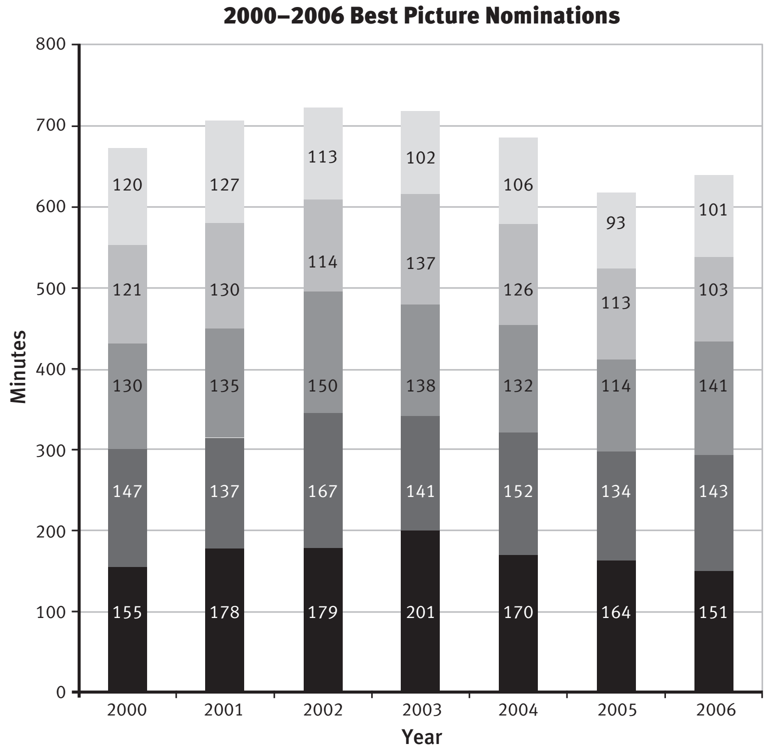


From this graph, you can easily see all the movies by name and how their lengths compare. You can also see which movie was the shortest or the longest and work out the median length (135) and information such as: 10 movies were $2\frac{1}{2}$ or more hours, 10 were 2 hours or less, and 15 were between 2 and $2\frac{1}{2}$ hours.

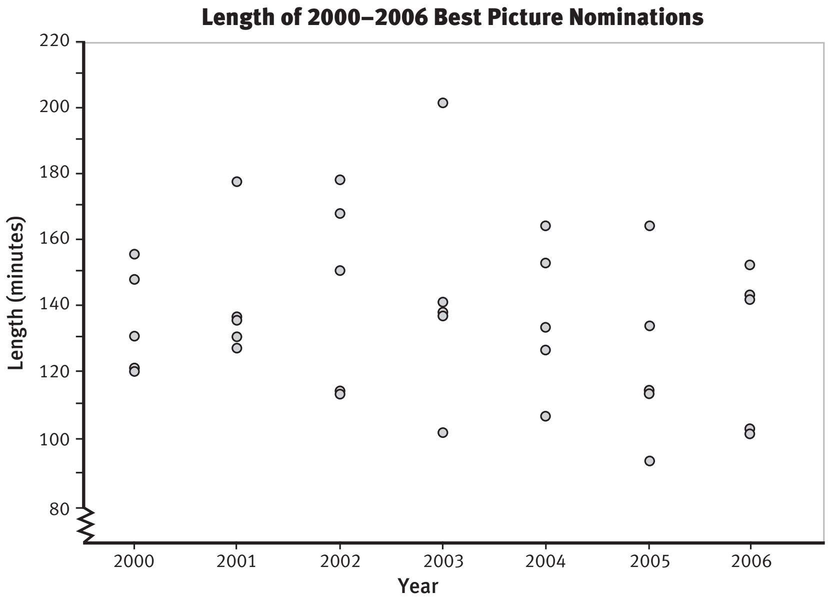
c. Answers will vary. If your graph shows similar information to the one above, other people can see from it that most of the nominated movies were more than 2 hours long.

d. Answers will vary. The graph above ignores the years of nomination, so you can't look for time-related patterns.

4. One way of comparing movie lengths by year is a stacked bar graph like this:



You could also make a dot plot to show the lengths for each year.



For a table such as the one below, you will need to do some mean and median calculations.

Best Picture Nominations

	Year							
	2000	2001	2002	2003	2004	2005	2006	Row summary
Mean	134.6	141.4	144.6	143.8	137.2	123.6	127.8	136.1
Median	130	135	150	138	132	114	141	134.3

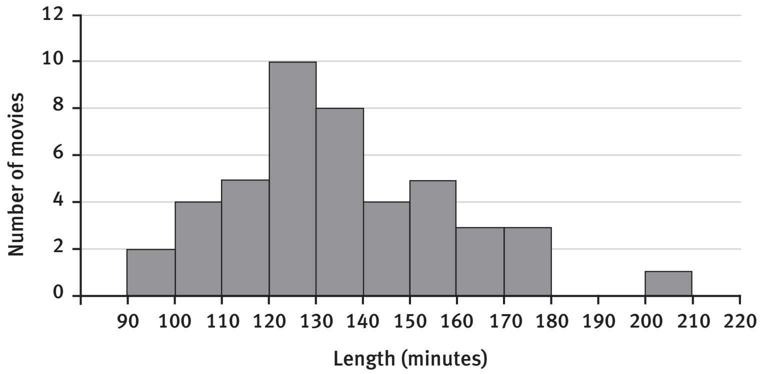
The graphs and the table show that there doesn't appear to be a trend toward increasing lengths in movies with each year. In fact, they may be getting shorter overall. Between 2001 and 2003 (and especially in 2002), there were longer movies, with one movie in each of those years being particularly long.

5. Your frequency table would look like this for data up to 2008:

Length (in minutes) of 2000–2008 Best Picture Nominations												
90–99	100–109	110–119	120–129	130–139	140–149	150–159	160–169	170–179	180–189	190–199	200–209	210–219
//	////	𝍸	𝍸 𝍸	𝍸 ///	////	𝍸	///	///			/	
2	4	5	10	8	4	5	3	3	0	0	1	0

Your histogram would therefore look like this:

Length of 2000–2008 Best Picture Nominations



The shape of the histogram basically stays the same, except that the peak is now the 120–130 interval rather than the 130–140 interval, which remains unchanged. (However, this is not enough extra data to be able to say that the trend is now towards shorter movies.)

Pages 14–15 Whose News?

ACTIVITY ONE

1. a. Practical activity

b. A possible tally chart for this information is:

	North Island	South Island	International
Friday	𝍸 /	/	///
Monday	////	𝍸 /	
Saturday	𝍸 /	/	///

A frequency distribution is:

	North Island	South Island	International
Friday	6	1	3
Monday	4	6	
Saturday	6	1	3

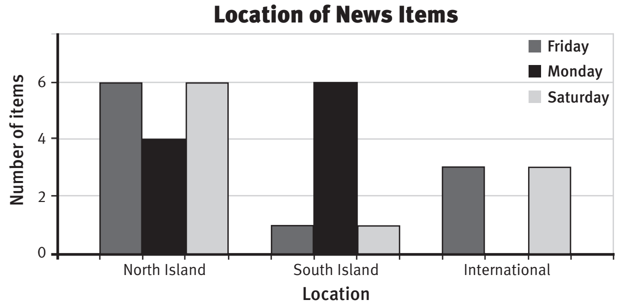
A percentage frequency distribution is:

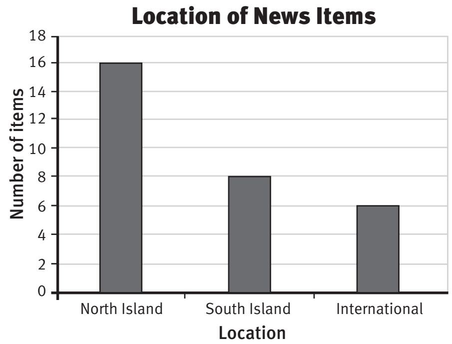
Location	Percentage of news items out of 30
North Island	53
South Island	27
International	20

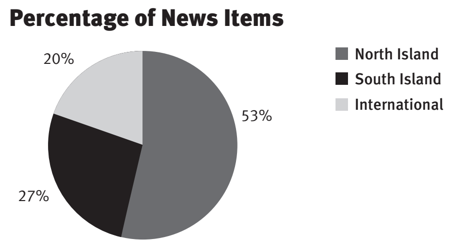
These could be combined as one chart:

	North Island		South Island		International	
Friday	𝍸 /	6	/	1	///	3
Monday	////	4	𝍸 /	6		
Saturday	𝍸 /	6	/	1	///	3
Total		16		8		6
Percentage	53		27		20	

Some possible graphs, which show relative proportions but in different ways, are:







You could also show these proportions as a strip graph.

c. Comments will vary. Both the North Island and the South Island get coverage in the first 10 items on all 3 days, but there were no international items in the first 10 items on Monday. The North Island (from the percentage figures and graph) gets more than twice as much coverage as the South Island or international news. The South Island has only one item in the first 10 on Friday and Saturday, but has more than the North Island on Monday (possibly weather-related).

2. Comments will vary. For the New Zealand items, there are more items for the North Island, mainly about events in Auckland and Wellington. Christchurch and Dunedin are the only South Island cities included as a single item on these nights. A small South Island town such as Fairlie may have featured as a bad weather/ snow story.

   If you take the perspective that we have two main islands, you may think that the news should be half and half. In that case, the news representation is not fair because the North Island has twice as much coverage as the South Island. On the other hand, given that there are 3 times as many people in the North Island as in the South Island, you could argue that there is an over-representation of the South Island because there are only twice as many stories for the North Island rather than 3 times as many.

3. If a major item rolled over more than 1 night, this might skew data collected on consecutive nights, especially if the story was an international one or one based in the South Island.

4. Most of the items in the first 10 are from the North Island, but there are a number of South Island ones as well. International items are represented less in the first 10 news items. International and South Island news items are often shown later in the news programme. If you were to rank the top 10 items (for example, 10 points for the first item through to 1 point for the last), the North Island outranks the South Island on every day. However, the smallness of the sample means that it is difficult to draw any conclusion about the relative importance of north versus south in relation to the order of items.

ACTIVITY TWO

1. Investigations and responses will vary.

2. The percentages that you mark on your map will vary depending on which year's data you use. Figures for 2008 are:

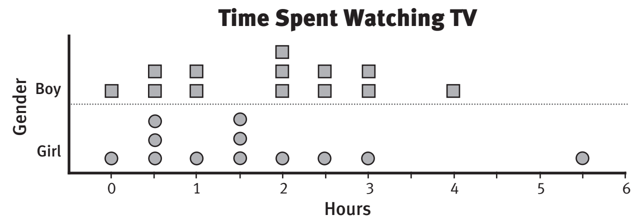
3. Results will vary depending on the results of your investigation in question **1**. (Note that the more data collected, the better for drawing conclusions.) For 2008, ask your teacher for the table and bar graph on pages 41–42 of the Teachers' Notes.

4. Answers will vary. The regions that appear to be invisible might be regions that don't contain a major city, that is, rural regions and regions with very small populations. Perhaps this is partly because most reporters live in the cities and it is easier to get stories from cities that include big business or government operations. However, if a major news story, such as a murder or a catastrophe, happens in a small area, the reporters and film crews get there as fast as possible!

Pages 16–17 Spin Doctor

ACTIVITY ONE

1.



2. Statements will vary. For example:
   - Most students watch between 30 minutes and 3 hours of TV per day.
   - One girl watches a lot more TV than most other students.
   - The time data is spread from 0 hours to $5\frac{1}{2}$ hours.
   - The spread for girls' and boys' TV watching times is similar.

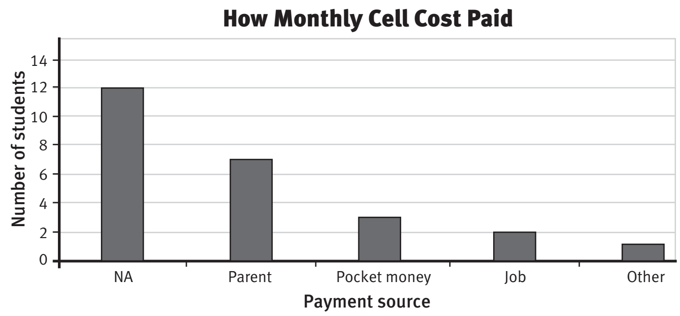
ACTIVITY TWO

1. Most of the headline statements have some basis in the data, but some of the interpretations are based more on opinion than fact.
   - **i.** Parents are cited as paying for cellphone costs more often than any other category (7 for parents, 3 for pocket money, 2 for job).
   - **ii.** Almost half the group ($\frac{12}{25}$) spent no money on cellphones (presumably because they didn't have one).
   - **iii.** The boys seem to watch TV more than the girls; 9 out of the 13 boys in the group have a TV in their bedroom (but only 6 out of the 12 girls).
   - **iv.** 8 of the 12 girls watched 1 hour or more of TV yesterday. However, the headline ignores the fact that 10 out of 13 boys also watched TV for 1 hour or more.
   - **v.** Only 2 students didn't watch TV on the day recorded, and one of these was a boy with a TV in his bedroom.
   - **vi.** The data does not support this statement – see **v** above.
   - **vii.** The mean time watching TV is greater for the boys than for the girls, but the difference is small. (The mean time for boys is 1.85 hours, and for girls it is 1.67 hours.)
   - **viii.** 1 girl watched 5.5 hours of TV over 24 hours (but this is only 1 day – perhaps she was sick in bed).
   - **ix.** 8 out of the 12 girls and 5 out of the 13 boys have a cellphone. Five of the 8 girls spend $20 or more per month on their cellphone. The mean spending for the girls with cellphones is $31.25 and for all the girls, $20.83; for the boys with cellphones, the mean is $16, and for all the boys, it is $6.15.
   - **x.** One boy had $40 of cellphone spending and a part-time job to pay for it. (Whether he is "exhausted" would depend on what the part-time job is!)

2. **a.–b.** Discussion will vary. The headlines often put an "emotional spin" on the data and don't incorporate all the information, so they can be misleading. There is not enough information given to justify any of the generalisations made (even for **ii**, as those without cellphones might wish they had them!).

