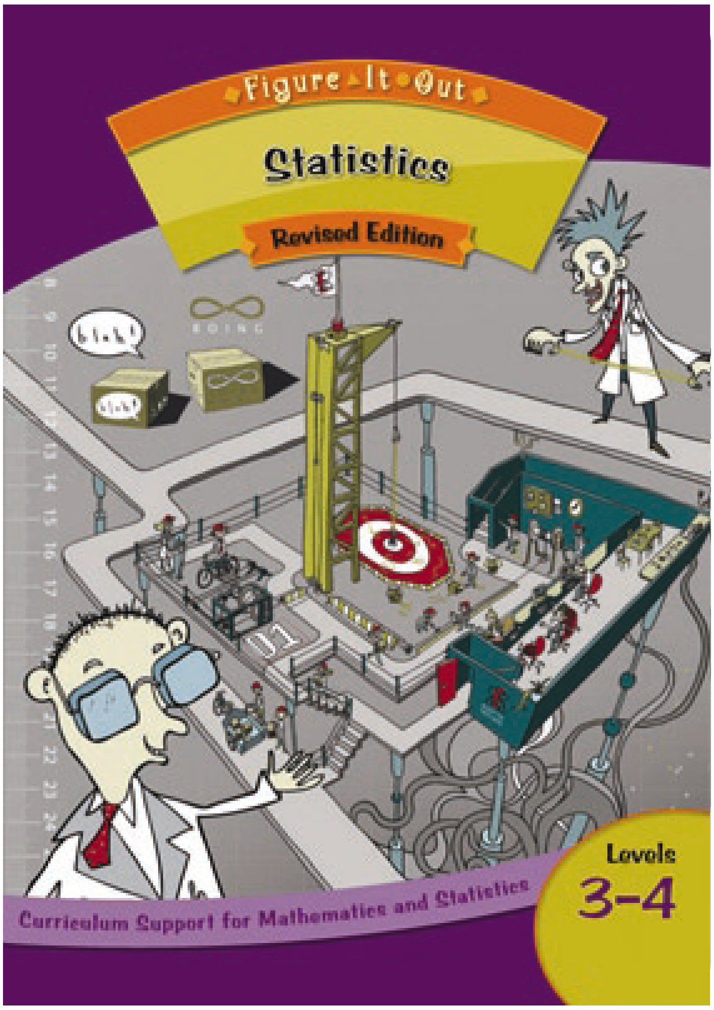
Answers and Teachers' Notes



MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

Contents

Introduction

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 Teachers' Notes* where appropriate.

The level 3–4 Figure It Out *Statistics* book that these *Answers and Teachers' Notes* relate to is an extensive revision, in line with the achievement objectives of the mathematics and statistics learning area of *The New Zealand Curriculum*, of the level 3–4 *Statistics* book published in 2001.

Student books

The activities in the Figure It Out student books are written for New Zealand students and are set in meaningful contexts, including real-life and imaginary scenarios. The level 3–4 contexts reflect the ethnic and cultural diversity and life experiences that are meaningful to students in year 6. However, teachers should use their judgment as to whether to use the level 3–4 book with older or younger students who are also working at this level.

The 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 the revised *Statistics* book 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 relevant achievement objectives, comments on mathematical ideas, processes, and principles, and suggestions on teaching approaches. The *Answers and Teachers' Notes* are also available on Te Kete Ipurangi (TKI) at www.tki.org.nz/r/maths/curriculum/figure/

Using Figure It Out in the classroom

Where applicable, each page of the students' book starts with a list of equipment that the students will need in order to do the activities. Encourage them 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. 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 the solution. 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.

Figure It Out

Statistics

Level 3–4

Answers

Page 1: Left to Chance

Activity

1. **a.** Practical activity

   b. Results will vary.

   c. Results will vary, but it is likely that there will be considerably more losses than wins (see **3** for explanation).

2. In this context, “fair” means that each person has an equal chance of winning. (Winners do get double their money back, but this is a different issue from “win or lose” chances.)

3. Win: LLL, RRR (L = left, R = right).

   Loss: RLL, LRL, LLR, LRR, RLR, RRL.

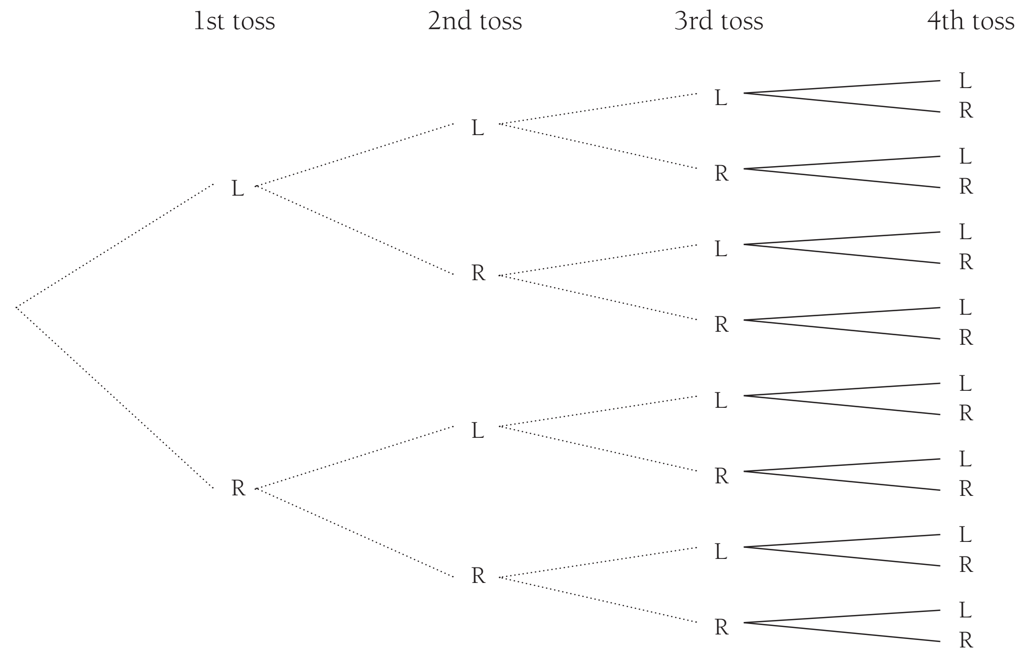
   There are 8 equally likely pathways, but only 2 of them lead to a “win”, so the probability of a win is $\frac{2}{8} = \frac{1}{4}$. So Simon’s game is not fair; he has a 6 in 8 chance of keeping the player’s money.

4. **a.** Practical activity. Results will vary, but you will probably still have more losses than wins. You might decide to pool your results with other classmates as you did earlier and see what difference this makes to the win:loss ratio.

   b. Win: LLLL, RRRR, RRLL, LRRL, LLRR, LRLR, RLRL, RLLR.

   Loss: RLLL, RLRR, RRLR, RRRL, LRLL, LLRL, LLLR, LRRR.

   On a tree diagram, the outcomes are set out (dotted lines for the Green Game) like this:



There are 16 equally-likely pathways. 8 of these give a “win”. So the probability of a win is $\frac{8}{16} = \frac{1}{2}$. The game couldn’t be any fairer, but from Simon’s point of view, it wouldn’t be very profitable as a fund-raiser (and in fact, he could make a loss, given that winners get double their money back!). He would raise more money from the 4-row game.

Pages 2–3: School Crossing

Activity

1. Discussion will vary, for example:

   i. True for the mornings but not the afternoons (in fact, the two quietest afternoons are the Fridays). True if you combine morning and afternoon data for each day.

   ii. True for every day except the second Thursday, so this is a reasonable generalisation.

   iii. True for every day except Friday

   iv. True. The graph shows this was the case on the second Thursday.

   v. True. The morning counts of approximately 90 (week 1) and 99 (week 2) are more than double the afternoon counts of approximately 41 and 39 respectively.

   vi. False. The number never reached 60. It would be truer to say: "About 50 vehicles went by on most afternoons."

   vii. You could argue true or false, depending on your definition of "not very different". On 3 of the 10 days, the counts differed by less than 10. On another 4 days, the difference was less than 20.

   viii. True. You can't get an exact number from the graph, but the approximate figures add up to about 1200.

2. Statements will vary. Possible statements include:

   "The morning and afternoon counts on the Wednesdays and Thursdays are closer than the counts on other days."

   "The Friday afternoon count was similar both weeks."

   "The morning count is lowest on Thursdays, but the afternoon count is lowest on Fridays."

3. Choices and explanations will vary. For example, for **ii**, it may be that most children get dropped off by their parents in the morning, but some catch a bus or a ride home with someone else in the afternoon. You also might want to think of possible reasons for the high morning and low afternoon Friday counts. Are more parents unavailable on Friday afternoons? (What happens on Fridays in your school?)

4. Suggestions will vary. Nathan and Malia have done a very good job of collecting data for two weeks, but patterns may be influenced by factors such as weather or major school activities such as sports days. The two classmates could continue collecting data over a longer time period (perhaps for another 2 or 3 weeks) and see if the patterns they have found continue.

Investigation

Data, patterns, and explanations will vary.

Pages 4–5: Change Ahead!

Activity One

1 **i.** From the tally chart, it can be seen that:

   - 67 responses favoured black; only 53 favoured grey
   - students were equally divided on black and grey (30 favoured grey, 30 favoured black) but parents strongly favoured black over grey, by 37 votes to 23.

   ii. In all, 24 responses favoured blue and black. (The next most popular combination was red and grey, with 22 responses.) But 28 students favoured red, a lot more than favoured blue (17).

   iii. The total number of votes for yellow and green (16 and 22 respectively) were much lower than for red and blue (42 and 40 respectively). But 12 parents favoured green and black, which is only 3 fewer than the most popular parent combination of blue and black.

2. Answers will vary. Paora might argue:

   - in terms of total votes (parents and students), red and grey (22) was second only to blue and black (24);
   - red was a lot more popular (28) than blue (17) with students;
   - it's the students who actually wear the uniform, so their preferences need to be

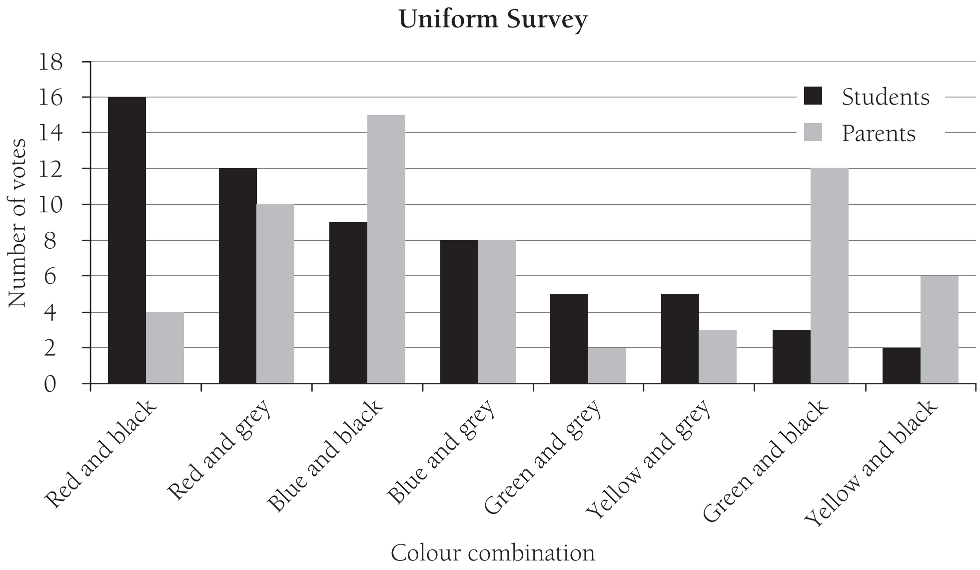
given some weight (although, in this case, red and black is the students' most popular choice and parents do seem to like black, even though they prefer it to be with blue).

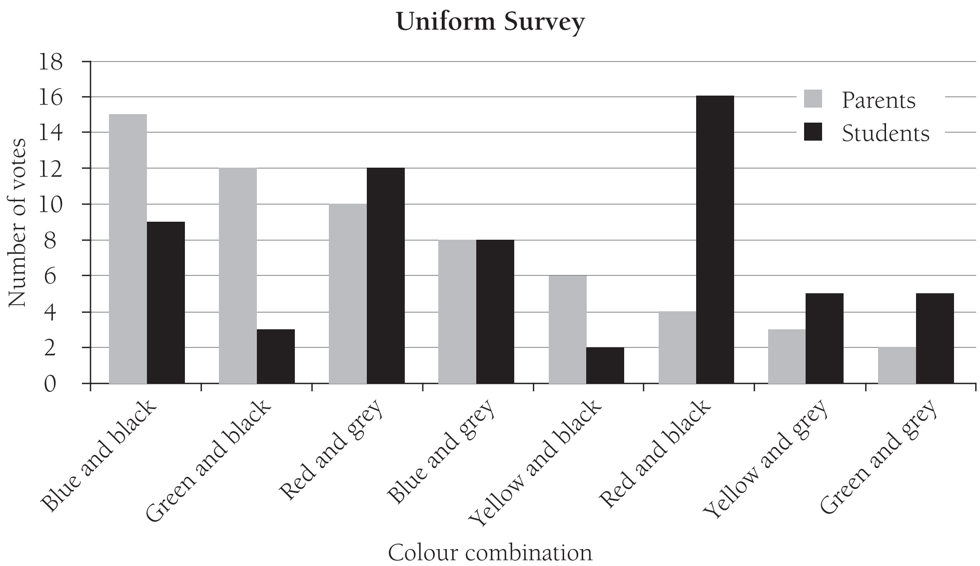
Activity Two

1. a. Ata's graph clearly shows how the different colour combinations stack up in terms of overall popularity (that is, when student and parent choices are added together).

   b. The graph disguises (to a certain extent) which combinations are more popular for either students or parents.

2. Graphs will vary. Here are two double bar graphs, the first arranged in order of student preference, the second in order of parent preference.





Activity Three

1. Forms will vary. The available data shows that the four colour combinations that got most votes involved either red or blue with black or grey, so it would make sense to include only these in a second survey. The focus of a second survey would be on trying to find a combination that both students and parents liked. It should be given either to equal numbers of students and parents or to all students and parents, in which case the votes should be shown as percentages.

2. Discussion and improvements will vary.

Pages 6–7: Opinion Polls

Activity One

1. a. Most students had an opinion for or against the statement in question. Approximately one-quarter of them (8) felt strongly about it (either for or against the statement).

   b. More girls (12) than boys (5) agreed with the statement.

   More people agreed with the statement (17) than disagreed (12).

   c. Comments will vary. It might be that this class has strong views on animal rights or that few class members have visited a modern zoo and seen the conservation work they do.

2. Graph **i** clearly shows that most people agree with the statement. It is also clear that most of the students in the two "agree" columns are girls.

   Graph **ii** clearly shows that more girls than boys agree with the statement.

   Graph **iii** clearly shows the percentage of boys and girls who responded in each of the five categories (for example, 70% of girls agreed or strongly agreed with the statement, compared with 37% of boys).

3. Discussion will vary. For this particular survey, even if there were 3 more boys to balance the genders and all 3 boys disagreed with the statement, there would still be more support than opposition. And even if all three boys supported the statement, there would still be more girls than boys who agreed with it. So the gender imbalance doesn't alter the overall opinion poll result.

Activity Two

1. Discussion will vary. Likert scales suit opinions on one particular issue. For example, you couldn't use a Likert scale to investigate choices such as favourite foods.

2. Statements will vary. For example:

   "There should be more sports time during the school week."

   "Every school should have a uniform."

   "People who use cellphones while driving should lose their licences."

   "Soft drinks should be banned from school premises."

Activity Three

Investigations, feedback, and improvements will vary. For example, in the case of the "time for sports" statement (above), it would need to be made clear whether sports time includes lunchtimes and/or after school, or just class time.

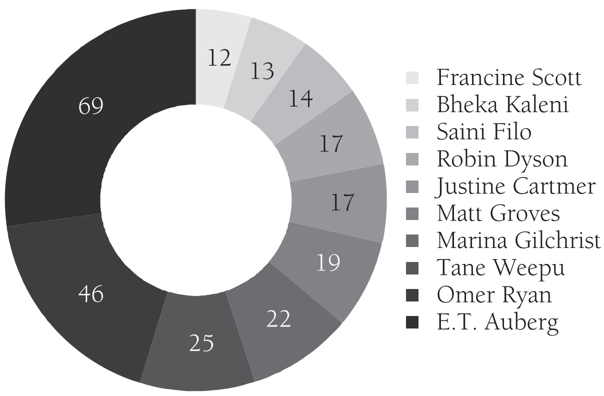
Pages 8–10: Reading Trends

Activity One

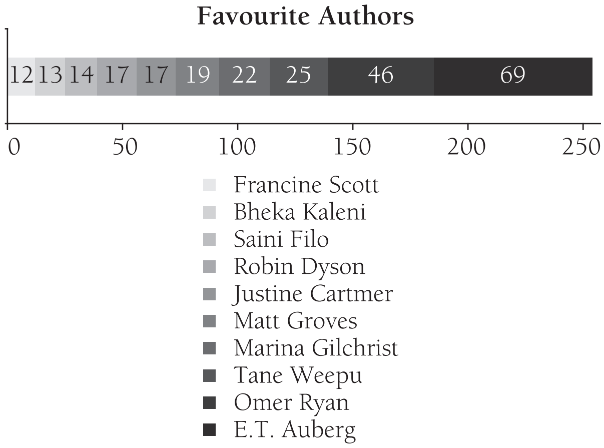
1. The bar graph shows the information more clearly because you can read appropriate numbers of votes from the left-hand axis. You can't do this from the pie graph, although the pie graph does clearly show the most popular authors.

2. Graphs will vary. Here are three possibilities:

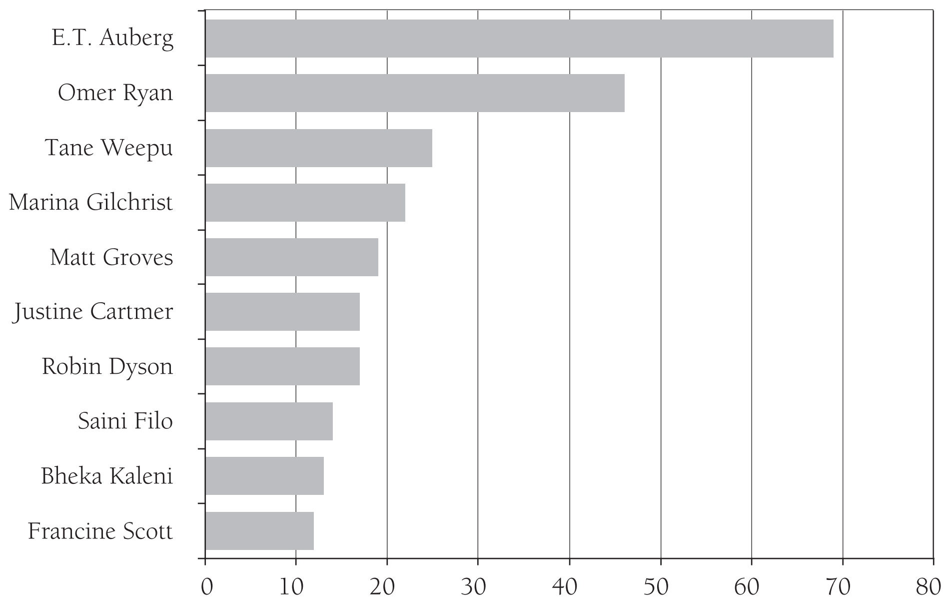
Favourite Authors



Favourite Authors



Favourite Authors



3. Opinions and explanations will vary. All will show that E. T. Auberg is the most popular, with Omer Ryan a clear second.

Activity Two

1. Possible statements include:

   Bheka Kaleni is popular only with boys in year 8.

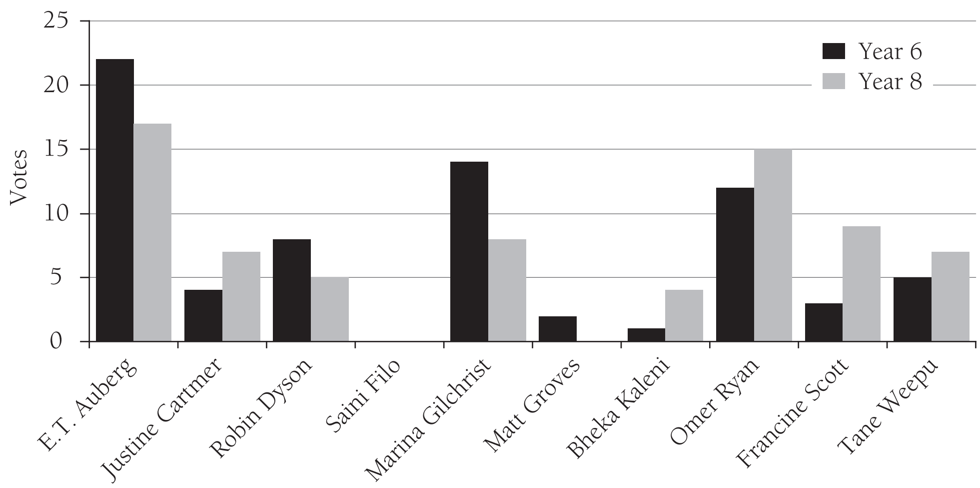
   Marina Gilchrist and Francine Scott are not popular with boys.

   E. T. Auberg is the most popular author for boys in both year 6 and year 8.

   Only 3 of the 10 authors (E. T. Auberg , Bheka Kaleni, and Omer Ryan) gain in popularity with boys between years 6 and 8.

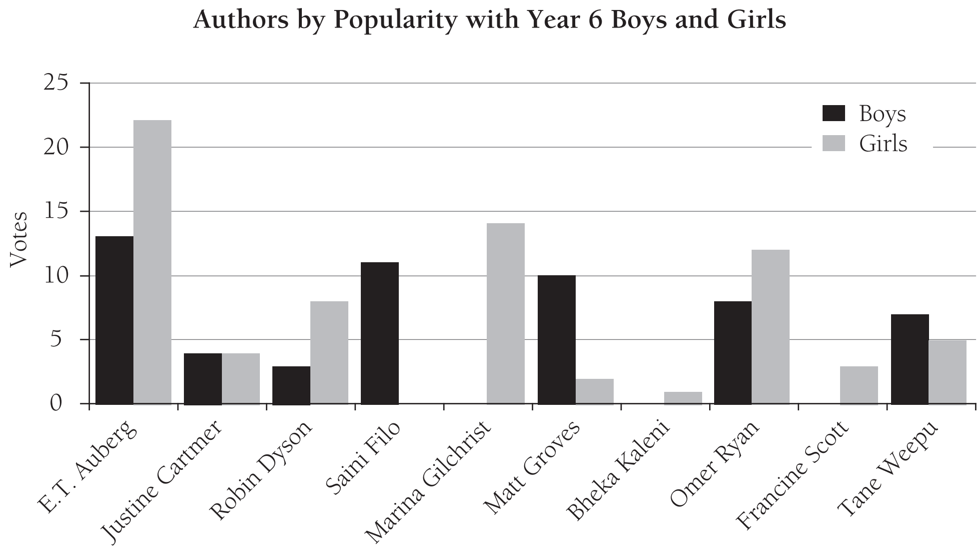
2. One possible graph (in this case, arranged alphabetically by surname) is:

Authors by Popularity (Girls)



Statements will vary. For example: E. T. Auberg is the most popular author with both year groups; Saini Filo is the least popular; Matt Groves and Bheka Kaleni are not very popular with girls.

3. One possible graph is:



Conclusions will vary. For example: For year 6 students, E. T. Auberg is much more popular with girls than boys; there are two authors that the boys are not interested in at all; there is no outstandingly most popular author for boys; the most popular author for girls is E. T. Auberg.