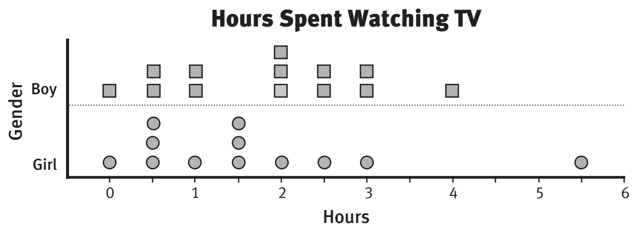
3. Answers will vary. Following are some examples:

   For "Parents foot bill for children's text habit":



   Conclusions:
   - Of the 13 students who spent money on cellphones, 7 of these had the bill paid by their parents. This is just over half of the students with cellphones.

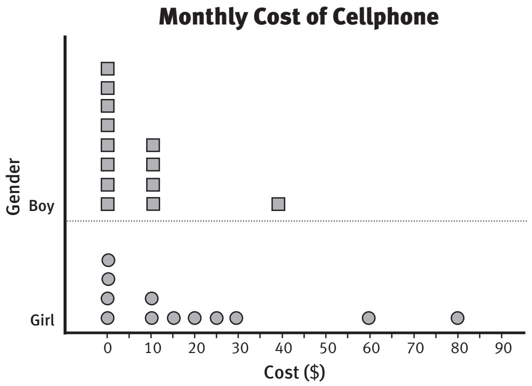
   For "Survey shows children need daily TV fix":



   - 67% of the girls (8 out of 12) watched at least 1 hour of TV yesterday.
   - 77% of the boys (10 out of 13) watched at least 1 hour of TV yesterday.

- 92% of the students (23 out of 25) watched some TV yesterday.
- Only 2 students watched no TV at all yesterday.

For "Girls fork out big bucks on cellphone habit":



How Monthly Cellphone Cost Paid

	Payment source					Gender	
	Job	NA	Other	Parent	Pocket money	Girl	Boy
Number of students	2	12	1	7	3	13	12
Mean	$25	–	$20	$32.86	$10	$6.15	$20.83

- More boys than girls don't have a cellphone.
- The girls' mean monthly cellphone spend is $20.83 compared with that for the boys of $6.15. The mean spend for the boys who have a cellphone is $16, and the mean spend for the girls who have a cellphone is $31.15.
- The mean spend ($32.86) for students whose parents pay the cellphone bill is higher than the mean spend for any of the other options.

4. Any of the data may be unreliable. Students may not tell the truth as to whether they have a TV in their bedroom, and they may not want to admit how much time they spend watching TV. Cellphone costs could vary greatly from month to month. If a parent is paying for the cellphone bill, students may not be fully aware of how much it costs each week. Students could only select one option for where their money came from. If it comes from more than one source, they've had to select just one.

INVESTIGATION

i. Practical activity. Results will vary depending on the data selected. Ask your teacher if you can look at the suggestions given in the Teachers' Notes for possible investigative questions using 2007 data. They include:

- I wonder what amounts girls typically spent on cellphones each month?

ii. The more data collected, the more accurate the conclusions are.

Pages 18–19 Free CD!

ACTIVITY ONE

1. a. No. Boxes contain CDs at random, so Pia could get 3 boxes with no CDs in them.

b. Yes, it is possible (but 3 CDs would be very unlikely!). (Note: there is a difference between "possible" and "guaranteed". See the answer for **Activity Two**, question **1a**.)

2. Practical activity. Results and comments will vary. Afterwards, you may want to tell Pia not to spend her money just in the hope of getting a CD (unless she really likes the cereal, in which case she might buy a few packets during the promotion). You may want to tell Zareb that you agree with him!

ACTIVITY TWO

1. a. Unless Pia knows exactly how many boxes are in the competition and buys more than two-thirds of them, there is no guarantee that she will win a CD. (For example, if there were 3 000 boxes in total in the promotion and Pia bought 2 000 of them, she might happen to pick the 2 000 boxes that don't win CDs. If she bought 2 001 boxes, she would only be *guaranteed* 1 CD. However, in reality, the more boxes she buys, the higher her *chance* of getting a CD. [In terms of spending money, there is no point in buying lots of the product in the hope of getting a CD if it means buying more of the product than you can use. You might be better just saving to buy the CD!])

b. Practical activity. Results will vary.

2. a.–b. Graphs and the features of the distribution will vary depending on the group results.

c. If you originally thought Pia would win a CD if she bought 3 boxes, you very likely will have changed your mind! There is no guarantee of a prize, and she would probably have to buy at least 5 boxes to win a free CD – and she still may not win!

Pages 20–21 Questionnaire Queries

ACTIVITY

1. All the replies give information, but it would be very difficult to collate that information meaningfully.
2. Answers will vary depending on interpretation of what having a school radio station would involve. For example, the cost of equipment and extras such as music CDs, the student and teacher time involved, and whether enough students would listen to make it worthwhile.
3. Answers may vary. The questions are not very good ones and could be tightened up as indicated below. A possible survey could include the following:
   - Year level
   - Gender
   - How often do you listen to the radio? (Days/Hours)
   - When do you mostly listen to the radio? (Day/Time of day)
   - Do you listen to the radio while you are doing your homework? (Rarely/Often)
   - What do you listen to on the radio?
   - Would you listen to a school radio station?

4.–5. Surveys and analysis will vary. Your survey will probably provide much more useful information than the one in the students' book. (Some of the suggestions above for question **2** go beyond a student survey.)

Pages 22–23 Gaming Choices

ACTIVITY

1. 1 (or 100%)
2. a.

		Wolves: will give fangs	Wolves: won't give fangs	Total
Rory	is stronger than the wolf	80	10	90
Rory	is weaker than the wolf	20	10	30
	Total	100	20	120

   b. 10

   c. For Rory to get the 2 fangs, he needs to overpower the wolf (so he has to be stronger) and the wolf has to give up his fangs. This is the top left box. $\frac{80}{120}$ = 67%.

   d. 80 of the 90 wolves that are weaker than Rory will give up their fangs; $\frac{80}{90}$ = 89%. 20 of the 30 wolves that are stronger than Rory will give up their fangs ($\frac{20}{30}$ = 67%) if they are defeated. So those that are weaker than Rory give up their fangs more easily.

3. a.

		Package: doesn't get broken	Package: gets broken	Total
Rory	delivers to the right person	58.5%	31.5%	90%
Rory	delivers to the wrong person	6.5%	3.5%	10%
	Total	65%	35%	100%

   b. 58.5%

4. The rabbit mission because he has 100% chance of succeeding. (He has a 67% chance of succeeding in the wolf mission and a 58.5% chance in the package mission.)
5. There is no simple answer to this question. The only mission that Rory can be sure to succeed with first time is the rabbit mission. But if speed really matters, Rory should probably try the wolf mission (see the answer to question **6**) because if he succeeds on the first or second attempt, he will have saved time. And even if it takes him 3 attempts before he succeeds, his decision won't have actually cost him time.
6. Based on time, Rory should choose the wolf mission because the probability of success is 67%, which is higher than the 58.5% for the package mission.

Pages 24 Advertising Claims

ACTIVITY

1. a. i. Because they are perceived as experts in this matter

   ii. Answers will vary. It could have been as few as 5.

   b. i. No

   ii. This product contains only 10% of the amount of fat that other (unnamed) brands do. (However, we don't know how much fat any of the products contain. The advertised product could still have more fat in it than health experts recommend.)

   iii. The advertiser might hope that people would think that the "other leading brands" would have 90% fat in them and that the advertised product is low in fat. "90% less fat" sounds healthy!

   c. i. In this case, it means the survey was not conducted by the manufacturers of Superfroth shampoo.

   ii. So that it sounds believable and unbiased

   iii. Answers may vary. We don't know what else was in the survey.

   d. i. The term "best seller" is used very loosely. It probably means that the book has sold a certain number of copies or more copies than other books have. (However, best-seller sales claims from trade associations are evidence-based.)

ii. Sales figures

iii. Not necessarily. The writing of many popular authors is not always of a high standard or well researched, although they may tell a good story. Some of the first books written by an author may be very good, but once an author is famous, their books may sell because of fame rather than quality.

iv. Many books could be, if "best seller" means a certain number of copies.

v. Answers will vary, for example, sales figures, the amount of promotion and advertising, whether it is a book by a well-known author (therefore eagerly awaited regardless of quality), whether the term is applied to the book by the publisher to promote sales, and so on.

e. i. Answers will vary. In this context, it could mean the current year's model.

ii. You can't tell from the statement.

iii. Answers will vary. If the survey is not an independent one, it could be very few and they could all be biased!

2. Critiques and discussion will vary.

Teachers' Notes

Overview Statistics in the Media

Statistics is defined in *The New Zealand Curriculum* as "the exploration and use of patterns and relationships in data". Like mathematics, it aims to equip students with "effective means for investigating, interpreting, explaining, and making sense of the world in which they live".

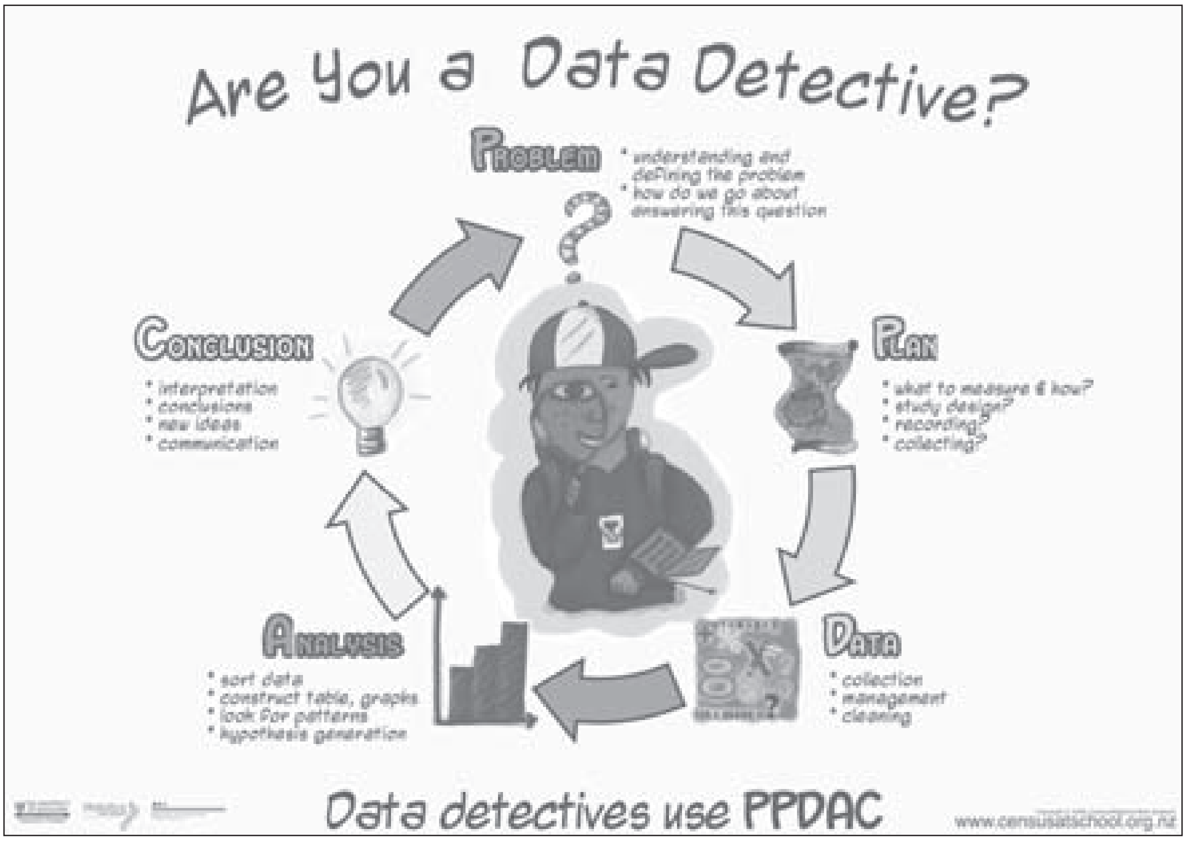
The New Zealand Curriculum goes on to say:

> Mathematicians and statisticians use symbols, graphs, and diagrams to help them find and communicate patterns and relationships, and they create models to represent both real-life and hypothetical situations. These situations are drawn from a wide range of social, cultural, scientific, technological, health, environmental, and economic contexts …
>
> Statistics involves identifying problems that can be explored by the use of appropriate data, designing investigations, collecting data, exploring and using patterns and relationships in data, solving problems, and communicating findings. Statistics also involves interpreting statistical information, evaluating data-based arguments, and dealing with uncertainty and variation.

page 26

The **PPDAC** (Problem, Plan, Data, Analysis, Conclusion) statistical enquiry cycle used for the New Zealand CensusAtSchool resources (see www.censusatschool.org.nz) provides a model for statistical investigation. This approach is used in this Figure It Out *Statistics in the Media* book and in the *Answers and Teachers' Notes* that accompanies it.

CensusAtSchool New Zealand makes available two posters (aimed at different age levels) for the PPDAC cycle. One version is:



The five steps (**PPDAC**) in this model are:

- **P**roblem – deciding what to investigate, and why, and how to go about it;
- **P**lan – determining how to gather the necessary data;
- **D**ata – collecting, managing, and preparing the data for analysis;
- **A**nalysis – exploring the data with the help of graphs and statistical tools and asking what it says;
- **C**onclusion – determining how the data answers the original problem and deciding what to do next.

CensusAtSchool New Zealand provides further information on all these steps in the form of a downloadable PDF (go to www.censusatschool.org.nz and select new curriculum, statistical enquiry cycle).

Much of the information in the following sections is adapted (with permission) from information available on CensusAtSchool New Zealand and Statistics New Zealand (www.stats.govt.nz).

TYPES OF DATA

Category data classifies data according to non-numeric attributes (variables), such as gender, colour, style, model, opinion, type, feel. For example, the category "foods" could be sorted into variables such as meat, fish, vegetables, fruit, and cereal.

Numeric data (which is sometimes referred to as **measurement data**) classifies data according to an attribute that can be counted or measured. Numeric data may be either discrete or continuous. **Discrete data** is whole-number, countable data, for example, the number of students in a class. **Continuous data** (also sometimes referred to as measurement data) is data obtained using measurement, for example, time, height, area, mass, age. (When continuous data is rounded to the nearest whole unit, it is effectively treated as discrete.) **Time-series data** is data that is collected from a series of observations over time, with a view to discerning time-related trends, for example, monthly profits, daily rainfall. It is usually numeric data.