4. Investigations, graphs, and stories will vary.

5. Recommendations will vary. For example, she should buy as many books as possible by E. T. Auberg , who is popular with both boys and girls. Perhaps she could buy multiple copies of books by this author.

Activity Three

1. a. E. T. Auberg (more votes from all four groups than for any other author), Omer Ryan (second in terms of total votes and a minimum of 8 votes from all four groups), and Tane Weepu (third in terms of total votes; votes were evenly spread across all four groups).

   b. The speaker is talking about total votes from each year group. The five authors are Justine Cartmer (8→9 votes), Bheka Kaleni (1→12), Omer Ryan (20→26), Francine Scott (3→9), and Tane Weepu (12→13).

   c. Bheka Kaleni. (This author received only a single year 6 vote.)

   d. Marina Gilchrist and Francine Scott. (Both authors received no votes at all from boys.)

   e. Saini Filo (no votes at all from girls) and Matt Groves (this author received only 2 votes from girls)

2. You can't tell because students could vote for as many authors as they wished. All that you can know for sure is that at least 69 students voted, because this is the greatest number of votes received by any one author.

Page 11: Where's My Bus?

Activity

1. Descriptions will vary, but possibilities include:

   - Aroha's bus was almost always late – by somewhere between 1 and 10 minutes. Half the time it was nearly 5 or more minutes late. But Aroha couldn't count on it being late because twice it came a minute or two early. It never came exactly on time.
   - Audrey's bus was generally between 4 and 8 minutes late. In fact, more than half the time, it was more than 5 minutes late. But on six occasions, it arrived within 2 minutes of the scheduled time.
   - Luke's bus always arrived somewhere between 1 minute early and 4 minutes after the scheduled time. On nearly half of the mornings, it was within about 1 minute either side of the scheduled time. As long as Luke arrived 1 minute

before the bus was due, he could expect to be on his way within 5 minutes.

- Jansen's bus almost always arrived somewhere between the scheduled time and 6 minutes later. On nearly half the mornings, he was on his way no later than 2 minutes after the scheduled time. On 1 day, the bus was nearly 2 minutes early; on another, it was 11 minutes late. This lateness was exceptional. There must have been some unexpected cause.
- Tina-Lee's bus never arrived more than 5 minutes after the scheduled time. On most days, it was at the stop no later than 2 minutes after the due time. It was a little early on 4 mornings, by up to about 1 minute.
- Culley's bus was always late. Half of the time, it was between $1\frac{1}{2}$ and $5\frac{1}{2}$ minutes late. The rest of the time, it was between 9 and 13 minutes late. On any given morning, Culley could never be sure whether the bus would be moderately late or very late.

2. Lists may vary. A possibility is:

Luke's bus (most reliable)

Tina-Lee's bus

Jansen's bus

Audrey's bus

Aroha's bus

Culley's bus (least reliable)

Explanations for the order may vary, but Luke's bus was most consistently close to the scheduled time, whereas Culley's was always more than 1 minute late and often 10 or more minutes late.

3. Discussion will vary.

Pages 12–13: Fish Figures

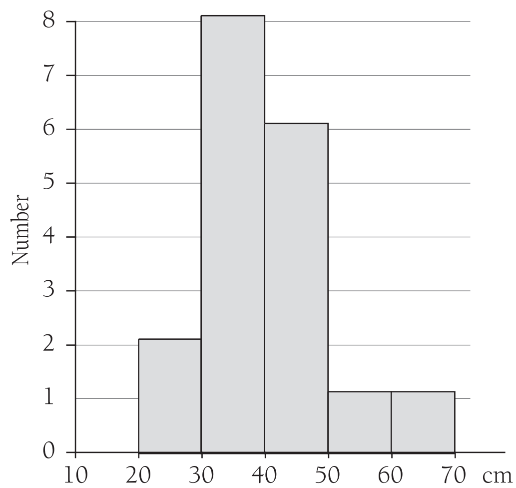
Activity One

1. The stem-and-leaf graph below shows the lengths of the fish caught.

Today's Catch (length)

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

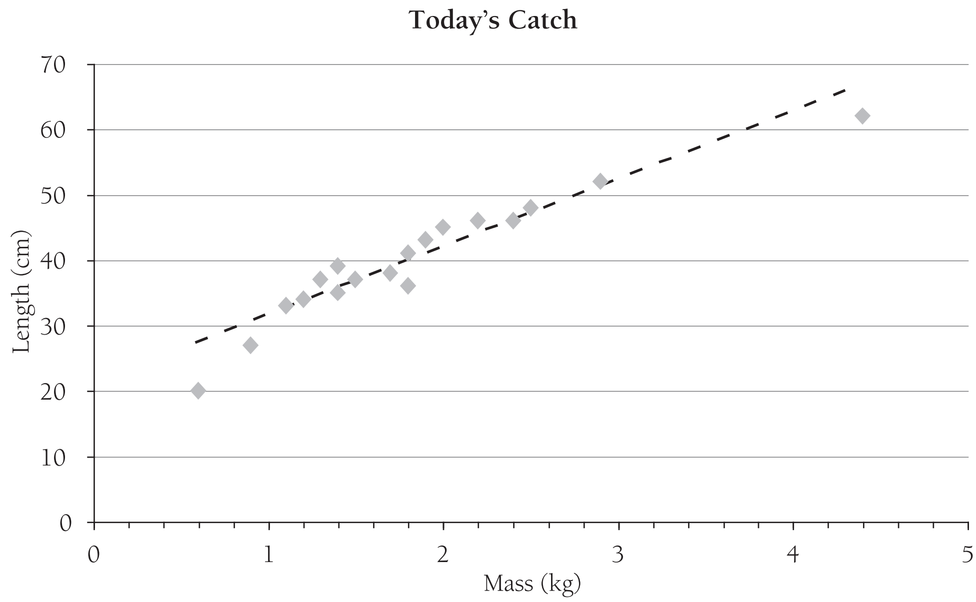
Today's Catch (length)



2. Descriptions will vary. For example, most of the fish weighed between 1 and 3 kg. Two fish weighed less than 1 kg and one weighed 4.4 kg. Most of the fish were between 30 and 50 cm long. The shortest was 20 cm, and the longest was 62 cm.

Activity Two

1. Your graph should look similar to this (except for the trendline, which has been added here in answer to question **2a**):



The pattern shows a definite relationship between length and mass: the longer the length of the fish, the heavier the fish is (within moderate variation); looked at the other way, the heavier the fish, the longer it is.

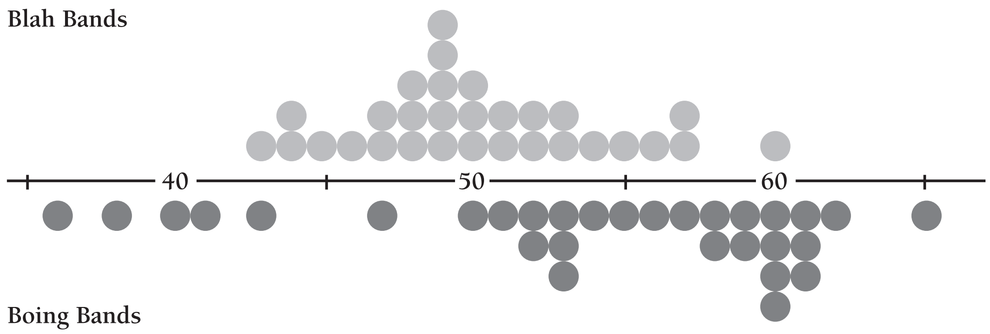
2. **a–b.** See the graph in question **1** for the trendline. The trendline should confirm your answer for question **1**.

c. On the basis of the trendline in the graph for question **1**, you could expect a fish weighing 3.5 kg to be between 50 and 60 cm in length; a fish that is 48 cm long could be expected to weigh between 2.5 and 3 kg.

Pages 14–15: Fully Stretched

Activity One

1.



2. Conclusions may vary, but the back-to-back dot plot does show that fewer Boing Bands broke before 50 cm and 5 of them stretched further than any Blah Bands. Half the Blah Bands broke before they stretched to 50 cm, while half the Boing Bands lasted until 55 cm. However, four Boing Bands broke sooner than any of the Blah Bands, so Boing Bands don't *reliably* last longer than Blah Bands.

Activity Two

Experiments, results, and reports will vary.

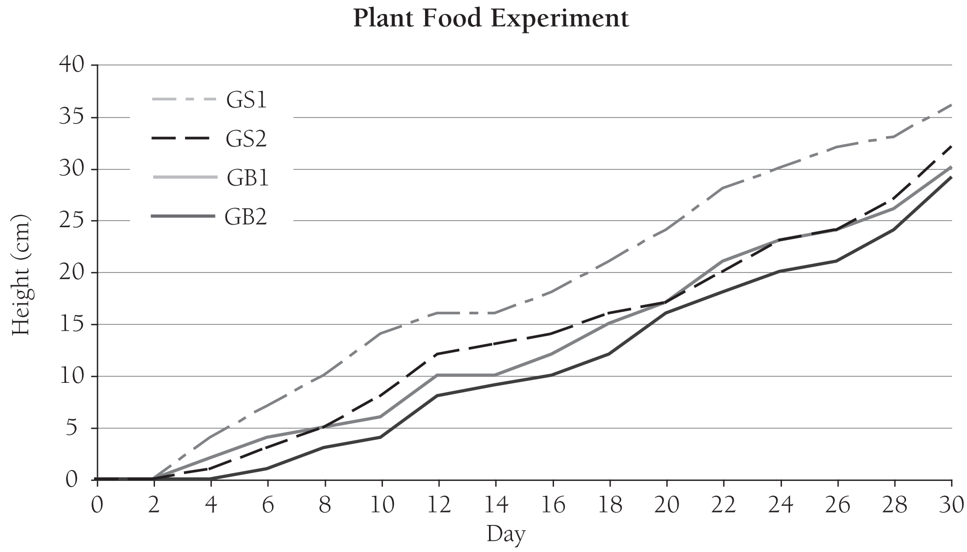
Activity Three

Suggestions and comments will vary, depending on the products.

Pages 16–17: Bean There

Activity

1\.



This is the best type of graph to use because it clearly shows the progress of the beans in relation to each other and, for any day, you can compare their heights by referring to the vertical axis.

2\. Discussion will vary. Possible points include:

i. The GrowSmart beans do grow *better* than the other two, but not more *steadily*. In fact, all four beans grow at a similar rate and have a similar pattern of plateaux and spurts. This is particularly true from day 10 onwards: GS1 gets away to an excellent start and manages to hold but not increase its lead.

ii. True. GS1 appears above ground 2 days earlier and quickly gains a lead. It maintains a lead of 8–10 cm for almost the entire period of the experiment, although the gap appears to be closing over the last 4 days.

iii. While these two beans *do* follow a very similar growth pattern (they keep swapping second and third position), it is not fair to ignore the other two when making a generalisation about the plant foods. (If a generalisation were based on just GS1 and GB2, it would strongly favour GrowSmart.)

iv. It is true that GB2 appears above ground 2 days later than the others, but this is fair in terms of the experiment, which measures growth from the time of planting.

v. Yes, you could. You could do the same for the other three beans, too, as all four are following a reasonably predictable growth path. Reasonable estimates (based on growth over the previous 4 days) would be GS1 38 cm, GS2 36 cm, GB1 33 cm, and GB2 33 cm.

vi. True. A longer experiment would have revealed, for example, if GS2's growth rate continued to exceed GB1's. It would also have shown if GS1's growth was slowing down.

vii. This is a fair comment, especially as factors other than the plant food will affect growth (for example, soil and the fertility of the individual beans).

viii. False. The beans within each pair in this experiment do not grow exactly the same even though they apparently receive the same treatment. It is difficult to ensure truly identical conditions and treatment. And like people, beans have their own individual characteristics.

ix. True. It may be that the beans would have grown just as well if they had been given just water.

x. Possibly true. It looks as if conditions are particularly favourable between days 10 and 12, as all four beans grow rapidly. But between days 12 and 14, they hardly grow at all. This suggests a couple of cold days.

Investigation

Experiments, conclusions, and discussion will vary. It might be useful to combine your results with those of other classmates who are using the same plant food as you. This will not guarantee your conclusions but may show some patterns, depending on how long you run your experiment for.

Pages 18–19: Take Five

Activity One

1. Answers will vary. Possible thoughts include:

   After the second selection: "Yes, it looks like they're all blue and red. Probably equal numbers of both."

   After the third selection: "Well, there's obviously a yellow tile, too. It looks like 1 yellow and roughly equal numbers of blue and red."

   After the fourth selection: "Wow, so there's a green, too. I'm surprised that didn't come out earlier. And 2 yellows! I'll have to revise my ideas."

2. This is the effect of probability (or chance). You can never be sure of an outcome when probability is involved.

3. **a.** There are several possibilities, but answers must include 3 blue, 3 red, 2 yellow, 1 green, plus one more. A good answer would be 4 blue, 3 red, 2 yellow, 1 green.

   b. For the suggestion in **a**, an explanation could be: We know from the four trials that this bag contains at least 3 blue, 3 red, 2 yellow, and 1 green, and since we know that there are 10 tiles in the bag, there is 1 tile we don't know the colour of. As blue appears more often than any other colour, it may well be blue (but it could be any of the other colours, too, or even a completely different colour).

   c. Yes. There is one tile we can't account for, which could be blue, red, yellow, green, white – or any other colour, for that matter.

4. This is the first selection to contain 4 blue tiles, so we now know for certain that the tiles in the bag are 4 blue, 3 red, 2 yellow, and 1 green.

5. There is no fixed number because it all depends on what comes out of the bag each time.

Activity Two

1. Practical activity

2. Practical activity

   a.–c. Predictions and results will vary.

   d. Combining results is useful because the more data you have, the clearer any patterns become.

Pages 20–21: Picking Pocket Money

Activity

1. Answers will vary. You may think that $5 guaranteed is best because at least Shelley then knows what she is getting. Or you may think that taking a chance on getting $7 or $10 is better because, some weeks, Shelley will have more to spend.

2. Practical activity. Results will vary.

3. Advice will vary, depending on the results of the experiment (and the values of the person making the recommendation!).

4. Results will vary. It is likely that different groups will have quite different results, so their advice may be quite different, too.

5. Results will vary, but it is likely that a large number of rolls of the dice will give a total that is very similar to the total of $5 per week for 12 months. If this is the case, your advice to Shelley might depend on whether you think it best to have the certainty of a regular income or the possibility of doing better (with the risk of doing worse).

Page 22: Crossing the Line

Activity

1. Practical activity. Results will vary.
2. Answers will vary.
3. There will most likely be considerable variation in results.
4. **a.–b.** If enough results are pooled, you should find that about twice as many of the toothpicks fall across a line as between the lines. This means that the probability of scoring a "hit" is about $\frac{2}{3}$.

Page 23: The Unit Fraction Game

Game

A game to improve your ability to operate with fractions

Activity

1. "Not fair", in a situation like this, means that the chance of losing is greater than the chance of winning.
2. While it might suggest that the chance of losing is greater than the chance of winning, it doesn't actually prove anything. (The cause might just be a run of bad luck.)
3. Comments will vary, depending on your experience of the game.
4. **a.** Methods will vary. Here is one:

   Each of the 6 operations can be carried out on each of the 6 numbers. That's a total of 6 × 6 = 36 possible results. These can be set out as an array or table, like this:

Number \ Operation	$+\frac{1}{2}$	$+\frac{1}{4}$	$+1$	$\times 2$	$\times 3$	$\times 5$
$\frac{1}{2}$	1	$\frac{3}{4}$	$1\frac{1}{2}$	1	$1\frac{1}{2}$	$2\frac{1}{2}$
$\frac{1}{3}$	$\frac{5}{6}$	$\frac{7}{12}$	$1\frac{1}{3}$	$\frac{2}{3}$	1	$1\frac{2}{3}$
$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$1\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	$1\frac{1}{4}$
$\frac{1}{5}$	$\frac{7}{10}$	$\frac{9}{20}$	$1\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	1
$\frac{1}{8}$	$\frac{5}{8}$	$\frac{3}{8}$	$1\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{5}{8}$
1	$1\frac{1}{2}$	$1\frac{1}{4}$	2	2	3	5

Key

1 point	Overs (15)
1 point	Unders (17)
2 points	Equals 1 (4)

This table shows that 17 of the possible outcomes are unders and 15 are overs. So the "unders" player has slightly more chance of scoring a point than the "overs" player.

b. If you were to play a large number of games, you should find that the game is biased slightly in favour of the "unders" player. Explanations will vary but could be similar to that given in **4a**.

Page 24: Dylan's Dominoes

Activity

1. 8, 9, 10, 11, and 12 are missing.
2. **a.**

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

   b. 6, 7, and 8. There are 6 ways to get a total of 7 and 5 ways of getting either 6 or 8. There are fewer ways to get any of the other totals.

3.–4. Practical activity. A bar graph or a dot plot could show your results clearly. Conclusions will vary.

Figure It Out

Statistics

Revised Edition

Teachers' Notes

Overview of Levels 3–4

What Is Statistics About?

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".

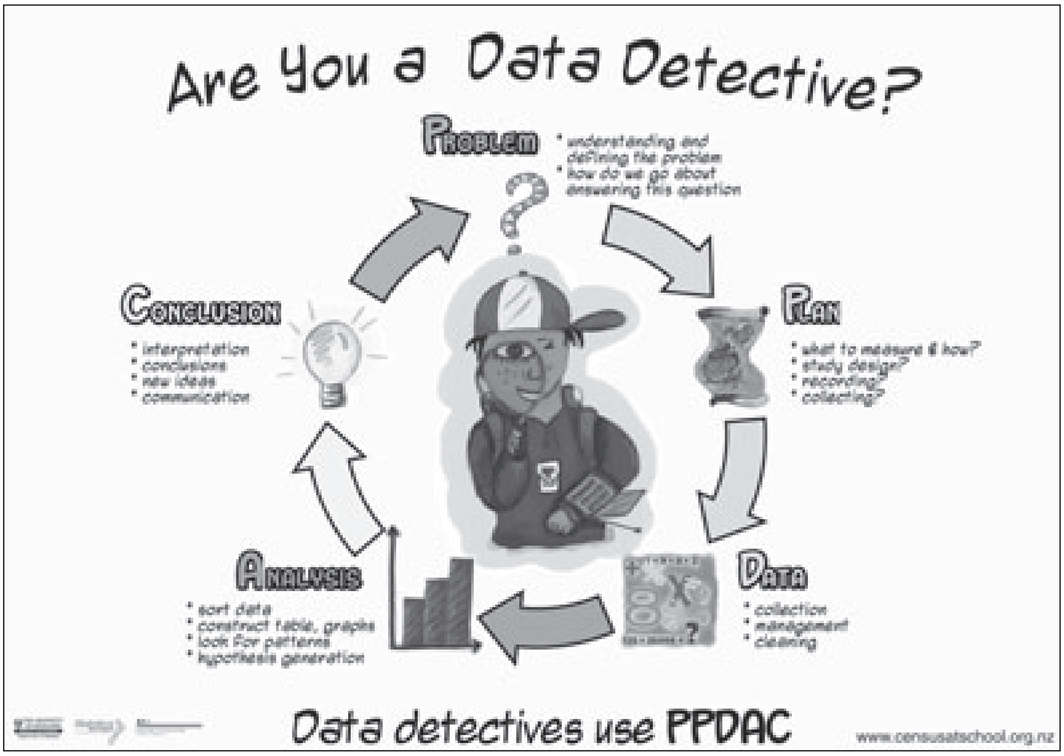
The New Zealand Curriculum goes on to say:

> 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 investigation cycle used for the CensusAtSchool New Zealand resources (see www.censusatschool.org.nz) provides an ideal model for statistical investigation. This approach is used in the revised level 3–4 Figure It Out *Statistics* 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 in this model are:

- Problem – deciding what to investigate, and why, and how to go about it;
- Planning – determining how to gather the necessary data;
- Data – collecting, managing, and preparing the data for analysis;
- Analysis – exploring the data with the help of graphs and statistical tools and asking what it says;
- Conclusion – determining how the data answers the original problem and deciding what to do next.

CensusAtSchool New Zealand provides this information in the form of a downloadable PDF.

Much of the information in the following sections is adapted (by 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 a non-numeric attribute such as gender, colour, style, model, opinion, type, feel, and so on. For example, foods could be sorted into categories such as meat, fish, vegetables, fruit, and cereal.

Numeric 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. **Continuous 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.

A **variable** is an attribute or factor that can take on different values, for example, time, colour, length, favourite author, number of items, cost, age, temperature. When a number of pieces of data are collected for a single object 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.

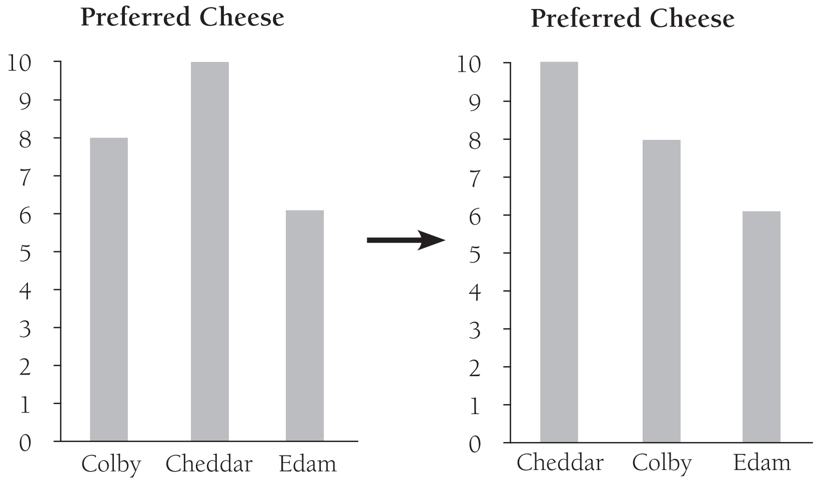
The activities in the students' book promote graphs as a means of exploring data and communicating findings. It is important that students learn to "read" graphs, question what they read or see, find the stories in the data, and ask more questions. Especially in the early years, students can devise their own graphical representations. 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. These basic conventions are designed to enhance the communicative power of graphs.

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. 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.

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.



Bar graphs can be used to show a second variable (such as gender or year). When this is done, the bars for the two variables are displayed side by side, without a gap. The two variables must be coloured or shaded distinctly and a key provided.

In the students' book, examples of bar graphs are found on pages **6** and **8–9**.

In a **stacked bar graph**, bars are divided into sections representing each data category. In a **100% stacked bar graph**, the height of the bar represents the entire data set. Just like the sectors in a pie chart, the size of each section is proportional to the frequency with which that category features in the data. But 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 data set.

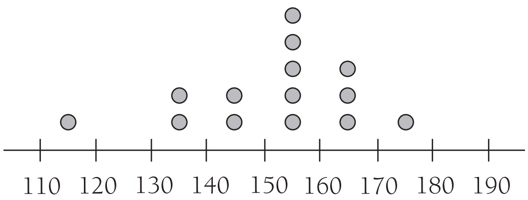
In the students' book, a stacked bar graph appears on page **5** and a 100% stacked bar graph on page **6**.

Pie charts 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.

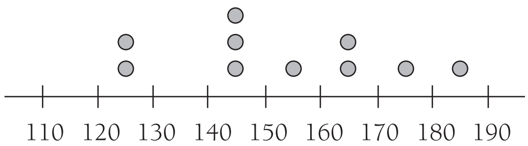
While they have their place, pie charts and strip graphs can only be used for a single variable and can only tell the simplest of stories. The only pie chart in the students' book is found on page **8**.