A **variable** is an attribute that can vary or take on different values, for example, gender (male or female), time, colour, length, favourite author, number of items, cost, age, temperature. When data on two or more variables is collected for each item or person, the result is a **bivariate** or **multivariate data set**. A list of movies by length is a **univariate data set** (that is, there is a single variable, time); a list of movies by length, genre, and country of origin is a multivariate data set. Multivariate data sets have much greater potential for exploration than univariate data sets.

GRAPHS

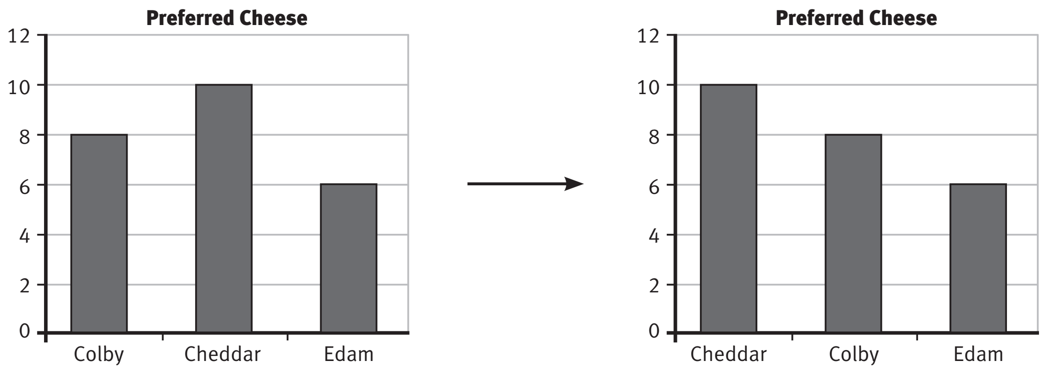
Graphs or charts? *Graph* is the more common usage in New Zealand (except in the case of pie *chart*) but *chart* is the term used by most graphing programs. In statistical contexts, these two terms are used virtually interchangeably.

The activities in the students' book promote graphs as a means of exploring data and communicating findings. It is important that students learn to make and/or "read" graphs, question what they read or see, find the stories in the data, and ask more questions. At any stage, students can devise their own graphical representations, but as they learn more, they need to become familiar with the standard types of graphs and associated conventions. All standard graphs should have a title that states the intent, **axes** (if used) should be labelled clearly, and the measures used should be consistent throughout. The use of these basic conventions enables the stories in graphs to be "read" more clearly. The following descriptions of graphs relate to those that occur in the students' book or that students may use in their answers and investigations.

Bar graphs are used to show the frequency of category data or discrete numeric data. Unlike dot plots and strip graphs, they have two axes, one labelled with the category and the other with the frequency. In basic bar graphs (see pages 15, 16, and 46 of the *Answers and Teachers' Notes*), there is always a gap between bars, showing that the categories are quite separate. The bars are normally vertical and, for category data, may be coloured or shaded differently as long as the colours or shades relate to separate category variables (that is, the use of different colours or shades must indicate differences).

On a well-constructed and labelled bar graph, it is easy to see which of the categories is most "popular" and to compare categories. Differences that appear insignificant in a pie chart or strip graph typically show up clearly in a bar graph.

Unless there is a good reason not to, the bars for category data are usually arranged in order of height, as in the example below.



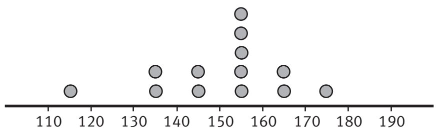
A bar graph can also be used to show multiple data sets. This is sometimes called a **clustered bar graph**. In effect, two or more bar graphs are sharing a common set of axes. The bars for the second variable are displayed side by side without a gap (for examples, see pages 8, *9*, and 35 of the *Answers and Teachers' Notes*). The two variables must be coloured or shaded distinctly and a key provided. This type of graph allows comparisons to be made both within and between data sets. Bar graphs can also be used to show how a numeric quantity (such as mean height) relates to a category variable (such as gender).

In a **stacked bar graph** (see page **4** of the students' book and page 13 of the Answers), bars are divided into sections representing each data category. In a **100% stacked bar graph**, as in the sectors in a pie chart, each section is shown as a percentage of the total of that category in the data. However, 100% stacked bar graphs have two advantages over pie charts: they use a percentage scale, which makes it easier to see the extent to which each category contributes to the whole, and their shape means that two or more stacked bars can be placed side by side so that data sets can be compared. Where two or more 100% stacked bars appear on the same graph, they are always the same length, regardless of the size of the groups being studied.

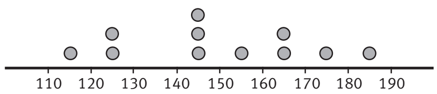
Pie charts (see page 15 of the Answers) and **strip graphs** show the relative size of the categories that make up a whole (whatever the whole may be). The categories are always labelled. The percentage value (and sometimes the actual data value) may also be shown on or alongside each region. Unlike bar graphs, pie charts and strip graphs do not show categories that contain zero data. Students find pie charts difficult to create by hand but easy to create in most graphing programs.

While they have their place, pie charts and strip graphs can only be used for a single variable and can only tell the simplest of stories. (Some statisticians suggest that pie charts should be avoided because angles are harder to relate to quantities than lengths are.)

Dot plots (see pages 16–17 and 37–38 of the *Answers and Teachers' Notes*) are a variation of the histogram. They are very easy to construct and clearly show the distribution of the data involved (that is, the way in which it is distributed and/or grouped). Dot plots suit discrete numeric data: each dot represents a single piece of data. Continuous (measurement) data is normally rounded to the unit used on the scale (for example, the nearest centimetre). The beginning and end of the scale are dictated by the least and greatest data value. Data can be grouped, as in the following dot plots.



Girls' Arm Span Length

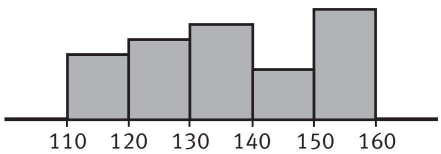


Boys' Arm Span Length

Stem-and-leaf graphs, explored on page **2** of the students' book, are a convenient means of organising and displaying discrete numeric data. Each individual data value retains its identity at the same time as overall patterns emerge. For further information on making and sorting stem-and-leaf graphs, see the section on graphs in the notes for *Statistics: Revised Edition*, Figure It Out, levels 3–4 (available online at www.nzmaths.co.nz/node/1992).

A **histogram** (see pages **2** and **12** of the students' book) looks similar to a bar graph but, in this case, the bars touch. Histograms are used for continuous data (for example, height).

In a histogram, the horizontal axis is a continuous number line. Its start and finish are defined by the data (there is no point starting at 0 if the first data value is, for example, 112 and the last is 158). The sides of the bars represent the limits of an interval. The bars are best labelled using the outer limits of each interval:

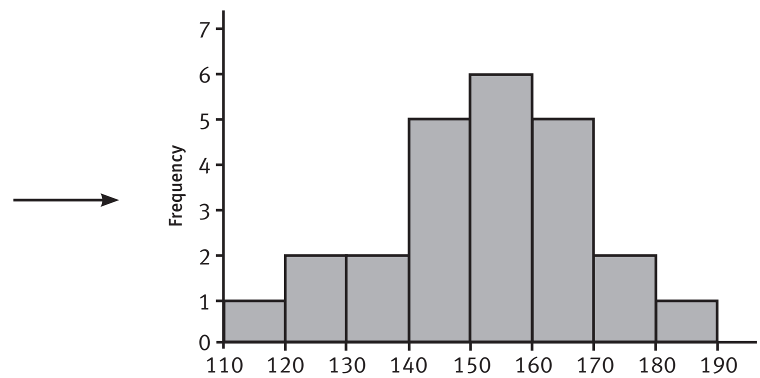


As a guide, 10 or 12 is a reasonable number of bars. If you want to communicate only the major features, you can use fewer bars, but if you want to find and communicate all (or most) of the detailed structure in the distribution, you may need lots of bars.

When deciding on a suitable interval, consider both the greatest and least values and the total number of values in the data set. Use "natural" steps, such as 2, 5, 10, or 20 (not steps of 3, 6, 7, 8, 9, and so on) and keep them the same throughout. As an example, the arm span data in the data set for the dot plots above ranges from 118 to 185 centimetres. There are 24 values in the data set.

This suggests a histogram with 8 intervals of 10 centimetres each. (Note that 110–120 is generally accepted as meaning from 110 up to, but not including, 120. 120 is included in the next interval.)

Arm span (cm)	Frequency
110–	1
120–	2
130–	2
140–	5
150–	6
160–	5
170–	2
180–190	1

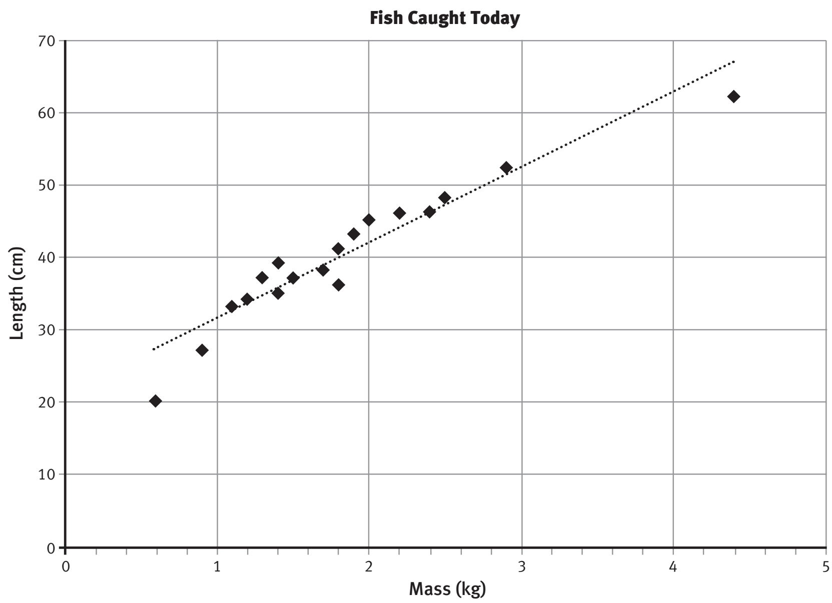


Histograms and bar graphs may show percentage (or proportional) frequencies instead of the frequencies themselves.

Line graphs relate to two variables, one plotted on the horizontal axis and the other on the vertical. Line graphs are useful for showing one variable in relation to another (for example, your height over time) or making predictions about the results of data that has not yet been decided or recorded (**extrapolations**), for example, visitor numbers to Abel Tasman National Park for the summer of 2012, based on previous visitor numbers for the same season and the general trend visible in several years' worth of data. Line graphs are most useful for displaying numeric data or information that changes continuously over time, so they are not suitable for category data.

A **time-series graph** is a line graph in which time is shown on the horizontal axis and the variable being observed is shown on the vertical axis.

A scatter plot graphs bivariate data (data with two numeric variables) as a series of separate points, as in the example below (see also page **3** of the students' book and pages 4, 13, and 35 of the *Answers and Teachers' Notes*). The horizontal axis shows one variable, the vertical axis the other. Some graphing programs call this kind of graph an XY scatter graph. Scatter plots are essential for showing the relationship between the two variables graphed. Students can use a scatter plot to look for clusters, outliers, trends (the shape may be straight or curved in some way), and changes in the vertical spread. It is sometimes useful to pencil in a curve or a straight line and then assess how this will fit the points. A computer can fit a line or various sorts of curves. In the example below, the computer-generated line gives less information than a curve would.



OTHER STATISTICAL TERMS

Axes (singular: **axis**) are the two lines, one horizontal and one vertical, that form the framework for most graphs. If frequency is one of the quantities mapped, it usually goes on the vertical axis.

Bias in statistics exists when the results differ from those that would reasonably be expected (for example, larger or smaller) due to some factor that was overlooked when data was being gathered (for example, a poorly worded survey question that was misunderstood by many respondents or a sample of the population that excluded some relevant age groups).

Collate means to collect and combine.

A **correlation** is said to exist between two variables when there appears to be some kind of relationship between them (for example, smoking and heart disease or latitude and temperature).

Data cleaning is the identification of incomplete, incorrect, inaccurate, or irrelevant parts of the data and the replacement, modification, or deletion of this data so that the remaining data represents its respondents as well as possible.

A **data display** is any way of displaying data, for example, a table or a graph.

To **extrapolate** is to go beyond the available data and make an educated prediction about what will happen "off the edge of the graph". (For example, using population data for the past few years, a reasonable prediction could be made for New Zealand's population next year or in 5 years.)

To **interpolate** is to estimate a value that lies somewhere among or within known data values. For the purposes of both interpolation and extrapolation, it is assumed that the observable pattern continues within and will continue beyond the available data. This will not necessarily be true. Extrapolation is generally less reliable than interpolation because there is no guarantee that a previous trend will be maintained.

A **frequency table** is a table that organises data by category or interval and gives the frequency for each category or interval. For example:

Weeks between haircuts	2	4	6	8	10	12	14+
Number of classmates	2	6	9	5	2	3	1

The number of data in a category or interval is known as its **frequency**. Frequency can be thought of as the count of individuals in a category.

"I wonder" questions are **investigative questions** – statistical questions or problems to be answered or solved. Investigative questions are questions asked of the data, whereas **survey questions** are questions that are asked to get the data.

Three types of investigative questions are of particular interest at this level:

- **Summary questions**, which usually involve a single variable and require the data to be described in some detail (for example, "I wonder how long it typically takes a year 6 student to run 100 metres?")
- **Comparison questions**, which involve comparing two or more subsets of data, for example, male and female, young and old, in relation to a common variable such as speed (for example, "I wonder whether year 6 girls are typically faster than year 6 boys?")
- **Relationship questions**, which are posed in order to look at the interrelationship between two paired numeric variables (for example, "I wonder if students who have an MP3 player are more likely to have their own cellphone as well?").

When continuous (measurement) data is grouped into a frequency table or histogram, the axis is divided into sections called **intervals**. An interval is defined by its limits, for example, "greater than or equal to 3 but less than 4". Intervals can be precisely and economically described using inequality signs and a symbolic variable, for example, $3 \leq d < 4$ (read [from the centre] as "distance is greater than or equal to 3 but less than 4").