Dot plots, used on pages **6**, **11**, and **15** of the students' book, are very easy to construct and clearly show the spread of the data involved (that is, the way in which it is distributed and/or grouped). They suit discrete numerical 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 also be grouped, as in the following dot plots, which are similar to a histogram.



Girls' Arm Span Length (cm)



Boys' Arm Span Length (cm)

Stem-and-leaf graphs, explored on pages **12–13** 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.

A stem-and-leaf graph is made by arranging numeric data in a display, using the first part of the number as the stem and the last digit as the leaf. For example, for 16 in the data set {16, 31, 25, 33, 27, 24, 14, 26, 31}:

- the tens digit (1) is the stem and appears to the left of the vertical line
- the ones digit (6) is the leaf and appears to the right of the vertical line.

Stem	Leaf
1	4 6
2	4 5 6 7
3	1 1 3

It is usual to sort the leaf data in number order, from least to greatest. This is often best done as a second step, particularly where there is quite a lot of data.

When constructing stem-and-leaf graphs, students should ensure that they space the digits equally. This is important because it makes it possible to observe features such as general shape, symmetry, gaps, and clusters. It also makes it easy to track down the median (and, in a larger data set, the quartiles).

If the data collected is three-digit numbers, such as height in centimetres, the hundreds and tens digits make the stem of the graph, with the ones digits as the leaves. For example, if graphing the following heights recorded in centimetres, 114, 122, 142, 116, 125, 127, 142, 144, the stem represents the hundreds and tens digits:

Stem	Leaf
11	4 6
12	2 5 7
13	
14	2 2 4

In Scandinavian countries, stem-and-leaf graphs are used for bus timetables. For example:

Hour	Minute
06	00 25 40
07	05 23 37 48 55
08	00 14 26 33 47 52
09	02 13 46

You could discuss with your students the suitability of this system for New Zealand. What system is currently used for our bus timetables?

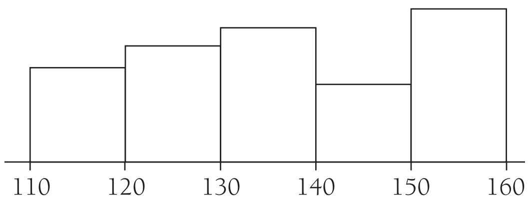
Two sets of numeric data can be displayed in a **back-to-back stem-and-leaf graph**, in which the stem is shared. This makes it possible to compare the shapes of the two distributions without losing the individual values. This graph is also useful for finding summary statistics, such as the median and quartiles.

Girls' Arm Span (cm)		Boys' Arm Span (cm)
8	11	
	12	5 8
0 8	13	
0 7	14	4 4 4
0 5 5 6 9	15	4
2 4 5	16	3 3
5	17	5
	18	2

Students should take care with the order of the leaves on back-to-back stem-and-leaf graphs; they must be consistent throughout.

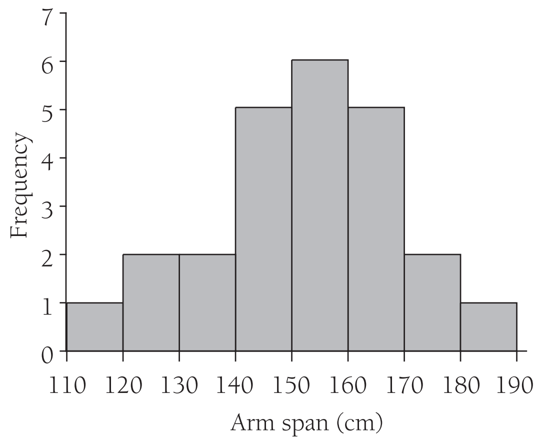
A **histogram** looks similar to a bar graph but, in this case, the bars touch. Histograms are used for continuous data (for example, height). Histograms are introduced on pages **12–13** of the students' book.

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, 127). The sides of the bars represent the beginning and end of an "**interval**". The bars are best labelled using the outer limits of each interval:



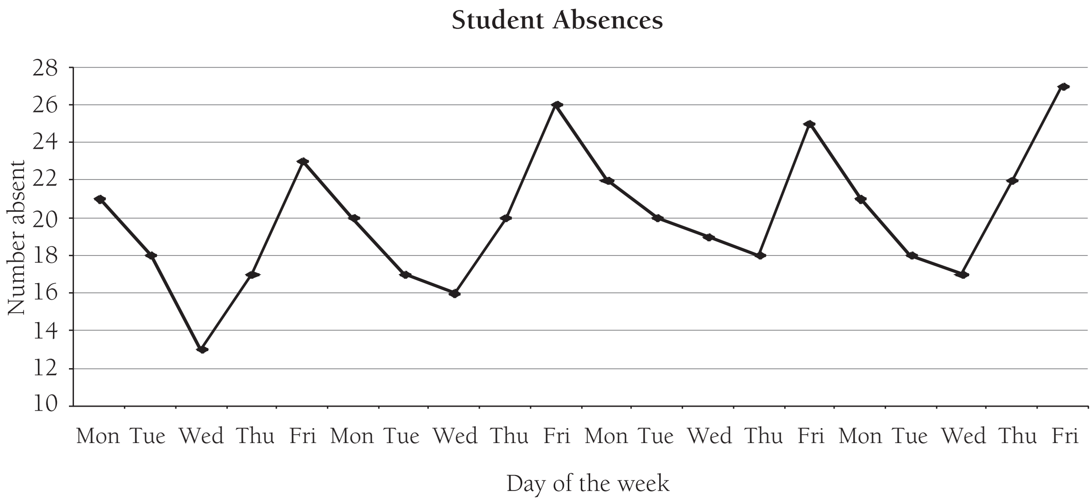
As a guide, 10 or 12 is a reasonable maximum for the number of bars. Any more, and it would be difficult to see patterns. When deciding on a suitable interval, consider both the greatest and least values and the total number of values in the data set. Use "natural" steps, such as 2, 5, 10, 20 (not 3, 6, 7, 8, 9, and so on), and keep them the same throughout. As an example, the arm span data in the stem-and-leaf example above ranges from 118 to 182 centimetres. There are 24 values in the data set. This suggests a histogram with eight intervals of 10 centimetres each:

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

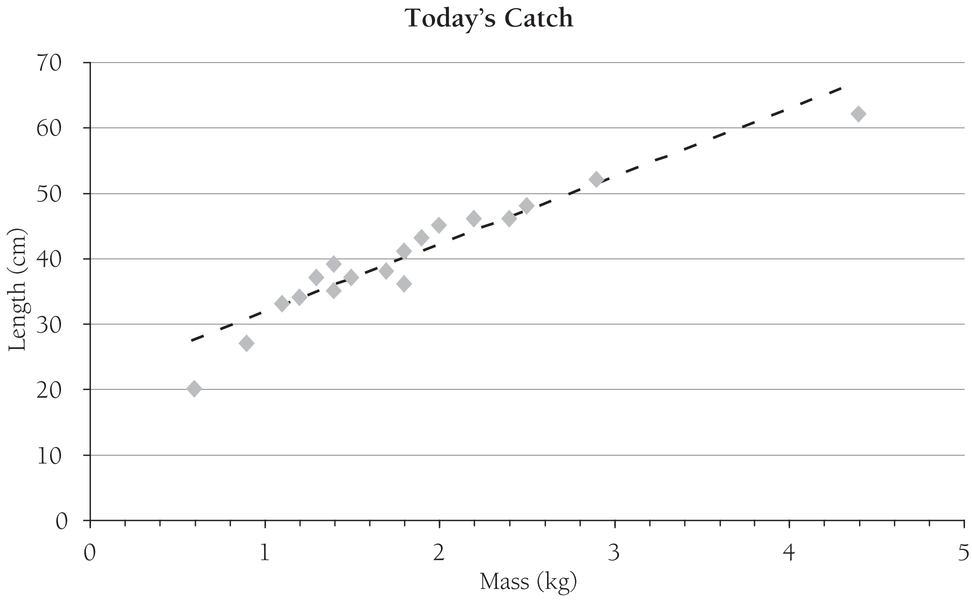


Line graphs compare two variables, one of which is plotted on the horizontal axis and the other on the vertical. Line graphs are useful for showing one variable in relation to another (for example, tree growth in relation to rainfall) or making predictions about the results of data that has not yet been 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 should not be used for category data.

Line graphs are excellent for showing how something changes over time. A **time-series graph** (see pages **2–3** in the students' book) is a line graph in which time is measured on the horizontal axis and the variable being observed is measured on the vertical axis.



A **scatter plot** graphs bivariate data (data in two variables) as a series of separate points, as in the example below (see page **13** of the students' book). 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 excellent for showing whether or not there is a relationship between the two variables graphed. A relationship is suggested when the points plotted are clearly concentrated along some invisible line, as in the example. The computer can attempt to draw this invisible line: choose Add Trendline from the Chart menu.



Other Statistical Terms

Axes (singular: **axis**) are the two lines, one horizontal and one vertical, that form the framework for most graphs. As a general principle, the vertical axis is used for frequency and the horizontal axis for categories, values, or time. This means that bars are equally spaced and vertical. (Pictographs and bar graphs do not always observe this rule.)

Bias is a statistical distortion created by some factor that was overlooked when data was being gathered (for example, a poorly worded survey question that was misunderstood by many respondents).

Collate means to collect and combine.

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

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

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 between known data values. For the purposes of both interpolation and extrapolation, it is assumed that the observable pattern continues between 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	4	6	9	5	2	3	1

"**I wonder**" questions are **investigative questions** – statistical questions or problems to be answered or solved. They consider the entire data set or population and do not involve locating an individual within a data set.

Two types of investigative question 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?").

When continuous (measurement) data is grouped in a frequency table or histogram, the groups are called **intervals**. An interval is defined by its limits, for example, "greater than or equal to 3 but less than 4". Intervals can, at a later stage, 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").

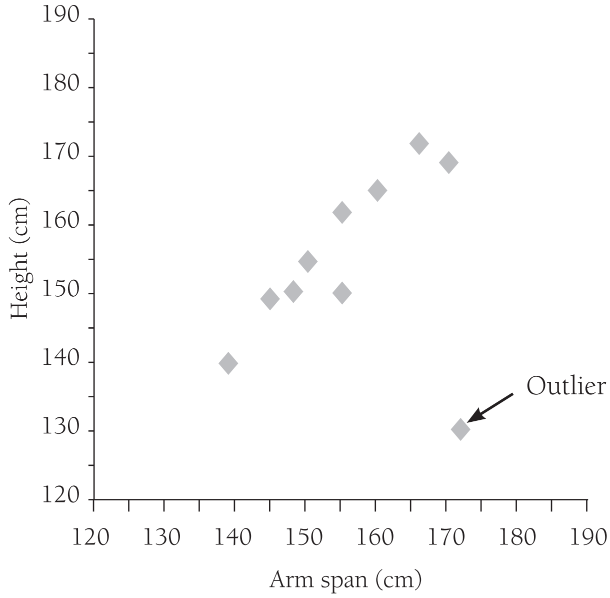
The **mean** of a data set is the sum of all the values in the set divided by the number of values. It is sometimes called the average.

The **median** is the middle value in a data set when all the values are arranged in order from smallest to biggest or biggest to smallest.

The **mode** is the most commonly occurring value in a data set.

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 some other factor.) Outliers can affect the average (or mean) quite considerably.

Relationship between Arm Span and Height



Probability:

- **Probability** and **chance** are the same thing, although one or the other term may be more usual in a particular context.
- **Outcome**: the result of a trial (for example, a match or no match)
- **Trial**: performance of an action or actions where the outcome is uncertain (for example, the toss of a coin)
- **Experiment**: sometimes used interchangeably with trial; otherwise, a series of trials
- **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.

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.

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. For example:

Footware	**Tally**	**Frequency**
Shoes	𝍸	5
Sandals	𝍸 II	7

A **trendline** is a line on a graph that indicates a statistical trend. The graphing program assesses the plotted values and marks a middle course through them.

Variation is the term used to refer to the differences between data, particularly differences from an expected pattern or trend. To illustrate: 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
Heads	41	44	52	53	42	49	47	50	38	45
Tails	59	56	48	47	58	51	53	50	62	55

Variation can be described and, at later levels, measured, using a variety of measures of spread from the simple to the sophisticated.

Links to *The New Zealand Curriculum*

Achievement Objectives

Achievement objectives in the teachers' notes for existing, revised, or new material in the *Statistics* students' book are from the mathematics and statistics area of *The New Zealand Curriculum*.

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 revised *Statistics* book suggest one or more key competencies that could be developed within those activities. (You may, of course, decide to focus on key competencies other than those suggested.)

Activity Notes

Page 1: Left to Chance

Achievement Objective

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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Left to Chance can be used to develop these key competencies:

- thinking
- participating and contributing.

Activity

This game explores the concept of "fairness" in relation to the chances of winning or losing a particular game. Even though the game, with its symmetrical board and equal number of "win" and "lose" destinations, looks as if it should be fair, the two **outcomes*** are not at all equally likely.

The Green Game uses the first four rows of the board and the extended game uses all five rows.

Unless you are helping those who may struggle to read the instructions, it is suggested that any introduction to the game should be minimal. The students are likely to get most out of it if they have not been told what to expect. Therefore, as the game progresses, they must try and find out for themselves what is going on.

What is important is that the students keep a record of outcomes. This is the data on which they base and justify their conclusions.

The pattern of results should become clear after a few games: players are much more likely to end up on either of the middle two (lose) circles than on either of the outside (win) circles. This pattern will be confirmed as further games are played and results collated with those of other students.

Although the students are not asked to do this, a suitable tool for finding and listing all the outcomes for both the games is a tree diagram. The Answers show how to set one out.

When the students list all the pathways or outcomes (question **3**), they will find that there are a total of 8 and that, of this total, only 2 (LLL and RRR) end in a win. All the outcomes are equally likely, so a player has only a $\frac{2}{8} = \frac{1}{4}$ **probability** of winning (and a $\frac{6}{8} = \frac{3}{4}$ probability of losing). This is the **theoretical** explanation behind the **experimental** findings.

Some students will realise that the probability of any particular outcome is given by the calculation $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. This is because, at every choice point on the game board, the probability that the desired path will be followed is $\frac{1}{2}$.

* For bolded terms, see introductory section.

Some students may also notice that the probability of a win plus the probability of a loss equals 1 ($\frac{1}{4} + \frac{3}{4} = 1$). 1 is certainty in terms of probability. It is certain that, if Simon's game is played, the result will be either a win or a loss; there is no other possible outcome.

Discuss what is meant by equally likely outcomes; not all outcomes are equally likely, and it is important that students realise this. Because of its fundamental symmetry, a coin is equally likely to give a head or a tail when tossed. A milk bottle top does not have such symmetry and is more likely to fall open side up.

Question **4a–b** extends the game with a fifth row of circles. The students will find, this time, that the probability of a win is $\frac{8}{16} = \frac{1}{2}$. From Simon's point of view, this would be a very bad fund-raising proposition!

Pages 2–3: School Crossing

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions;
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

School Crossing can be used to develop these key competencies:

- thinking
- using language, symbols, and texts
- participating and contributing.

Activity

This activity requires the students to interpret a **time-series graph*** and assess the validity of statements made about it.

The graph shows **discrete numeric data** in two series (number of vehicles in the morning and in the afternoon). Although time (horizontal **axis**) is a continuous **variable**, the data has been grouped by day, and so the time data is effectively treated as discrete.

* For bolded terms, see introductory section.

As they work through the different statements, the students will:

- make comparisons between traffic flows on different days
- make comparisons between traffic flows in the morning and the afternoon
- make generalisations, based on the limited data available.

As an orientation, have the students look quickly at the graph and briefly describe a few obvious features.

Like a number of the other activities in this book, question **1** is designed to give students models of the kinds of statements they can make when looking at graphs. It is also designed to make them search for data that will either support or refute the statements and, if necessary, use data to resolve disagreements.

Question **2** is designed to have students look for other data-supported patterns.

Question **3** asks students to think about the human factors that may be behind two of the patterns that have been noted. Whatever explanations they come up with, these will be hypotheses. Challenge them to suggest how they might test these hypotheses.

Question **4** relates to the issue of variability. There are almost always fluctuations in data, even in carefully controlled situations. These fluctuations may or may not be significant. In this case, some of the **variation** may be random. The only way to find out is to gather data for a longer period.

Investigation

Encourage the students to come up with a purpose for this investigation – a reason for going to the trouble (apart from that the activity told them to!). The purpose may be as simple as having fun investigating something of interest or as serious as investigating a safety concern. You could ask:

- Who would be interested in the results of your survey?
- How would these results be useful?

Before the data gathering begins, ask the students to predict and write down what patterns they might find.

Pages 4–5: Change Ahead!

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions;
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Change Ahead! can be used to develop these key competencies:

- thinking
- using language, symbols, and texts
- relating to others.

Activity One

In this activity, the students interpret a **tally chart***. Although they will have met tally charts before, this one is likely to be new to them. Tally charts typically involve a **variable** (for example, favourite reading), but this one involves three: colour 1, colour 2, and parent/student. The boxes in the grid have been divided in two to cope with the third variable.

Tally charts usually have a column for totals and students may find it useful, when answering question **1**, to create a table of totals, something like this:

	Red & grey	**Yellow & grey**	**Green & grey**	**Blue & grey**	**Red & black**	**Yellow & black**	**Green & black**	**Blue & black**
Students	12	5	5	8	16	2	3	9
Parents	10	3	2	8	4	6	12	15
Total	22	8	7	16	20	8	15	24

Paora's view, expressed in question **2**, is worth discussion. The survey does not produce a clear-cut option that is favoured by students and parents alike, and Paora's suggestion gives greater weight to student opinion than parent opinion:

- Red is the feature colour most popular with the students (28 votes) and, although red and black gets a few more votes (16) than red and grey (12), overall black and grey are equally popular (30 votes each).

* For bolded terms, see introductory section.

- Blue (23 votes) is considerably more popular than red (14 votes) with parents, who also overwhelmingly prefer black (37 votes) to grey (23 votes).

An alternative would be to weight the decision in favour of the parents' views and recommend blue and black. After all, they usually pay for the uniform (and virtually everything else!).

A compromise suggestion would be red and black on the grounds that students overwhelmingly prefer red to blue and parents prefer black to grey.

Activity Two

In question **1**, the **stacked bar graph** highlights the total support for each colour option. Each bar of a stacked graph is like a **strip graph**. Although the sub-totals for parents and students can be calculated or estimated from the graph, the reader's attention is focused on the combined totals.

For question **2**, the students will probably follow Ata's suggestion and use a double bar graph. A double bar graph will facilitate comparisons between the two groups surveyed.

Activity Three

This activity requires the students to develop their own questionnaire. To do this, they should (i) work out what information they need and (ii) work out what questions will be most likely to get them that information. The exact wording is important. Also important is how responses are to be gathered. It may be that a five-point Likert scale (see next activity) is the best way of judging the strength of people's views. It is very difficult to ask good survey questions and provide appropriate options for response. Even professional research organisations ask ambiguous or otherwise unclear questions and may frustrate respondents by not providing options that fit them. For this reason, students should always try their questions out on several classmates before using them "in the field".

Pages 6–7: Opinion Polls

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions;
  - identifying patterns and trends in context, within and between data sets (Statistics, level 3).
- 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 (Statistics, level 4).

Statistical literacy

- Evaluate the effectiveness of different displays in representing the findings of a statistical investigation or probability activity undertaken by others (Statistics, level 3).
- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Opinion Polls can be used to develop these key competencies:

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

In these activities, students interpret different graphs showing the results of an opinion poll. They also devise questions suitable for an opinion poll that uses a Likert scale.

Activity One

This activity makes use of a kind of **category data***: the categories are the five opinions that respondents are allowed to hold; but unlike most category data, the categories come with an inbuilt order that needs to be maintained. Many surveys use a 5-point scale like the one in this activity (often referred to as a Likert scale – see the notes for **Activity Two**).

Questions **1** and **2** relate to what the graphs show and how well they show it. The **tally chart** and graphs **i** and **ii** are just different ways of presenting the same thing: in each, a tally mark, symbol, or unit of height represents a vote. Suggest to your students that they imagine graph **i** rotated 90 degrees clockwise and then compare it with the tally chart. Suggest also that, by cutting and pasting, graph **i** might easily be turned into graph **ii**. How?

Graph **iii** is a completely different kind of graph: a **100% stacked bar graph**, which is a form of **strip graph**. Like a **pie chart**, this type of graph shows the proportion (rather than the number) of data belonging to each category. Be aware that students find it difficult to manually create

* For bolded terms, see introductory section.

100% stacked bar graphs and pie charts because they must calculate and then mark off the correct parts of the strip or circle. This requires proportional reasoning as well as measurement and construction skills. Both types of graph are, however, very easy to create by computer.

Graph **iii** is the only graph that takes account of the gender imbalance. In effect, each boy gets slightly more of the strip's length than each girl. This is why the two "strongly disagree" sections are of slightly different length even though the data shows that the number of votes in this category was the same (2) for both boys and girls.

Activity Two

Opinion polls often present assertions and ask the respondent to agree or disagree with varying degrees of emphasis.

The Likert Scale, named after its orginator Rensis Likert (an American educator and organisational psychologist, 1903–1981), is an ordered, one-dimensional scale, typically offering between four and seven options for respondents to choose from (five is particularly common).

Numbers (such as 1 to 5) are usually assigned to each option. The options typically also include labels, although sometimes only the extremes are labelled. The Likert scale is also known as the "summative scale" since a result is often obtained by summing the number values attached to the various response options.

Examples:

I like going to Chinese restaurants

Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree

When I think about Chinese restaurants, I feel:

Good		Neutral		Bad

A benefit of the Likert Scale is that it offers unambiguous options and so leads to consistent answers. A disadvantage is that the Likert scale offers only a few options and respondents may not fully agree with any of them.

There is much debate about how many options should be offered. An odd number allows respondents to sit on the fence, while an even number forces them to take a side even when this does not reflect their true position. The more options offered, the greater the chance respondents will be able to match their opinions to the available options, but a greater number of options means less clarity and more work for the administrator.

As with any other measurement tool, the questions and statements and the options offered should be carefully calibrated so that together they give a useful and coherent picture.

If carrying out a survey that uses a Likert scale (or even a three-point scale: agree / neither agree nor disagree / disagree), students need to be aware that negatively phrased questions or statements (for example, "Birds should never be kept in cages") are often misinterpreted and invite respondents to circle options that actually say the reverse of what they mean. Positively expressed, non-extreme statements or questions are preferable.

Activity Three

This activity is best done as a homework exercise. The students can then share their findings with their classmate the following day. For more able students, this activity could lead to discussion about **bias**.

Pages 8–10: Reading Trends

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- Plan and conduct investigations using the statistical enquiry cycle:
  - comparing distributions visually;
  - communicating findings, using appropriate displays (Statistics, level 4).

Statistical literacy

- Evaluate the effectiveness of different displays in representing the findings of a statistical investigation or probability activity undertaken by others (Statistics, level 3).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Reading Trends can be used to develop these key competencies:

- thinking
- using language, symbols, and texts.

Activity One

In this activity, **multivariate category data*** is summarised in a **frequency table**. The data is merged so that year-group and gender distinctions disappear and two graphs highlight the relative popularity of different authors.

In question **2**, the students use a computer to create two further graphs that use the same (merged) data. Three possibilities are given in the Answers. Out of all the options, a **bar graph** (vertical or horizontal) with the authors sorted according to popularity may be the best.

Be aware that, when using a computer to graph **category data**, it is not obvious how to get the category labels on the graph. In some graphing programs, the procedure is to forget the category names and select only the number (frequency) data when choosing and creating the graph. Add the category labels as a second stage: click on the graph you have created to select it, go to the *Chart* menu, click on *Source data*, then the *Series* tab. Insert the cursor in the *Category (X) axis labels* box and then select the cells in the spreadsheet that contain the category labels. Click OK.