The **mean** of a numeric variable (in a data set) is a way of defining the "centre" of a distribution (or set of values). It is the "centre of gravity" of a graph. The mean is calculated by dividing the sum of all the values in the data set by the number of values. The mean is sometimes called the **average**.

The **median** is the middle value in the distribution for a numeric value when all the data values are arranged in order from smallest to largest or largest to smallest.

The **mode** is the most commonly occurring data value (if there is one) in a data set.

When analysing any variable (for example, life expectancy) in a data set or distribution, it is important to consider both **central tendency** and **spread**. Central tendency refers to the extent to which data clusters around a "middle" value. The most common measures of central tendency are median and mean. Spread refers to the extent to which data is spread out or dispersed. Simple measures of spread include range and interquartile range, both of which can be clearly seen in a box plot.

An **outlier** is an outlying value in a data set. It is a term that is often used in statistics. An outlier in a scatter plot is a point that is a long way from the rest. (It may be the result of a counting or measurement error, or it may be a record for an unusual individual.) Outliers can affect the average (or mean) quite considerably.

Population means the entire group that a particular investigative question relates to.

A **sample** is a subset of the population.

Probability (see pages **18–19** and **22–23** of the students' book):

- **Probability** and **chance** relate to the same concept, although one of the terms may be more usual in a particular context.
- **Trial**: performance of an action or actions where the outcome is uncertain (for example, 8 tosses of a coin)
- **Outcome**: the result of a trial (for example, 5 heads and 3 tails, where the trial is a toss of 8 coins)
- **Experiment**: sometimes used interchangeably with trial
- **Experimental probability**: the likelihood that something will happen, based on a number of trials
- **Theoretical probability (expectation)**: the likelihood that something will happen, based on reasoning or calculation from assumptions about the process.

A **questionnaire** is a form containing a set of questions designed to gain statistical information.

A **ratio** is a mathematical comparison between two numbers or quantities, indicating their relative (rather than absolute) sizes. A ratio can be expressed: in words ("2 of this to 3 of that", "2 out of 5"), using ratio notation (2:3), as a percentage (40%), or as a fraction ($\frac{2}{5}$). Ratios depend on their contexts for their meaning.

A **simulation** is the use of a mathematical model to recreate a situation, often repeatedly, so that the likelihood of various outcomes can be more accurately estimated. Many simulations (see the notes for pages **18–19** of the students' book) can be generated by calculator or computer.

Tally marks (/) are used when counting or categorising data by hand. Every fifth stroke is drawn across the previous four, facilitating skip-counting by 5s and 10s. For example, ~~////~~ ~~////~~ // stands for 12.

In a **tally chart**, information is presented in three columns: category, tally, and frequency (the tally total). For example:

Footware	Tally	Frequency
Shoes	~~////~~	5
Sandals	~~////~~ //	7

A **trendline** (see the example under scatter plot graphs) is a line on a graph that indicates a statistical trend. Some computer programs can assess the plotted values and mark a middle course through them. Any such line must always be assessed against the data to see how well (if at all) it fits the data.

Variation is the term used to refer to the differences among values of the same variable, particularly differences from an expected pattern or trend. Variation can be described and, at later levels, measured, using a variety of measures of spread, from the simple to the sophisticated. (Variation can be considered in Cheap Reads, pages **10–11**.)

Here is a simple example of variation in probability, involving coin tosses: If a coin is tossed a very large number of times, we would expect that the numbers of heads and tails would be approximately equal (because the two outcomes are equally likely). But in practice, if we were to toss a coin 100 times and then repeat this experiment 10 times, we would almost certainly get widely differing results, for example:

Experiment	1	2	3	4	5	6	7	8	9	10
H	41	44	52	53	42	49	47	50	38	45
T	59	56	48	47	58	51	53	50	62	55

Achievement Objectives

Achievement objectives in the Teachers' Notes are from the mathematics and statistics area of *The New Zealand Curriculum*. In the notes for each set of activities in the student book, the relevant Achievement Objective headings and steps of the **PPDAC cycle** (see pages 21–22 of the Teachers' Notes) are shaded in the box diagrams below the list of achievement objectives.

Key Competencies

The New Zealand Curriculum identifies key competencies that students will develop over time and in a range of settings. Schools can develop the key competencies within the mathematics and statistics learning area as well as encouraging and modelling values for students to explore.

The five key competencies identified in *The New Zealand Curriculum* are:

- thinking
- using language, symbols, and texts
- managing self
- relating to others
- participating and contributing.

The notes for the student activities in this *Statistics in the Media* book suggest one or more key competencies that relate to each set of activities and give suggestions for how these could be developed. (You may, of course, decide to focus on key competencies other than those suggested.)

Pages 1–4 Texting Olympics

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy		Probability
P	P	D	A	C

Key Competencies

Texting Olympics can be used to develop these key competencies:

- thinking: justifying and verifying, discerning if answers are reasonable, exploring and using patterns and relationships in data, and designing investigations
- using language, symbols, and texts: interpreting statistical information, communicating findings, using ICT as appropriate, interpreting visual representations, and demonstrating statistical literacy
- relating to others: sharing ideas.

Statistical Ideas

Texting Olympics involves the following statistical ideas: using the **PPDAC cycle***, **multivariate data sets**, **stem-and-leaf graphs**, **scatter plots**, and **histograms**.

The activities in Texting Olympics are good examples of ways to use multivariate data sets. Students at this level need to experience exploring statistics using more sophisticated techniques than those used with simple **category** or **univariate data**.

Activity One

Before the students start the activity, it may be useful to have them collect their own data relating to the Texting Olympics. Doing this will raise questions regarding the logistics of the data collection process and will make it easy for them to answer question **1**. The class data can later be graphed in the same way as the graphs produced for the supplied data and then compared. (Note: The data in this activity is real and was collected from year 9 students [names changed]. Also note that 1:54.94 is one way of recording 1 minute, 54.94 seconds. Watch for students who get confused by the "out of 60" for seconds and the "out of 100" decimal notation for part-seconds.)

* For bolded terms, see What Is Statistics About?, pages 21–27.

For question **2**, decide if you want the students to work out the seconds manually or whether, if computers are available, you want to challenge them to explore formulae to do the task for them. If they are to do the data conversion (and later, the graphs) on the computer, it would be sensible to have the data available for them as a spreadsheet file rather than each student entering the data separately. (Although this activity is particularly suited to the use of computers, students without that access can still make the required graphs, although the one for **Activity Three**, question **1a** may need to be simplified, as suggested later.)

The process of turning minutes into seconds in this activity is necessary before the students can analyse results and make **graphs**.

A stem-and-leaf graph, as shown in question **3**, is an excellent way for the students to display the data as they organise it. Encourage them to realise that it is more efficient to put the "leaves" into the graph in the order they appear in the raw data than to try to order the data before drawing up the graph, as they may otherwise miss some data out. However, as suggested in the student book for this activity, it is usually best to reorder the leaves of the stem-and-leaf graph into numerical order before the display is used for analysis.

Question **4** requires the students to analyse a different sort of graph (in this case, a histogram) based on the same data. The idea of making different graphs with the same data is that, sometimes, by regrouping the data you get a different or a more useful display. In this case, the intervals for the histogram are 2 seconds, whereas the stem-and-leaf graph, necessarily, has 1-second intervals. With a histogram, individual values are less visible, so the graph is more abstract. You could discuss the usefulness of each type of graph in the Texting Olympics context.

Activity Two

You could start with a discussion as to why the students in the data set provided were comparatively better in hurdles. This discussion may lead to new **investigative questions**. Your students may wish to repeat the investigation with different people doing different events first to see if results are similar.

If students struggle to understand how the scatter plot works, have them plot some points by hand. If they timed themselves for these events (as suggested in the notes for **Activity One**), they could locate where their own results would be placed.

Activity Three

Although this activity can be done by hand, the students will find it much easier to use a computer spreadsheet program, especially if they have already entered the data for **Activity One**, question **2** into a spreadsheet. A stacked bar graph, as shown in the Answers, clearly shows the order of the students overall, and you can also see where individual event times were faster or slower. (Students without computer access could make a stacked bar graph by hand, but it might be simpler if they just added up the 3 times for each student and then put all the total times in order.)

The different way of selecting a team proposed in question **2** should lead to some interesting discussion and help the students to realise that data can be analysed in different ways, each of which will have advantages and disadvantages.

Investigation

Whenever students do statistical investigations, it is important to revisit the **PPDAC cycle**. Students tend to have most difficulty in converting from an idea or problem to a statistical question. The planning phase is arguably the most vital. Students need to decide what they will measure and how they will measure it.

Extension

Students could:

- do a parallel investigation activity with real athletic events.
- investigate how long they can balance on their right foot and then on their left foot (or other timed activities).
- investigate the links between typing speed and cellphone texting speed or texting speed to cellphone bill size – the possibilities are endless!

Pages 5–7

Singing Star

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  – determining appropriate variables and data collection methods;
  – gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  – comparing distributions visually;
  – communicating findings, using appropriate displays (Statistics, level 4).

Statistical investigation	Statistical literacy	Probability

P	P	D	A	C

Key Competencies

Singing Star can be used to develop these key competencies:

- thinking: discerning if answers are reasonable, hypothesising, thinking critically, and making decisions
- using language, symbols, and texts: interpreting statistical information, capturing thought processes, interpreting visual representations such as graphs and diagrams, and demonstrating statistical literacy
- relating to others: sharing ideas and understanding others' thinking.

Statistical Ideas

Singing Star involves the following statistical ideas: using the **PPDAC cycle***, **multivariate data sets**, **bar graphs**, **medians**, and **means** and using statistics for decision making.

ACTIVITY ONE

Before the students begin this activity, you could discuss with them a current or past TV show that has a talent quest format. The students may have noticed that different judges have different judging personalities or varying consistency with their scoring. You could also discuss the relative merits of using a total, the median, or the mean (see the definitions on page **6** of the students' book) of the judges' scores to decide the winner and which would be fairer in the case of the TV show they are discussing. Alternatively, you could have the second part of this discussion after the students have completed Singing Star.

For question **1**, discuss with the students whether the bar graph gives any more information than the raw data. The purpose of a graph is to make data more meaningful and to give a picture that may not be apparent when numbers alone are reported. If a graph does not provide a better picture, it shouldn't be used. In this case, although the graph is difficult to read in terms of the contestants, it does show patterns in terms of the judges' scoring that are not so apparent from the table of raw data. (Question **2b** shows how the bar graph can be enhanced to be more useful in terms of the contestants. This is a useful learning point for the students to keep in mind before they decide not to use a certain sort of graph.) Perhaps you and your students could create graphs that reveal the story better. There are several possibilities.

For question **2b**, make sure that the students understand the definitions given for median and mean. The example showing the use of colour and a bar on the student page is one way to record this information. Students may decide to show both the mean and the median as lines on the graph, using different colours, or to place symbols on the graph to represent the median and mean figures.

Question **3** provides an opportunity to explore the functions of the median and the mean. They are both measures of **central tendency** (numbers used to give an idea of "middle"). In instances such as this, where each contestant has their own data set, a convenient first step when comparing them is to compare their "middles". This would also be true for comparing large data sets.

* For bolded terms, see What Is Statistics About?, pages 21–27.

The median of the 3 scores may be the best measure of a contestant's skill because it minimises the effect of **outlier** judges. However, a disadvantage of the median is that it is blind to all numbers except for the one in the middle: the median of {1, 1, 3, 3, 3, 4, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9} is exactly the same as the median of {1, 1, 3, 3, 3, 4, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 21}, yet the range of numbers is quite different because of the difference between the two numbers at the higher end. The advantage of the mean is that it does take account of every number in a data set (generally, if any number in a data set is changed, this will affect the mean). Its main disadvantage is that it is unduly influenced by outliers (as in the second range above).

The students may come up with some novel methods for ranking the students in question **3b**. Methods using the median and the mean are given in the answers. Discuss with your students the advantages and disadvantages of mean and median as a basis for ranking the contestants. They may want to change their minds about their chosen ranking system!

Questions **3c**, **4**, and **5** focus on elimination of contestants – a hotly contested part of any show of this type! Various choices and reasons for them are provided in the answers. You could give the students the following situation to consider: *What if Ethan's scores were 3, 4, 5? His median would still be 4, but his mean would be 4 as well. Therefore, Ariana and Ethan would both have the lowest medians and means. Who should be eliminated?* (In this situation, it might be better to ditch each person's worst score and find the mean of the other two scores.)

Extension

Students could investigate the judges' scoring rather than the contestants' results. The students will need to make the judges the subject of a graph (see the answers for **Activity One**) or rank contestants by each judge and compare lists. Have them look at how the judges' opinions differ – is there a "nasty" judge? Any crossover in the graph may show that a judge dislikes a particular person.

ACTIVITY TWO

Students may be interested in exploring whether the producers of the show are more interested in finding talent or making money. High viewer numbers mean more advertising revenue, so a contestant who gets a bigger audience is more valuable than the others, regardless of talent or judges' scores. Data is important in decision making – but the right data needs to be collected! (You may need to suggest to the students that they look again at the performance order of contestants given on page **5**.)

Question **2** should provide a good discussion point about what makes a "good" graph. Encourage the students to look critically at their own graphs to see if they can find better ways to tell the stories involved.

Extension

Data could be collected from your school speech competition (or other judged event) and analysed in a similar way.

Students could do a food-tasting experiment – judging three different flavours of a particular food. Data could then be analysed similarly.

Pages 8–9

Billboards

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy	Probability	
P	P	D	A	C

Key Competencies

Billboards can be used to develop these key competencies:

- thinking: exploring and using patterns and relationships in data, and making decisions
- using language, symbols, and texts: communicating findings, using ICT as appropriate, interpreting visual information such as graphs and diagrams, and demonstrating statistical literacy
- relating to others: sharing ideas and communicating thinking
- participating and contributing: working in pairs with both people participating.

Statistical Ideas

Billboards involves the following statistical ideas: using the **PPDAC cycle***, evaluating statements made by others, and using **multivariate data sets** and **clustered bar graphs**.

Data collection must always have a purpose. Traditionally, students in statistics classes have often been sent outside to count cars. Billboards provides a meaningful situation in which someone might need to collect data about traffic flows.