Discourage your students from using the unusual and three-dimensional graphs that may be on offer. These are almost always more difficult to read than two-dimensional graphs, particularly when it comes to making comparisons between two categories or sets of data. Clarity and simplicity should be the aim.

Activity Two

This activity requires the students to create and/or interpret three double bar graphs. Encourage them to sort the data in the spreadsheet in ways that might be useful (for example, order of popularity with boys). As they do this, the bars of their graphs will automatically rearrange

---

* For bolded terms, see introductory section.

themselves to match. By sorting one of the **variables**, the students will find it considerably easier to spot significant features and patterns.

Before starting this activity, students should have had some practice at writing statements about data in graphs. For examples, see the earlier activities in this book, School Crossing and Change Ahead! See also the Level 3 *Statistics* book.

Activity Three

In question **1**, the students go back to the tables and graphs they have used and/or created and do a little detective work. In addition to naming the author(s) in each case, they should cite the supporting data (as in the Answers).

Question **2** is designed to make the students think about the size of the sample from which these statistics have been drawn. To answer the question, they need to have read the introductory paragraph and have seen that students can vote for as many authors as they wish. This means that they can't know how many students voted. It does not mean, however, that we know absolutely nothing about the sample size: 69 votes were cast for E. T. Auberg, so we know that at least this number of people took part in the survey (assuming, of course, that no one voted more than once for E.T. Auberg!)

Page 11: Where's My Bus?

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- Plan and conduct investigations using the statistical enquiry cycle:
  - communicating findings, using appropriate displays (Statistics, level 4).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Where's My Bus? can be used to develop these key competencies:

- thinking
- managing self
- relating to others.

Activity

In this activity, students compare six graphically presented data sets. It may be the first time that they have encountered a graph of this kind, but all it consists of are six **dot plots*** with a common **axis**. The different presentation should remind them that there is more to graphing than **pie charts** and **bar graphs**. When confronted with data and no specific instructions about how to graph it, they should be flexible and, if necessary, creative when it comes to determining a style and format. This particular graph has a title and axis labels that are sufficient and clear in the

* For bolded terms, see introductory section.

context, has minimal clutter, immediately conveys meaning, and invites detailed scrutiny. These are all characteristics of good graphs.

With your students working in pairs, challenge them to make meaning of this graph with a minimum of assistance from yourself.

Note the use of "–2" on the time axis to denote an arrival that is 2 minutes early. Although this use may be unexpected, could it have any other meaning?

If help with question **1** is necessary, collaboratively develop a description for Aroha's bus and then have the students write the other descriptions themselves. Encourage them to begin with the most important generalisation they can make ("Aroha's bus was almost always late") rather than details ("Twice, Aroha's bus was between 1 and 2 minutes early"). They can then tease out their generalisation and add details.

Question **2** hinges on students knowing what "reliable" means in this context. It will help if they can imagine themselves waiting for each of these buses. Which would frustrate them most?

Question **3** is designed to encourage mathematical disputation. While "answers" are often not clear-cut (and will therefore vary), this is not the same as saying that any answer is good enough. The students should expect to have to justify their ordering of the buses on the basis of the data and, if necessary, change it.

Using the **PPDAC cycle**, the students could conduct their own survey of local bus or train arrival times. Prompt them to consider who would be interested in this information.

Pages 12–13: Fish Figures

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- Plan and conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends;
  - communicating findings, using appropriate displays (Statistics, level 4).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Fish Figures can be used to develop these key competencies:

- thinking
- using language, symbols, and texts.

In these activities, which involve measurement data, the students convert **stem-and-leaf graphs*** into **histograms**, create a **scatter plot**, and use **trendlines** to help them make predictions.

* For bolded terms, see introductory section.

Activity One

Question **1** is designed to show the relationship between a stem-and-leaf graph and a histogram: rotate the stem-and-leaf graph 90 degrees anticlockwise, adjust the labelling, and there you have it.

This process illustrates an important advantage that a stem-and-leaf graph has over a histogram. In the first, it is possible to see the shape (or distribution) of a data set *and* still identify individual data values. In the second, the shape of the distribution is just as clear but the individual data values are submerged in the groups (**intervals**); detail has been lost.

A further advantage of the stem-and-leaf graph is that the data count and the **median** (and, in larger data sets, the quartiles) can easily be found. The data count cannot be found from a histogram and, for **continuous** (measurement) **data**, the median and quartiles can only be estimated.

The students should ensure that the leaves on their stem-and-leaf graph are evenly spaced so that, when it is rotated, it matches the histogram and it is easier to see which intervals have the greatest frequency.

Activity Two

The students need to have access to and an understanding of how to use a computer spreadsheet/graphing program for this activity (though, for a small data set such as this, a **scatter plot** can also be drawn by hand reasonably easily).

This may be the first time that your students have come across scatter plots. This type of graph (also known as an XY scatter graph) is designed to show what relationship (**correlation**), if any, exists between two **variables** (in this case, the mass and length of fish). One of the most useful features of a scatter plot is that each individual data value retains its identity – no data whatsoever is lost in the graphing process. This means that the reader can see both the pattern generated by the data as a whole *and* trace exactly where every individual piece of data (fish, person, book, or whatever) has got to.

So that the students have an idea what to expect, page **13** in the students' book shows an appropriately formatted graph, with two data values in place. (Needless to say, if the graph is created by computer, all the values appear simultaneously. Note that you can add values and update the computer graph at any stage.)

It is clear from the pattern of dots in the completed graph that as the fish gets longer, the mass increases. If there were no such relationship, the dots would be randomly scattered around the graph instead of clustered very obviously along an imaginary line.

Question **2** involves adding a trendline to the graph and then using it to make two predictions. The predicting provides a check that the students understand the meaning of the graph and the relationship it reveals. (You will probably find *Add trendline* under *Chart* on the menu bar of your graphing program.)

Pages 14–15: Fully Stretched

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions;
  - identifying patterns and trends in context, within and between data sets
  - communicating findings, using data displays (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Fully Stretched can be used to develop these key competencies:

- using language, symbols, and texts
- participating and contributing.

In these activities, the students analyse the results of an experiment designed to test the relative strengths of two kinds of rubber band.

Activity One

Question **1** involves creating a **back-to-back dot plot***.

Question **2** involves comparing the two distributions and relating this information to the rubber band scenario. When comparing the distributions, students should consider:

- spread
- maximum and minimum values
- clusters
- **medians**
- **outliers** (for the more able students).

The students should base their conclusions about the rubber bands on this analysis. Encourage them to use words and phrases such as “reliable”, “more consistent”, “**variable**”, and so on, backing all their statements with reference to the dot plots.

* For bolded terms, see introductory section.

After they have reached their conclusions, a whole-class discussion could be a valuable means of giving students feedback on the quality of their thinking.

Activity Two

This activity involves setting up an experiment that is best undertaken in pairs or small groups. It should follow all five steps of the **PPDAC**. It could conveniently be linked with the technology curriculum. With careful planning, it could provide a rich learning experience.

Some thought needs to be given as to what your role will be in terms of giving advice and providing resources. Ideally, the students will be given maximum independence. If so, they may come up with very different approaches to the problem. You are obviously responsible for setting up an environment in which genuine experimentation is encouraged, but dangerous or foolish conduct is a no-no.

The activity calls for the students to write a brief report. This report should cover:

- the reason for the research
- the methodology
- the results
- an analysis of the results
- notes on any problems encountered
- a comment on the validity of the results
- a final conclusion
- reflection (for more advanced students).

The depth of this report will vary considerably from group to group, depending on the depth of understanding and the literacy skills available in the group.

Activity Three

This broadly-defined activity invites students to generalise from the experience they have gained in **Activity Two**.

Pages 16–17: Bean There

Achievement Objectives

Statistical investigation

- Conduct investigations using the statistical enquiry cycle:
  - gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions;
  - identifying patterns and trends in context, within and between data sets;
  - communicating findings, using data displays (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Bean There can be used to develop these key competencies:

- thinking
- relating to others.

Activity

This activity involves creating, reading, and interpreting a four-series **time-series graph*** in some detail. It is followed by an investigation in which students replicate an experiment, following the stages of the **PPDAC**. This statistical activity could easily be linked to the science learning area of the curriculum.

In question **1**, the students use a computer spreadsheet and graphing program to create a time-series graph. Depending on the graphing program, steps may include:

- Open a spreadsheet and enter the data from the table on page **16** of the students' book. The data can be entered just as it appears in the table but without the title or the GrowSmart and GardenBest headers.
- Select (by dragging the cursor) the data in the four columns headed GS1, GS2, GB1, and GB2 (select the column headers, too).
- Select the *Line graph* option from the *Chart* menu. Add the title and **axis** labels to your graph.
- Go to the *Source data* menu, click on the *Series* tab, click cursor in the *Category (X) axis labels* box, highlight the numbers in your spreadsheet below the header Day (0, 2, 4, 6 …). Click OK in the *Source data* dialog box.

* For bolded terms, see introductory section.

- If necessary, use the *Chart options* menu to add or fix labels.
- If necessary, change the format of the four data lines (for example, make the two GS lines dotted) so that it is easy to see which is which (right-click on the lines for a menu of line types and colours).

In question **2**, the students work in pairs to assess the accuracy of various statements made by Jyoti's friends. They should write their judgments down and justify them with reference to their graphs.

This is a reasonably sophisticated reading and interpretation task. Depending on their experience of graphs and mathematical understanding, the students will need varying levels of support. Four key areas are:

1. Interpreting the lines: the horizontal axis shows the days of the experiment; each line represents one bean; upward slope indicates growth; the steepness of a line indicates the rate of growth (steeper is faster); horizontal parts of a line mean no growth; we can read how much growth (in centimetres) by referring to the vertical axis.
2. Comparing growth: wherever one line is above another, that bean is taller than the other; where two lines cross each other, one bean is overtaking in height; you can determine the relative heights of the beans on any particular day by reading values off the vertical axis.
3. Making generalisations: while the table and the students' graphs contain a lot of data, a picture does emerge; statements that broadly describe this picture are known as generalisations; false generalisations can be refuted with reference to the data or to experimental procedure.
4. Experimental procedure: certain conditions must be met if a science experiment is to be fair. Some of the statements on page **17** require some awareness of these conditions (for example, statement **ix**, which relates to the notion of a "control").

Note that statement **v** requires the students to consider what might happen if the data gathering were extended. This involves **extrapolation**. Extrapolation is generally much less reliable than **interpolation** because there can be no guarantee that previous trends will continue beyond the known period.

Once students have answered question **2** in pairs, have them discuss their answers with another pair of students and, where there is disagreement, encourage data-based disputation, with the aim of reaching an agreed position.

Investigation

For this investigation, the students need to have a clear understanding of the PPDAC.

Encourage the students to use statistical language in their analyses and their conclusions. They should be gaining confidence in using words such as theory, investigate, data, justify, predict, **ratio**, distribution, middle, range, spread, cluster, **outlier**, frequent, sample, **scatter plot**, random, relatively, generalisation, systematic, biased, selection, **bivariate**.

Pages 18–19: Take Five

Achievement Objectives

Probability

- Investigate simple situations that involve elements of chance by comparing experimental results with expectations from models of all the outcomes, acknowledging that samples vary (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Take Five can be used to develop these key competencies:

- thinking
- participating and contributing.

In these activities, the students use sampling as the basis for predictions about the nature of an unseen whole.

Activity One

This activity is about **probability***, sampling, and **variation**. No one selection of 5 tiles can reveal what the bag contains. Each selection is likely to be different from the last. But the more samples we take, the closer we get to the truth about the contents. Question **1** is there to get the students engaged in extracting and synthesising the useful information that progressively becomes available in the sequence modelled by Malia and Natalie. This is a very good exercise in logical deduction.

The fifth sample provides the clinching piece of information. Up until now, we had been able to account for 9 of the 10 tiles. It turns out that the missing tile was blue.

The answer for question **5** implicitly warns against falsely generalising that 5 samples will always be enough. The students need to understand that Malia and Natalie could have gone on sampling the tiles for a very long time without getting a sample with 4 blue tiles in it.