Activity One

A good starting point would be to relate these questions to the students' own world. For each billboard, ask: *Who do you know who would be interested in this billboard?* and *Where in our school neighbourhood would you want to place this billboard for these people to see it?*

Activity Two

Encourage your students to always ask questions of data that is presented to them (even if there are no firm answers). For example, they might ask: "How was the data collected?" "What about between 7 a.m. and 8 a.m.?" "Why is the last category more than 1 hour?"

When the students are checking the statements that other students have made, remind them to make their criticism fair and polite. If they think a particular statement is not firmly based on the data, they could discuss with the classmate what they did base the statement on and whether that data could be interpreted differently.

* For bolded terms, see What Is Statistics About?, pages 21–27.

For question **2**, students are making recommendations on behalf of Supa Signs. However, you could also encourage them to write a report in the format Supa Signs might give the owner of each product. They can use the graphs if they feel they lend weight to the recommendations. As well as the data and findings, the report may include factors such as visibility, busy roads (so drivers need to concentrate on their driving), or anything else is in those areas competing for driver, passenger, or passer-by attention.

Extension

Students could design their own billboards for a product targeted at schoolchildren. They could then examine traffic and pedestrian flows near the school and use the data to select an appropriate location. Alternatively, students could examine traffic and pedestrian flows for the location of an external school sign/noticeboard.

Pages 10–11 Cheap Reads

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4)

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy		Probability
P	P	D	A	C

Key Competencies

Cheap Reads can be used to develop these key competencies:

- thinking: designing investigations and thinking flexibly
- using language, symbols, and texts: interpreting statistical information and communicating findings
- relating to others: listening actively
- managing self: working independently, planning, and persevering.

Statistical Ideas

Cheap Reads involves the following statistical ideas: using the **PPDAC cycle***, evaluating statements made by others, posing **investigative questions**, using **multivariate data sets**, and investigating **samples** and **population**.

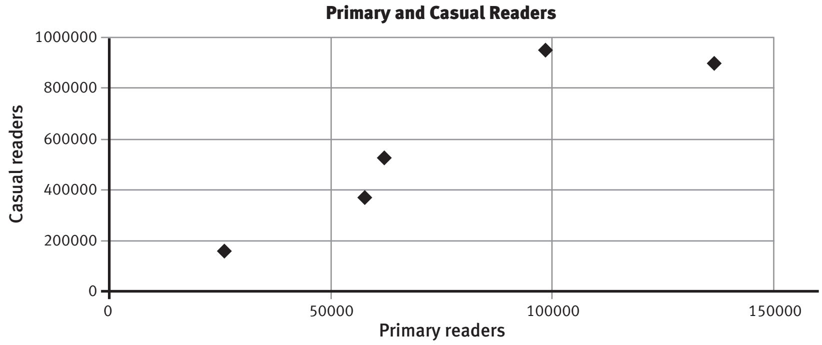
ACTIVITY ONE

Before the students answer question **1**, discuss with them why and how magazines might want to use the data about readership numbers. You may need to point out to them that the largest part of a magazine's revenue is from advertising rather than from the cover price. Encourage them to work out for themselves that larger readership numbers means a magazine can attract more advertisers or charge them more. Magazines would therefore want to paint a favourable picture of their readership numbers and can sometimes use readership data to help them do this.

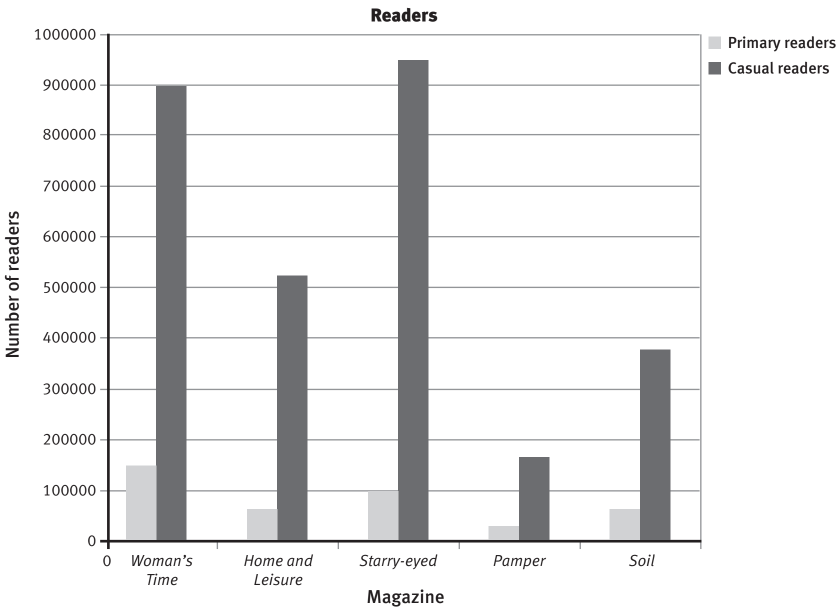
* For bolded terms, see What Is Statistics About?, pages 21–27.

Have the students go on to look at the readership table in question **1** in the light of this discussion. For example, *Woman's Time* would want to focus on the primary readership figure (which represents actual sales), whereas *Starry-eyed* would want to focus on overall readership, including casual readers who don't buy the magazine (advertisers are still interested in this figure). In each case, one of the magazines has the largest figure and thus would want to emphasise the point with potential advertisers.

For question **2**, the students can make comparisons simply by looking at the numbers, but they could also make a **scatter plot** to see how closely related primary and casual readership are.



A bar graph is an alternative way of highlighting the variation in casual readership.



If the students have access to the Internet, a search on "magazine readership New Zealand" gives information that may help them answer the various parts of question **3**.

ACTIVITY TWO

This activity can be done as a precursor to the **Investigation** or you may decide to combine the two. Your students may prefer to go straight to collecting new data using real magazines that they are familiar with. You could conduct a focus-group survey of a number of familiar magazines and record the data in a similar manner to that in the activity. 20 students were involved in the focus-group survey in this activity, so it may be useful to use the same number of students for comparison purposes.

Questions **1 a–c** can also be asked of your class's own data.

For question **2**, a discussion about how samples can be used to make predictions about the wider population would be useful. It is important to discuss with the students what a representative sample is; the samples from your own class (or the one used in the student text) are not representative of all students because the students in the sample form a very particular demographic.

For question **3**, note that Sarah or your own students can only make conclusions based on their sample groups. If Sarah follows the PPDAC cycle, she may now have new questions to investigate, based on the data she has already collected and analysed. For example: "Is there a difference between what magazines girls and boys read at this age?" Brainstorm with your students possible follow-on investigative questions.

Extension

Here are the 2008 advertising prices for spreads in a popular New Zealand magazine:

Double page	$15,000
Full page	$8,000
Half page	$5,000
Third page	$4,000
Quarter page	$3,000
Inside front cover	$18,000
Outside back cover	$10,000

Using these prices as a benchmark, your students may like to examine the total advertising revenue for magazines that they are familiar with. They could compare how much advertising revenue is contained in the first 10 pages and how much in the last 10 pages and discuss any differences.

Students could also get two copies of the same magazine and cut out all the advertising, taking note of the page number. They could make a big display of this, calculate the area and percentage of the page taken up in ads, and then make a line graph (using the page numbers on the horizontal axis).

Information on readership may be found on media research websites, which often provide current figures on categories such as magazine and newspaper readership.

INVESTIGATION

See the comments for **Activity Two**.

Pages 12–13

Blockbusters

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy		Probability
P	P	D	A	C

Key Competencies

Blockbusters can be used to develop these key competencies:

- thinking: exploring and using patterns and relationships in data, and making decisions
- using language, symbols, and texts: interpreting statistical information, interpreting visual representations such as graphs and diagrams, and communicating findings.

Statistical Ideas

Blockbusters involves the following statistical ideas: using the **PPDAC cycle***, **multivariate data sets**, **dot plots**, **histograms**, **tally charts**, **stacked bar graphs**, **means** and **medians,** and evaluating statements made by others.

The following activities are an excellent example of the ways in which graphs can be used to analyse data. The movie lengths do not tell a story when looked at as a series of numbers, but by experimenting with different graphs, a number of different pictures can be painted. For example, the two histograms showing 1980s films and 2000s films appear to show an increase in movie length overall. Looking at the mean or median nominated film length for every year from 1980 to now would possibly show a different picture.

activity one

Encourage your students to look at the shape of the graph when they note the features. It is more useful to look at a larger interval than simply at the longest or shortest bar when making statements.

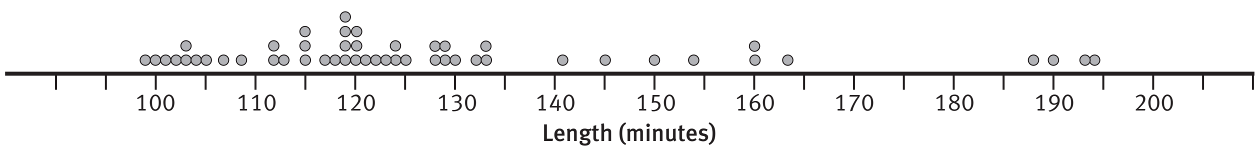
Ask: *We could estimate a median of 125 from the histogram, but 123 would be better. Can you see why?* (There are 50 films, so the middle, or median, is between the 25th and 26th values in the 120–129 interval. Although 124.5 is the middle of that interval, the 25th and 26th values are the 4th and 5th values out of the 14 values in the interval. This is about one-third of the way through the interval, so 123 might be a better estimate than 125.)

For question 3, the students are not required to make a graph. In case you want them to do so, here is the 1980s data (which could also be used later for comparison in **Activity Two**):

Year	Title and Length (in minutes) of 1980s Best Picture Nominations				
1980	*Coal Miner's Daughter*	*The Elephant Man*	*Ordinary People*	*Raging Bull*	*Tess*
	125	124	124	129	190
1981	*Atlantic City*	*Chariots of Fire*	*On Golden Pond*	*Raiders of the Lost Ark*	*Reds*
	104	123	109	115	194
1982	*ET*	*Gandhi*	*Missing*	*Tootsie*	*The Verdict*
	115	188	122	119	129
1983	*The Big Chill*	*The Dresser*	*The Right Stuff*	*Tender Mercies*	*Terms of Endearment*
	105	118	193	100	132
1984	*Amadeus*	*The Killing Fields*	*A Passage to India*	*Places in the Heart*	*A Soldier's Story*
	160	141	163	112	101
1985	*The Colour Purple*	*Kiss of the Spider Woman*	*Out of Africa*	*Prizzi's Honor*	*Witness*
	154	120	150	130	112
1986	*Children of a Lesser God*	*Hannah and Her Sisters*	*The Mission*	*Platoon*	*A Room with a View*
	119	103	120	120	117
1987	*Broadcast News*	*Fatal Attraction*	*Hope and Glory*	*The Last Emperor*	*Moonstruck*
	133	119	113	160	102
1988	*The Accidental Tourist*	*Dangerous Liaisons*	*Mississippi Burning*	*Rainman*	*Working Girl*
	121	119	128	133	115
1989	*Born on the Fourth of July*	*Dead Poets Society*	*Driving Miss Daisy*	*Field of Dreams*	*My Left Foot*
	145	128	99	107	103

A possible dot plot, showing the data in minutes, is:

Length of 1980s Best Picture Nominations

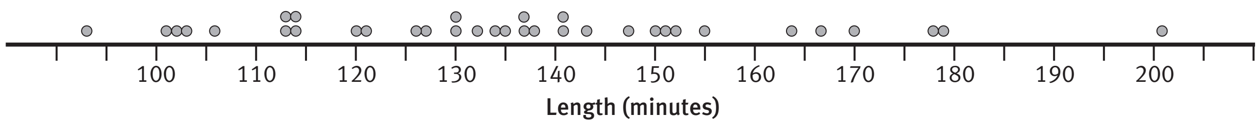


* For bolded terms, see What Is Statistics About?, pages 21–27.

ACTIVITY TWO

Although the students are asked to make a histogram to compare with the one in the students' book, also encourage them to play around with different ways of graphing this data, for example:

Length of Best Picture Nominations 2000–2006



To create a stacked bar graph using a computer spreadsheet, as shown in the answers for question **4**, the students will need to order the data:

	A	B	C	D	E	D
1	2000	155	147	130	121	120
2	2001	178	137	135	130	127
3	2002	179	167	150	114	113
4	2003	201	141	138	137	102
5	2004	170	152	132	126	106
6	2005	164	134	114	113	93
7	2006	151	143	141	103	101

When first creating the graph, select only the movie lengths (shaded). This is best done as a stacked column graph, rather than a bar graph, so that you can add the year labels as category labels on the x axis.

A useful exercise would be to have the students calculate the total length of the 5 nominated movies for each year {673, 707, 723, 719, 686, 618, 639} and the mean length of the nominated movies in each year {135, 141, 145, 144, 137, 124, 128}. Using this information, they will see that the total and mean lengths were greatest in 2001–2003. Further, they should see that this was because the three very long Lord of the Rings movies (*The Fellowship of the Ring, The Two Towers,* and *Return of the King*) featured in those years. The mean length for each of the years 2000–2004 was between 135 and 145, a range of just 10 minutes. For the years 2005–2006, the mean length was 124 and 128, so clearly the movies for these latter 2 years were shorter overall than for the previous 6 years. However, this information is not enough to make any assumptions about trends.

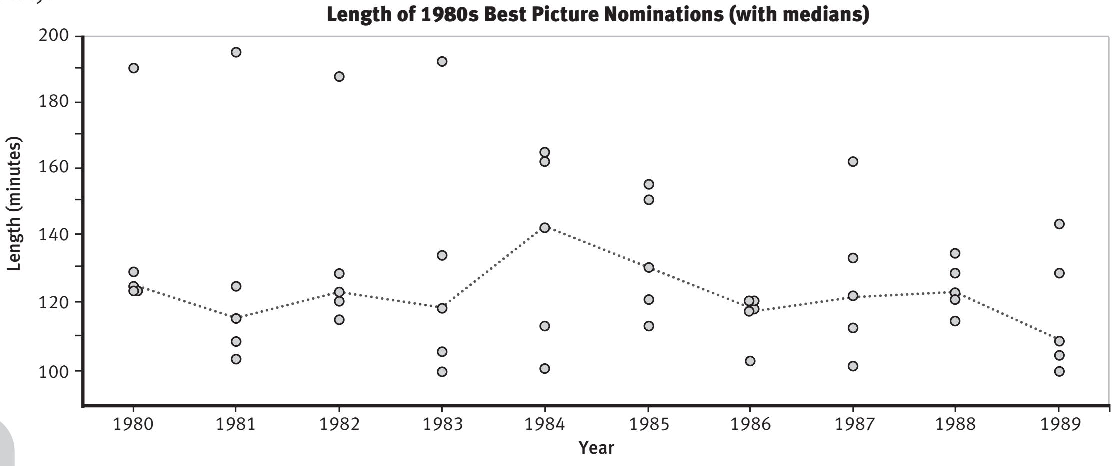
As an extension to question **5**, later years than 2008 can be accessed on the Internet when available. Nominations for all available years can be accessed at www.oscars.org/awardsdatabase/index.html
To find the movie length, do an Internet search on the movie name; the best option is the Wikipedia (film) version, where each movie listed has a box on the right with key information about the movie, including length.

Extension

The students could do a longitudinal study of the change in movie lengths over a period of time and turn the data into graphs such as those shown in the answers for this activity (but over a much longer time period).

The line on the dot plot below shows the median for each year in the 1980s (see the information in the notes for **Activity One**).

Length of 1980s Best Picture Nominations (with medians)



They could discuss what including the director's cuts would do to the length of film data.

To find the information about the Best Pictures for a particular year or range of years, go to the website www.oscars.org/awardsdatabase/index.html. For example, to get the 1990s information, use the Basic Search, select Best Picture in the Award Category and 1990 to 1999 in Award Years. This gives a complete list of all the movies that were nominated for Best Picture during those years.

Students may also be interested in doing some research on the genre of the winner of Best Picture. One site that had summary information (as a percentage) from 1927 to 2001 (at time of writing) was www.filmsite.org/bestpics2.html

Pages 14–15 Whose News?

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical investigation	Statistical literacy	Probability		
P	P	D	A	C

Key Competencies

Whose News? can be used to develop these key competencies:

- thinking: exploring and using patterns and relationships in data, and designing investigations
- using language, symbols, and texts: communicating findings using visual representations such as graphs
- relating to others: collaborating
- participating and contributing: working in groups with everyone contributing, and sharing equipment and resources.

Statistical Ideas

Whose News? involves the following statistical ideas: using the **PPDAC cycle***, **investigative questions**, **multivariate data sets**, **tally charts**, **bar graphs**, and **pie charts**, and doing a statistical investigation.

Activity One

As an introduction, you could have your students watch the news for 1 night and note the location of the first 10 items. This process could raise questions about the data collection; it is good to encourage this sort of questioning by students. The data collected could be very different at different times of the year, such as during periods of unusual weather phenomena, major sporting events, or during election time.

Question **1b** involves the students in creating various charts and graphs. You may need to give your students some help with these. Examples of these are shown in the Answers.

* For bolded terms, see What Is Statistics About?, pages 21–27.

The ranking system mentioned in the answers for question **4** takes into account the importance an item may have based on its placement in the news order. It highlights the differences between North and South Island distribution perhaps more fully than a simple tally chart does. Here is a points table based on the system of 10 for first news item and 1 for the 10th. The spread for this small sample is not an even one; the students may have ideas about why this might be. (In the data gathered for the students' book, the occurrence of snowstorms in the South Island on the Monday was a big news item.)

Points	Friday	Monday	Saturday
10	Dunedin	Auckland	Auckland
9	Australia	Otago	Auckland
8	Wellington	Otago	Auckland
7	Rotorua	Wellington	Tauranga
6	Wellington	Wellington	Auckland
5	Wellington	Wellington	Wellington
4	Auckland	Christchurch	Christchurch
3	Whakatane	South Island	UK
2	Texas, USA	Christchurch	USA
1	UK	Fairlie	India
Totals	NI = 33 points SI = 10 points	NI = 28 points SI = 27 points	NI = 45 points SI = 4 points

ACTIVITY TWO

This activity provides an opportunity to work through the PPDAC cycle, with the students focusing mainly on the problem, planning, and data stages of the cycle. You will need to allow the students time to work through this.

Problem

Examples of the types of investigative question that students could pose are: "Where do news items on the 6 o'clock news on TV1 typically come from?" "Where do news items on prime-time TV in New Zealand typically come from?"

Important features of the investigative questions include specifying which news slot and which channel and that the location of the news item is of interest.

Plan

The students need to:

- make decisions about what information they are going to collect;
- decide how they will categorise the location, for example, is the news item about the South Island as a whole or is it more a specific South Island location?
- consider what is prime time (if using the second question above) and what channels should be considered (for example, only free to air or include SKY);
- decide who is collecting on which day, from which channel, in which time slot. (This part of the planning will depend on what the investigative question posed was. The data collection needs to be organised in a way that ensures there are no double-ups of data.)

At this stage, have the students consider if there is further information they could be collecting for future use, for example, information such as what the news item was about or how long it was. Collecting the additional information at the same time makes this a multivariate data set that can be used for other investigations.

Data collection

The students may find a table such as this one useful to complete. They can adapt the headings to fit the data that they decide to collect.

Channel			
Day			
Location: Topic: Time:			

Analysis

The students will draw a variety of graphs depending on the data they have selected. They should be writing "I notice …" descriptive statements about the shape of the data, the spread of the data, the middle of the data, and any appropriate statistics that may have been calculated.

Conclusion

In this part of the PPDAC cycle, the students should be answering the investigative question(s) they posed and using evidence from their analysis to support their answers.

Census statistics on the population of each region and other interesting information can be found on the Statistics New Zealand website. The population figures in the table below were sourced from www.stats.govt.nz/datasets/population/populationestimates.htm, under the Subnational Population Estimates heading in the Tables section. The figures in the last two columns were used to draw a bar graph (shown after the table below) that compares population and news items (the South Island item was put under Canterbury because that area has the highest population in the South Island; the percentage is of the New Zealand content).

Estimated Resident Population (1)
Regional Council areas
At 30 June 2006–2008

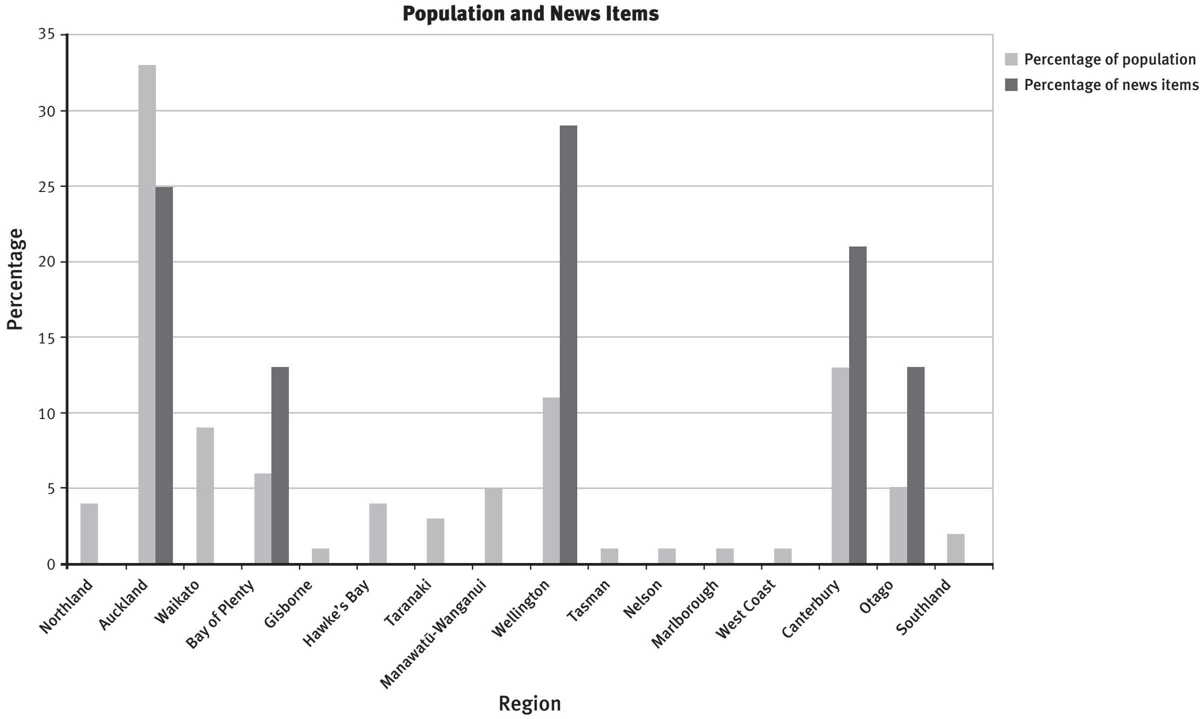
Regional Council area (2)	At 30 June			Using 2008 estimates and Activity One stories	
	2006	2007	2008	Percentage of total population	Percentage of news stories
Northland	152 700	153 800	154 700	4	
Auckland	1 371 000	1 394 000	1 414 700	33	25
Waikato	395 100	398 600	402 200	9	
Bay of Plenty	265 300	267 700	269 900	6	13
Gisborne	46 000	45 900	46 000	1	
Hawke's Bay	152 100	152 500	152 800	4	
Taranaki	107 300	107 200	107 500	3	
Manawatū-Wanganui	229 400	229 000	229 200	5	9
Wellington	466 300	470 300	473 800	11	29
Tasman	45 800	46 100	46 500	1	
Nelson	44 300	44 400	44 700	1	
Marlborough	43 600	44 000	44 500	1	
West Coast	32 100	32 200	32 400	1	
Canterbury	540 000	546 900	552 900	13	21
Otago	199 800	201 700	203 500	5	13
Southland	93 200	93 000	93 000	2	
North Island regions	3 185 100	3 219 200	3 250 800	76	
South Island regions	998 800	1 008 400	1 017 400	24	
Area outside regions(3)	650	650	650	0	
New Zealand	4 184 600	4 228 300	4 268 900	100	

(1) The estimated resident population is based on the census usually resident population count, updated for residents missed or counted more than once by the census (net census undercount); residents temporarily overseas on census night; and births, deaths, and net migration between census night and the date of the estimate.

(2) Based on 2006 regional council area boundaries.

(3) Includes the population of Kermadec Islands, Chatham Islands Territory, and people on oil rigs.

Note: Individual figures may not add up to the stated totals due to rounding.
All derived figures have been calculated using data of greater precision than published.



Question **2** provides a good opportunity for a discussion about when it is necessary to "clean" the data. For example, Masina labelled one news item as "South Island". If the students do likewise, they would need to decide what to do with it later in their investigation: leave it out, choose a likely South Island region (if they know what the item is about), put it into every South Island region, or put it into the region with the largest population (as suggested earlier). Unless the region is known, assigning it to the highest population area is best; putting it into every South Island region would make most regions over-represented.

As an extension to question **3**, you could ask the students if any region (Wellington in Masina's data) stands out as having many more news items than its population would suggest and why that might be.

Extension

Students may like to investigate related representation issues in the media such as male versus female representation in the sports section of the newspaper or ethnic representation in TV advertising.

Pages 16–17 Spin Doctor

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy		Probability
P	P	D	A	C

Key Competencies

Spin Doctor can be used to develop these key competencies:

- thinking: discerning if answers are reasonable and designing investigations
- using language, symbols, and texts: demonstrating statistical literacy, and communicating findings
- managing self: planning
- relating to others: working co-operatively in groups
- participating and contributing: sharing equipment and resources.

Statistical Ideas

Spin Doctor involves the following statistical ideas: using the **PPDAC cycle***, evaluating statements made by others, and using **investigative questions**, **multivariate data sets**, and **dot plots**.

The following activities provide a good introduction to concepts of statistical literacy that will be explored in greater depth at later levels (in particular level 7), when students will need to question the statistical sources used by the media when producing headlines.

Activity One

In this activity, students are asked to reorganise the data into a dot plot in order to "paint a picture" of the situation. Graphs can highlight the features of a data set more easily than looking at data in a table. Encourage your students to look at the whole shape of the graph (that is, the whole distribution) rather than just at the highest points.

Activity Two

The purpose of this activity is to highlight some of the ways that the media can skew data for their own purposes. Students may be unaware that newspaper or magazine headlines can be inflammatory, taking only the parts of the data that back up their argument. Usually this data is not untrue – but sometimes it doesn't tell the whole story. Encourage your students to always ask questions of the statistics presented to them. For example, in media reports, they should want to know how the data was obtained and what else the data may show. Sometimes, the same data can be used to support contradictory statements. Encourage your students to find examples of these in the students' headlines in this activity (for example, "Survey shows girls glued to screen" and "Boys more likely to watch tube").

It may be useful to have students answer the following questions themselves before beginning the task:

- Do you have a TV in your bedroom?
- How much time did you spend watching TV yesterday?
- About how much do you spend on your cellphone each month?
- How do you pay for your cellphone usage (parent, job, pocket money, other)?

This may give them more understanding of the situation and help answer questions such as that in question **4** regarding unreliable data. As a follow-up activity, the students could compare their own class's responses with the examples in the book and write headlines (funny or otherwise) to justify or refute them. Drawing another dot plot underneath the one they drew for question **1** in **Activity One** is a helpful way to compare two data sets. After they complete question **3**, have the students share their graphs with a classmate and discuss whether the graph is effective in communicating the desired information.

Investigation

Before your students begin their investigations, you may need to revise the PPDAC cycle with them. The main purpose of Spin Doctor is to encourage students to explore the CensusAtSchool website. You may need to allow the students time to try out some of the different aspects of this site, such as the random sampler, building summary tables, and the data viewer.

* For bolded terms, see What Is Statistics About?, pages 21–27.

Note that although the 2008 CensusAtSchool questionnaire contained media-use questions, this may not be true for subsequent years.

After the students take a random sample of responses, remind them to keep the identifying data (such as age, gender, year, and region) and the questions relating to media and to delete the data for all the other questions.

The survey section of CensusAtSchool lists the questions asked. Encourage the students to explore this section, especially the media questions, because it will help them to interpret the data they download from CensusAtSchool. (As noted in the students' book, most of the headings in the table on page **16** differ from the short forms used on the CensusAtSchool website. They make sense on the website, where additional information is provided via the questionnaires, but fuller versions are needed for clarity in the students' text.)

Problem

The students need to pose investigative questions. Using the starter "I wonder …" is useful. Questions can be of the following types: **summary**, **comparison**, or **relationship**. Examples of these in relation to the CensusAtSchool database for 2007 are:

Summary:

- I wonder what TV shows typically appealed to year 7 students in 2007?
- I wonder what websites year 9 students typically liked to access in 2007?
- I wonder what amounts girls in 2007 typically spent on cellphones each month?

Comparison:

- I wonder if year 8 girls tended to send more text messages than year 8 boys?
- I wonder if more younger students tended to get the money from their parents for their cellphone bills than older students?
- I wonder if students in high school in 2007 sent more text messages than intermediate school students?
- I wonder if students in years 9 and 10 in 2007 sent more text messages than those in years 11 and 12?

Relationship:

- I wonder if there is a relationship between the age of the student and the amount they spend on their cellphone?
- I wonder if students who have an MP3 player are more likely to have their own cellphone as well? (The students could probably use a two-way table [as in the notes for pages **22–23**] to answer this.)
- I wonder if there is a relationship between the amount of money students spent on cellphones and the number of text messages they sent?
- I wonder if there is a relationship between the number of text messages that the students sent and the number of text messages they received?

Plan

Have the students look at the CensusAtSchool survey questions to help them pose investigative questions. Remind them that investigative questions are questions we ask of the data; survey questions are questions we ask to get the data. Note that region and year level happen by default, so the students should keep these two variables.

Data

The data can be downloaded from the CensusAtSchool database www.censusatschool.org.nz

Analysis

The students will draw a variety of graphs depending on the data they have selected. They should be writing "I notice …" descriptive statements about the shape of the data, the centre of the data, the spread of the data, and any appropriate statistics that may have been calculated.

Conclusion

The students should be answering the investigative question(s) they posed and using evidence from the analysis to support their answers.

Spin Doctor uses a small sample of 25 for practical reasons. Encourage your students to think about the advantages and disadvantages of larger sample sizes.

One of the main advantages of using the data obtained from CensusAtSchool is that it is a large database and students are therefore not limited by the time constraints involved in collecting a large amount of data themselves.

Pages 18–19 Free CD!

Mathematics and Statistics Achievement Objectives

Probability:

- Investigate situations that involve elements of chance by comparing experimental distributions with expectations from models of the possible outcomes, acknowledging variation and independence (Statistics, level 4).
  - Use simple fractions and percentages to describe probabilities (Statistics, level 4).

Statistical investigation		Statistical literacy	Probability	
P	P	D	A	C

Key Competencies

Free CD! can be used to develop these key competencies:

- thinking: solving problems in new situations, making decisions, engaging in making sense, making conjectures, and dealing with uncertainty and variation
- using language, symbols, and texts: interpreting statistical information
- relating to others: co-operating and communicating thinking
- participating and contributing: taking on appropriate roles in different situations.

Statistical Ideas

Free CD! involves **probability***, **simulations**, probability misconceptions, and using statistical **graphs**.

This activity challenges some of the probability-related misconceptions students often have – particularly when it comes to winning competitions that are based on chance. Ask your students why a company might offer a free prize. The reason is transparent: to get consumers to choose their product or to buy more of their product. Receiving prizes might make consumers feel good about the company, but the company's bottom line is money, not friends.

Activity One

Simulations such as the one in this activity can be a good way to challenge students' assumptions about probability. However, the simulations do need to provide a good representation of the actual situation and be understood by the students. In this case, the key is to ensure that the students understand that the counters represent cereal boxes and that 1 counter in 3 is a different colour to represent the boxes with CDs in them.

Note that this simulation is not perfect because each bean is taken from the bag and not replaced until all 3 beans are selected, so the probabilities for the next selection change slightly: the probability that the first bean will be colour A is $\frac{10}{30} \approx 0.33$; for the second (assuming the first was colour A), it is $\frac{9}{29} \approx 0.31$; and for the third (assuming the first two are colour A), $\frac{8}{28} \approx 0.29$.

* For bolded terms, see What Is Statistics About?, pages 21–27.

A simple, more accurate simulation involves the use of a standard dice. Let the numbers 1 and 2 represent a box with a CD and the numbers 3 to 6 a box without a CD. Rolling the dice represents purchasing a box of cereal. Three rolls of the dice (or 1 roll of 3 dice) models the purchase of 3 boxes of cereal, as in question **2**.

Another simulation involves the use of the random number function (Ran#) that can be found on most scientific calculators. Have the students initiate the random function and then keep pressing the = key; they will find that they obtain a series of random numbers between 0 and 1, usually to 3 decimal places. For the purpose of this simulation, they can ignore all but the digit after the decimal point. If this digit is {1, 2, 3}, it represents a box with a CD; if it is {4, 5, 6, 7, 8, 9}, it represents a box without a CD. If the digit is 0, they should ignore it. Here are 30 numbers generated in this way. Each column represents a trial (in this case, the purchase of 3 boxes of cereal). Each shaded cell represents a box containing a CD:

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
0.75	0.481	0.498	0.41	0.411	0.274	0.939	0.871	0.817	0.446
0.309	0.239	0.604	0.594	0.745	0.122	0.199	0.155	0.294	0.663
0.99	0.435	0.966	0.276	0.154	0.187	0.199	0.151	0.887	0.316

It can be seen that 1 trial (column 6) produced 3 CDs, 2 trials (columns 7 and 8) produced 2 CDs, 6 trials produced a single CD, and only 1 trial (column 3) produced 0 CDs. Overall, $\frac{13}{30}$ of the boxes contained CDs and $\frac{9}{10}$ of the trials produced at least 1 CD.

By pooling their results, the students will be able to approximate the theoretical probabilities for this experiment. (For your information, approximate percentages for these are: No CDs ≈ 30%; 1 CD ≈ 44%; 2 CDs ≈ 22%; 3 CDs ≈ 4%.)

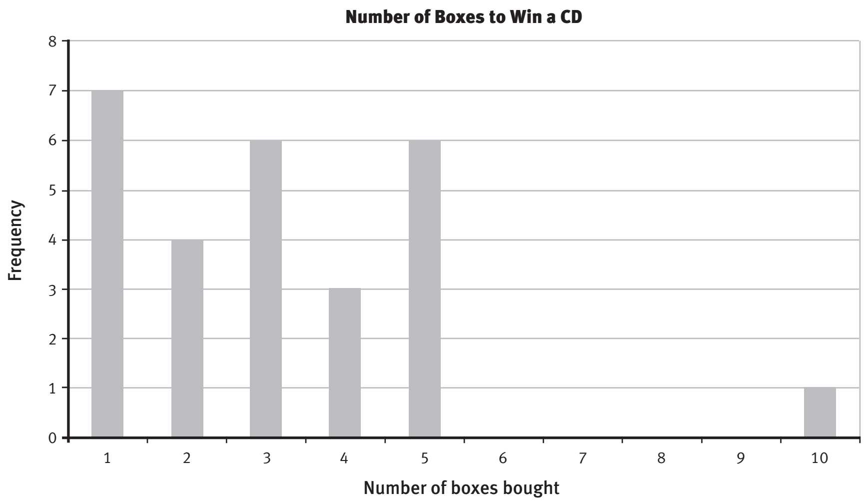
ACTIVITY TWO

By the end of question **1**, the students should have data from 100 trials to display and discuss. As in the previous activity, a simulation involving dice or random numbers can be used instead of beans. This time, if the students are generating the random numbers on the calculator, have them write down all the non-zero digits. Each 1–3 digit represents a box with a CD in it (a "win"). Mark off the sequences of digits that end in "wins". Here is the sequence obtained from one simulation (each trial stops as soon as a CD is found):

75481 | 49841 | 41 | 1 | 2 | 7493 | 9871 | 81 | 74463 | 92 | 3 | 9645947451 | 2 | 2 | 1 | 991 | 552 | 94663 | 99943 | 59662 | 761 | 541 | 871 | 991 | 51 | 8873 | 1

In this particular simulation, 5 boxes had to be purchased to win the first CD, another 5 to win the next CD, 2 to win the third CD, and so on. In all, this simulation contained 27 wins. The data can be summarised in a table and graph like this:

Boxes bought	1	2	3	4	5	6	7	8	9	10
Frequency	7	4	6	3	6	0	0	0	0	1



When a group of 5 pairs graphs their accumulated 100 trials, they are likely to find that an orderly graph appears out of randomness. This is well worth sharing and discussing. (The graph of 100 trials will have a more orderly distribution than the trial graph shown above or than any of the graphs that pairs of students make from their own data.)

In answer to the question, the simulation shown suggests that a person should not normally have to buy more than 5 boxes of cereal to win a CD. But clearly, 5 boxes will not *guarantee* a CD because, in one case, it took 10 purchases. Would 10 boxes guarantee a CD? No. When the outcome is a matter of chance, you can never guarantee anything!

Extension

You could ask your students to guess what would happen if they did a million trials. The students could also consider: Did the cereal company cost this promotion out using, for example, experiments, simulations, or theory?

Pages 20–21 Questionnaire Queries

Mathematics and Statistics Achievement Objectives

Statistical investigation

- Plan and conduct investigations using the statistical enquiry cycle:
  - determining appropriate variables and data collection methods;
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation		Statistical literacy	Probability	
P	P	D	A	C

Key Competencies

Questionnaire Queries can be used to develop these key competencies:

- thinking: designing investigations, predicting/visualising outcomes, and thinking creatively
- using language, symbols, and texts: collecting and interpreting statistical information
- relating to others: listening actively, negotiating meaning, understanding others' thinking, and working co-operatively
- managing self: making decisions and reflecting.

Statistical Ideas

Questionnaire Queries involves the following statistical ideas: using the **PPDAC cycle***, **investigative questions**, and **survey questions**, planning an investigation, designing a **questionnaire**, and doing a pilot study.

Activity

Students at this level often find it difficult to write survey questions and can finish with more responses in the "other" category than in any other. This activity provides an opportunity to discuss ways to avoid this pitfall.

Before the students begin, encourage them to decide on an interpretation of the problem. They need to collectively decide on their answers to questions such as: "What does a school radio station entail?" "What will be the hours of operation?" "What are the rules of use?" Agreeing on these will make the purpose for this activity more meaningful.

* For bolded terms, see What Is Statistics About?, pages 21–27.

Question **1** demonstrates the need for survey questions to be specific and focused in terms of the possible responses required. Ask the students how they could categorise the answers given – they won't find it easy!

Question **2** asks students to clarify the problem in order to make a plan. It is an excellent opportunity to focus on the "problem" and "plan" parts of the PPDAC cycle.

Problem

The problem or investigative question is basically: "What are the radio-listening habits of students at Motutapu School?"

Plan

Questions to prompt students' thinking about the planning phase could be:

- What evidence is needed?
- What variables should be collected?
- How will they be measured?
- Why would we want to collect information about gender and age?
- Who should we ask?
- What listening habits should we ask about, weekdays or weekends?
- How will we record the data?

One way to approach question **3** is for the students to trial this questionnaire with a few classmates before deciding which survey questions to keep and which to dismiss or change. They will need to clarify the "measure". For example, for "How often do you listen to the radio?", they need to ask themselves: "What does 'often' mean?" "What does 'sometimes' mean?" One way to find out is to start by asking some classmates the original open question ("How often do you listen to the radio?") and use their responses to refine the categories and what they mean by them.

For question **4**, the students need to understand that questionnaires can be designed to collect either qualitative or quantitative data. These terms may vary in meaning depending on the context. In general terms, qualitative (subjective) data is gleaned from interview questions. The less directed the question, the more qualitative it is. Quantitative (or measurable) data is gleaned from surveys that have been given to a large number of people. Both types of data have advantages and disadvantages (you may like to discuss these with students). When designing a questionnaire, it may be helpful to interview a smaller number of people in order to generate some likely responses (qualitative data), then use these to create the response categories for a survey that can be given to a large number of people to get quantitative data.

Extension

Use a similar process for a topical issue in your school.

Pages 22–23

Gaming Choices

Mathematics and Statistics Achievement Objectives

Probability:

- Use simple fractions and percentages to describe probabilities (Statistics, level 4).

Statistical investigation		Statistical literacy	Probability	
P	P	D	A	C

Key Competencies

Gaming Choices can be used to develop these key competencies:

- thinking: making deductions, dealing with uncertainty, and making decisions
- using language, symbols, and texts: interpreting statistical information and interpreting word problems.

Statistical Ideas

Gaming Choices involves **probability*** and two-way tables.

activity

Many students will be very familiar with computer gaming, particularly strategy games. Strategy games typically involve completing missions to get to the next level. There is always a probability factor involved.

In this activity, the students explore probability in a gaming context. The first task is straightforward, but the other tasks involve two-way tables.

Two-way tables are usually of this form:

	A	A^1	Total
B			
B^1			
Total			

(Note that A^1 is the set of all things that are not in the set A.) Two-way tables summarise two events (for example, a coin toss and a dice roll). They can be used to answer probability questions such as “What is the expected number of times you would get both heads and a 6 from 600 rolls and tosses?” The two-way table can be started by filling in the expected number of 6s (100 out of 600) and the expected number of heads (300 out of 600).

	Heads	Not heads (tails)	Total
A 6			100
Not a 6			
Total	300		600

From here, it is a matter of keeping totals correct to fill the gaps and find an answer of 50 (see the shaded box):

	Heads	Not heads (tails)	Total
A 6	50	50	100
Not a 6	250	250	500
Total	300	300	600

You can then turn these into percentages. $\frac{50}{600}$ chances of getting both a 6 and heads is $8.\dot{3}\%$.

* For bolded terms, see What Is Statistics About?, pages 21–27.

It's a characteristic of computer gaming that there's a probability built into every "mission". The "10" provided in the two-way table in **Task 2** for stronger wolves who won't give up their fangs and the "35%" in **Task 3** of missions where the package gets broken are factors that are built into the game. The students use these factors to complete the tables in a logical way and then make deductions from them.

Task 1 in Gaming Choices is an opportunity to look at certainty as a probability concept. If something is certain, the probability is 1 or 100%. **Task 2** requires students to complete and then "read" the two-way table. To answer question **2b**, they need to locate the cell that means Rory is successful. That is, they need to connect "Rory will win" with "Wolves will give fangs" – this is the top left cell. You may wish to discuss with the class what the other cells represent. All numbers will come from cells in the two-way table. For the wolves weaker than Rory, 80 out of a total of 90 weaker wolves will give up their fangs; similarly, 20 out of a total of 30 stronger wolves will give up their fangs.

Task 3 is very similar to the above, but this time, percentages are given.

Questions **4**, **5**, and **6** are about using probability to help with decision making. Note that a high probability does not guarantee success; for example, a 90% chance of rain may still result in a sunny day!

Extension

Do a class brainstorm of the other games that use probability and discuss ways of calculating probabilities from known information.

Page 24 Advertising Claims

Mathematics and Statistics Achievement Objectives

Statistical literacy

- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Statistical investigation			Statistical literacy	Probability
P	P	D	A	C

Key Competencies

Advertising Claims can be used to develop these key competencies:

- thinking: challenging assumptions, engaging in making sense, and evaluating
- using language, symbols, and texts: demonstrating statistical literacy
- relating to others: accepting and valuing different viewpoints and negotiating meaning.

Statistical Ideas

Advertising Claims involves the following statistical ideas: evaluating statements made by others and thinking critically about the statistics presented in the media.

ACTIVITY

The purpose of this activity is to get students to think critically about the statistics presented to them in the media or, as in these examples, advertising that claims statistical backing.

For each part of question **1**, encourage your students to discuss what the statement is saying and, perhaps more importantly, what is it not saying. Discuss how the statistic or the data may have been obtained. Students at this level may not be fully aware of the ways in which data can be manipulated in order to present a desired picture for advertising (for more on this see Spin Doctor on pages **16–17**).

Some examples of questions and activities to get students thinking critically are:

- For **a**, ask the students what they think the dentists were actually asked in the survey. The question Which brand of toothbrush do you use?" and the question "Have you ever used a Bubbles toothbrush?" will result in a differing amount of Bubbles "users". You could ask your class these kinds of questions and see if the results are significantly different. You could also ask them about the sample of dentists. "How big was it?" "What population of dentists did it come from?" (New Zealand? America?) "Were the dentists selected randomly?" "Is it possible to select dentists randomly?"
- For **b**, get the students to find some "low-fat" products and compare with the full-fat counterparts to see how much less fat each low-fat product really has. It is also useful to look at sugar content because some low-fat products replace fat with extra sugar.
- For **c**, this statement provides a good lead-in to a blind taste test. For example, get two known brands of orange juice and have 10 students do a blind tasting and state which they prefer. Then repeat the experiment, but have 10 different students comparing the losing brand to a watered-down version of the winning brand from the first experiment. Discuss how the preference for one brand depends on what it is being compared with.
- For **d**, ask the students to find out what number of books must be sold to define a best seller in New Zealand and how best-seller ratings are done in New Zealand. They can then debate how reliable or meaningful the ratings are. (They could do an Internet search on "New Zealand best seller sales figures".) Have them compare this information with that for another country, such as the United States, and discuss why the numbers are different.
- For **e**, discuss the ways in which a magazine's own ranking system may not be purely objective. Ask the students if they think magazines get paid by manufacturers to promote their product.

When the students are critiquing real advertisements in question **2**, have them ask and attempt to answer questions such as:

- Who was surveyed?
- How many people were surveyed?
- How were the participants chosen?
- What survey questions were asked?
- What was tested?
- What did the test involve?
- What is this statistic actually saying? What is it not saying?
- What assumptions have been made?

The students may find it useful to use the PPDAC cycle when they are critiquing because every stage of the cycle can be questioned in relation to the quoted statistic. In other words, the advertiser is just presenting the "conclusion" phase – the other steps must be assumed and can therefore be questioned.

Copymaster **Texting Olympics**

Name	Sprint	Marathon	Hurdles
Rebecca	0:05.44	0:49.67	0:42.40
Aki	0:12.91	3:42.78	1:47.06
Aoife	0:18.75	2:55.59	1:46.65
Mere	0:05.00	1:00.00	0:49.56
Nina	0:06.00	1:03.00	0:53.44
Ariana	0:05.00	1:15.00	0:46.74
Thomas	0:04.66	1:01.00	0:40.96
Laurel	0:08.65	1:19.85	0:55.13
Abhay	0:11.10	1:44.75	1:06.62
Hua-Ling	0:12.19	1:20.43	1:21.75
Fiona	0:12.16	1:45.94	1:07.35
Ming	0:04.06	0:46.07	0:43.13
Rangi	0:06.69	1:12.08	0:43.47
Siri	0:08.22	0:51.31	0:28.41
Quaid	0:05.97	1:11.53	1:53.12
Andrew	0:05.47	1:48.82	1:14.84
Matthew	0:07.09	1:28.75	1:07.83
Yvette	0:04.29	1:48.50	1:28.69
Susan	0:05.50	1:30.00	0:41.72
Tariq	0:08.12	0:43.31	0:39.56
Chris	0:07.28	1:54.94	0:54.90
Toline	0:04.00	0:55.88	0:30.50
Taylah	0:07.88	1:39.94	1:00.72
Molly	0:04.97	1:01.38	0:45.03
Quinten	0:10.43	1:20.25	1:13.66
Koria	0:16.94	3:57.22	2:53.34
Mary	0:06.69	0:37.88	0:25.13
Aketu	0:07.38	1:28.50	0:40.44
Pania	0:04.81	1:10.00	0:53.72
Mike	0:05.46	1:03.59	0:51.59

Name	Sprint	Marathon	Hurdles
Rebecca	0:05.44	0:49.67	0:42.40
Aki	0:12.91	3:42.78	1:47.06
Aoife	0:18.75	2:55.59	1:46.65
Mere	0:05.00	1:00.00	0:49.56
Nina	0:06.00	1:03.00	0:53.44
Ariana	0:05.00	1:15.00	0:46.74
Thomas	0:04.66	1:01.00	0:40.96
Laurel	0:08.65	1:19.85	0:55.13
Abhay	0:11.10	1:44.75	1:06.62
Hua-Ling	0:12.19	1:20.43	1:21.75
Fiona	0:12.16	1:45.94	1:07.35
Ming	0:04.06	0:46.07	0:43.13
Rangi	0:06.69	1:12.08	0:43.47
Siri	0:08.22	0:51.31	0:28.41
Quaid	0:05.97	1:11.53	1:53.12
Andrew	0:05.47	1:48.82	1:14.84
Matthew	0:07.09	1:28.75	1:07.83
Yvette	0:04.29	1:48.50	1:28.69
Susan	0:05.50	1:30.00	0:41.72
Tariq	0:08.12	0:43.31	0:39.56
Chris	0:07.28	1:54.94	0:54.90
Toline	0:04.00	0:55.88	0:30.50
Taylah	0:07.88	1:39.94	1:00.72
Molly	0:04.97	1:01.38	0:45.03
Quinten	0:10.43	1:20.25	1:13.66
Koria	0:16.94	3:57.22	2:53.34
Mary	0:06.69	0:37.88	0:25.13
Aketu	0:07.38	1:28.50	0:40.44
Pania	0:04.81	1:10.00	0:53.72
Mike	0:05.46	1:03.59	0:51.59

Copymaster

Singing Star

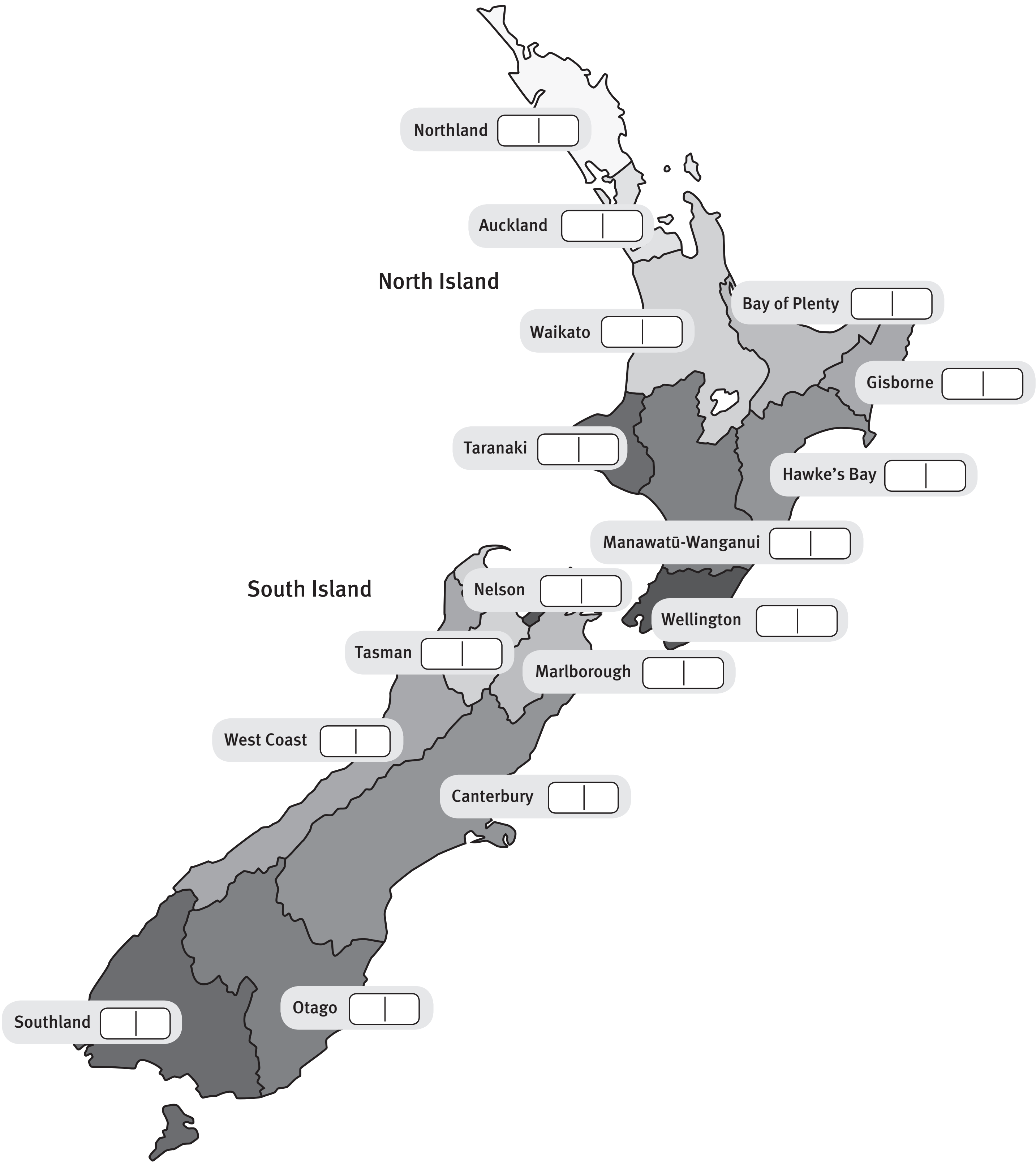
Order	Contestant	Judge 1	Judge 2	Judge 3	Total	Median	Mean (1 d.p.)
1	Alexia	4	7	8			
2	Hone	4	8	4			
3	Ding	2	6	5			
4	Lesieli	5	8	6			
5	Tāne	3	7	4			
6	Simon	3	9	5			
7	Ariana	2	4	6			
8	Ethan	4	4	5			
9	Whetū	4	8	5			
10	Aoife	3	6	7			

Order	Contestant	Judge 1	Judge 2	Judge 3	Total	Median	Mean (1 d.p.)
1	Alexia	4	7	8			
2	Hone	4	8	4			
3	Ding	2	6	5			
4	Lesieli	5	8	6			
5	Tāne	3	7	4			
6	Simon	3	9	5			
7	Ariana	2	4	6			
8	Ethan	4	4	5			
9	Whetū	4	8	5			
10	Aoife	3	6	7			

Order	Contestant	Judge 1	Judge 2	Judge 3	Total	Median	Mean (1 d.p.)
1	Alexia	4	7	8			
2	Hone	4	8	4			
3	Ding	2	6	5			
4	Lesieli	5	8	6			
5	Tāne	3	7	4			
6	Simon	3	9	5			
7	Ariana	2	4	6			
8	Ethan	4	4	5			
9	Whetū	4	8	5			
10	Aoife	3	6	7			

Regional Council Areas of New Zealand

Regional Council Areas of New Zealand



Copymaster **Whose News?**

2008 Population Estimates	
Region	**Population**
Northland	154 700
Auckland	1 414 700
Waikato	402 200
Bay of Plenty	269 900
Gisborne	46 000
Hawke's Bay	152 800
Taranaki	107 500
Manawatū-Wanganui	229 200
Wellington	473 800
Tasman	46 500
Nelson	44 700
Marlborough	44 500
West Coast	32 400
Canterbury	552 900
Otago	203 500
Southland	93 000
North Island regions	3 250 800
South Island regions	1 017 400
Area outside regions	650
New Zealand	4 268 900

2008 Population Estimates	
Region	**Population**
Northland	154 700
Auckland	1 414 700
Waikato	402 200
Bay of Plenty	269 900
Gisborne	46 000
Hawke's Bay	152 800
Taranaki	107 500
Manawatū-Wanganui	229 200
Wellington	473 800
Tasman	46 500
Nelson	44 700
Marlborough	44 500
West Coast	32 400
Canterbury	552 900
Otago	203 500
Southland	93 000
North Island regions	3 250 800
South Island regions	1 017 400
Area outside regions	650
New Zealand	4 268 900

2008 Population Estimates	
Region	**Population**
Northland	154 700
Auckland	1 414 700
Waikato	402 200
Bay of Plenty	269 900
Gisborne	46 000
Hawke's Bay	152 800
Taranaki	107 500
Manawatū-Wanganui	229 200
Wellington	473 800
Tasman	46 500
Nelson	44 700
Marlborough	44 500
West Coast	32 400
Canterbury	552 900
Otago	203 500
Southland	93 000
North Island regions	3 250 800
South Island regions	1 017 400
Area outside regions	650
New Zealand	4 268 900

Source: Statistics New Zealand (see details on page 41)

Copymaster **Free CD!**

ACTIVITY ONE

Number of CDs	Tally	Frequency	Fraction	Percentage
No CD				
1 free CD				
2 free CDs				
3 free CDs				
Total				

ACTIVITY TWO

Pairs' Trial Record											
Trial		1	2	3	4	5	6	7	8	9	10
Number of boxes to win	Classmate 1										
	Classmate 2										
	Total										

Number of boxes to win	Totals					Times out of 100	Fraction out of 100	Percentage
	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5			
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								