Activity Two

Question **1** asks the students, working in pairs, to repeat Malia and Natalie's **experiment**, using a different selection of 10 coloured tiles. If a third party makes up the contents of the bag and then gives it to the pair carrying out the experiment, the latter two can share the thinking.

Question **2** involves a similar experiment, but with double the number of tiles in the bag. The chances of correctly predicting the contents of this bag after taking just 5 samples of 5 are extremely low. But each 5 × 5 sampling will reveal important information. After several more 5 × 5 samples, there will be even more information but probably no certainty.

It is important that the students realise that if they add all their samples together, it is likely that the most common colours in the bag will have been drawn out more frequently than the less common. This information can be used to predict how many of each colour are in the bag.

* For bolded terms, see introductory section.

This prediction should be refined by cross-checking against the information gained from the separate 5 × 5 samples. Finally, the bag is tipped out and the contents compared with the prediction.

It is important that students understand that:

- a prediction is not the same as a guess: it is a carefully considered estimate based on the best use of the available information;
- not all predictions are equally good or valid: a prediction that ignores a piece of useful information is a careless prediction;
- using exactly the same information, people may come up with different predictions, either because the information is scanty or because of different (but reasonable) assumptions;
- a well-thought-out prediction can still be wrong: if there is enough information to be certain, it's not a prediction, it's a logical deduction (as in **Activity One**, question **4**).

Either or both of these experiments could usefully be carried out a number of times.

Note that sampling a bag of 20 tiles, 5 at a time, does not give as much information as sampling a bag of 10 tiles. If your students find it too difficult to make reasonable predictions when working with the larger number of tiles, tell them to increase the sample size to 8 or 10 and see how they get on.

Pages 20–21: Picking Pocket Money

Achievement Objectives

Statistical literacy

- Evaluate the effectiveness of different displays in representing the findings of a statistical investigation or probability activity undertaken by others (Statistics, level 3).
- Evaluate statements made by others about the findings of statistical investigations and probability activities (Statistics, level 4).

Probability

- Investigate simple situations that involve elements of chance by comparing experimental results with expectations from models of all the outcomes, acknowledging that samples vary (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Picking Pocket Money can be used to develop these key competencies:

- thinking
- managing self
- relating to others.

Activity

In this activity, the students carry out a simulation and give advice based on the results of that simulation.

Shelley must choose between two options: one involves certainty; the other, uncertainty.

In question **1**, the students have to think about what the money dice option might mean for Shelley. At the one extreme, she could get as much as $520 in one year (by throwing $10 on the dice every week); at the other, she could end up with as little as $52 (by throwing $1 every week). Is a guaranteed amount ($260) the best choice? Listening to the students' discussion and reasoning will give you some insight into their understanding of **probability***. Students of this age sometimes think that some numbers on a dice are harder or easier to throw than others.

Question **2** is the simulation, with the 52 trials each representing a week. When the students have completed their dice throwing, they need to calculate the grand total for the year and compare it with the $260 that would otherwise be guaranteed. Depending on the **outcome** of the simulation, either option may turn out to be the winning one.

Question **3** asks the students to reconsider their advice to Shelley. This is a test of their understanding of probability. Those who don't understand it will probably advise the money dice option if their simulation produces a total of more than $260. Those who do understand probability may advise Shelley to take the guaranteed amount even if the simulation favours the dice, on the grounds that a second simulation might turn out quite differently. Then again, the students' advice may also reflect their own values and ways of operating. ("Sure, but it's worth the risk".)

Questions **4** and **5** involve seeking out a larger pool of data. By now, the students should be aware that, when a situation involves **experimentally** determining probability, the bigger the data set, the better.

For your information, if a very large number of **trials** were to be conducted and the results collated, they would show that, in the long run, the money dice would return slightly less than $5 per week. **Theoretically**, the amount is $4.83. This amount, known as the expected value, is the **mean** of the amounts on the six faces of the dice ([1 + 2 + 4 + 5 + 7 + 10] ÷ 6 = 4.83).

This activity may have given students their first experience of a simulation. Simulation is a procedure developed for answering questions about real problems by running experiments that closely resemble the real situation. TTRRC is a model that is commonly used for probability simulations:

Tool:	Select a model to generate outcomes with probabilities that match those of the real situation.
Trial:	Define and conduct a trial, for example, one toss of the dice.
Record:	Record the outcome of the trial; that is, whether it was favourable or its numerical outcome.
Repeat:	Repeat the process a large number of times. The accurate estimation of probability requires the experiment to contain many trials.
Conclusion:	Summarise the information and draw conclusions.

The probability of an event is estimated by totalling the number of favourable outcomes and dividing this result by the total number of trials.

---

* For bolded terms, see introductory section.

Data from a computer simulation

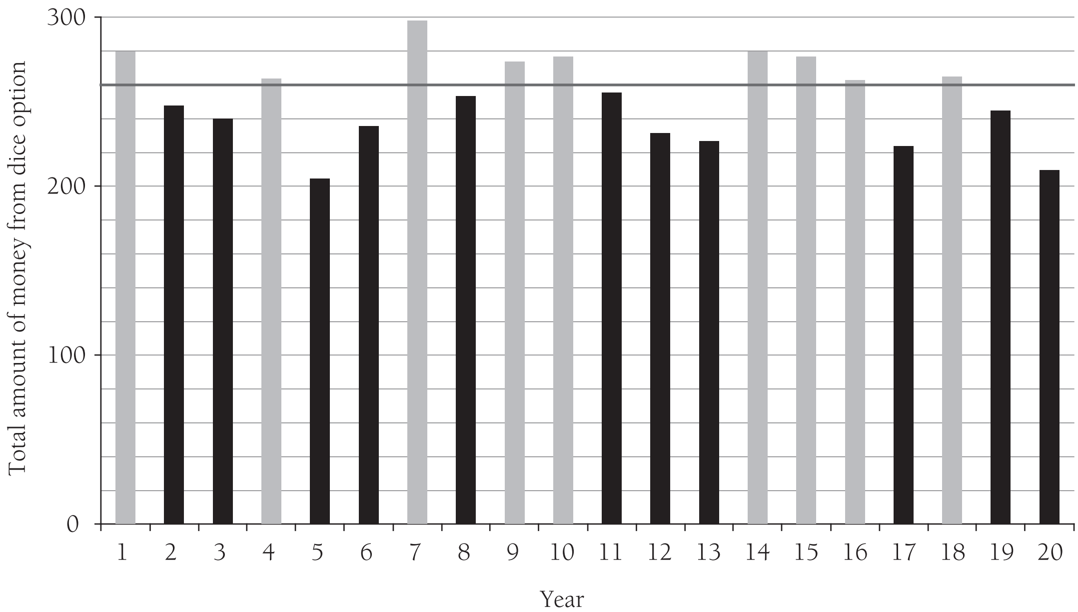
The following data from a computer simulation of the money dice is included for your information and for possible sharing with students. It clearly shows (i) the very wide **variation** from year to year and (ii) the expected return over an extended period of time:

Year	1	2	3	4	5	6	7	8	9	10
$1	6	9	6	6	14	10	4	6	5	5
$2	4	6	12	9	10	11	9	9	8	10
$4	10	11	12	6	9	9	7	12	7	2
$5	11	7	8	12	7	5	10	7	13	15
$7	10	14	6	12	4	9	7	11	10	10
$10	11	5	8	7	8	8	15	7	9	10
Yearly total	280	248	240	262	213	236	299	254	274	278

Year	11	12	13	14	15	16	17	18	19	20
$1	9	12	10	7	7	7	13	9	9	11
$2	11	5	11	5	5	7	8	4	13	13
$4	5	10	9	11	12	12	7	11	7	8
$5	10	12	9	10	8	10	12	12	4	8
$7	5	7	6	7	9	6	5	5	10	7
$10	12	6	7	12	11	10	7	11	9	5
Yearly total	256	231	225	280	278	261	222	266	243	208

The above data can be graphed in a **bar graph** as below and a line added (using the line tool from the drawing toolbar) to show the $260 break-even point. The graph shows that, over the 20 simulated years, the money dice gave a better return than the $5 per week on 9 occasions. But on 3 of these 9 years (4, 16, and 18), the difference was trivial, and on only one occasion (year 7) was the increased return better than $20.

Simulation Data (20 years of 52 weeks)



Over the 20 simulated years, the money dice earned a total of $5,063, while the guaranteed $5 per week brought in $5,200.

Page 22: Crossing the Line

Achievement Objectives

Probability

- Investigate simple situations that involve elements of chance by comparing experimental results with expectations from models of all the outcomes, acknowledging that samples vary (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Crossing the Line can be used to develop these key competencies:

- thinking
- participating and contributing.

Activity

This page involves a **probability*** **experiment**. Unlike coin-tossing or dice-throwing experiments, the probabilities can't be deduced or calculated by simple reasoning.

Before beginning, it would be useful to have a class discussion about the process and the **outcomes** that might be expected because many students will not have done anything like this before. It could be a good idea to conduct a pre-experiment to establish the best height for dropping the toothpick. This will be a height that allows the toothpick to hit the paper in a random fashion without frequently skittering off the edge of the sheet of paper. If it proves difficult to find such a height, use an A3 rather than an A4 sheet of paper. The size of the sheet is immaterial; what matters is that it is carefully ruled up with parallel lines one toothpick length apart. While the height of the drop should be fairly consistent, consistency is not of critical importance.

Questions **2–4** all involve assigning a simple fraction to **experimental** results. Some students may be very uncertain as to how they should do this, in which case the process should be discussed and modelled.

Here is one way of doing this:

- Brainstorm which fractions might be included under the term "simple fraction" ($\frac{1}{2}$, thirds, quarters, fifths, and tenths). Write these on the board.
- Noting that each experiment involves 100 trials, put up a range of "out of 100" fractions on the board (for example, $\frac{26}{100}$, $\frac{71}{100}$, $\frac{67}{100}$).
- In pairs and then as a class, decide which simple fraction best represents each of the "out of 100" fractions on the board.

This experiment, known as Buffon's Needle, was first devised in the 18th century by Georges-Louis Leclerc, Comte de Buffon. It has attracted a surprising amount of mathematical interest. There are a number of very good computer-generated simulations available on the Internet (type Buffon's Needle into your browser). Statistically and mathematically, it has been shown that the probability of a "hit" is close to $\frac{2}{3}$.

---

* For bolded terms, see introductory section.

Page 23: The Unit Fraction Game

Achievement Objectives

Number strategies

- Use a range of additive and simple multiplicative strategies with whole numbers, fractions, decimals, and percentages (Number and Algebra, level 3).

Probability

- Investigate simple situations that involve elements of chance by comparing experimental results with expectations from models of all the outcomes, acknowledging that samples vary (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

The Unit Fraction Game can be used to develop these key competencies:

- thinking
- managing self.

Game and Activity

The game and activity on this page can be used to improve your students' understanding of fractions, in particular, their understanding that fractions can function both as numbers and as operators. It can also be used to improve their confidence when it comes to adding, multiplying, and simplifying fractions.

In the game, the students must toss a pair of dice. One dice gives a number, and the other dice gives an operation to perform on that number. **Probability*** concerns the **chance** of a particular outcome occurring in relation to all the possible outcomes (in this case, 36) that might occur in a given situation.

Check that the students:

- understand unit fractions and mixed numbers (this is vital for this exercise);
- recognise equivalent fractions (for example, that $\frac{1}{2} + \frac{1}{2} = \frac{2}{2}$ or 1);
- know that fractions can operate on other numbers multiplicatively (for example, they can find $\frac{1}{4}$ of another number);
- know that, when adding fractions with the same denominator, the denominator stays the same (for example, $\frac{1}{4} + \frac{1}{4} = \frac{2}{4}$ or $\frac{1}{2}$);
- know that, when adding fractions with unrelated denominators (that is, when one denominator is not a multiple of the other), the fractions need to be rewritten as equivalent fractions with a common denominator (for example, $\frac{1}{3} + \frac{1}{4} = \frac{4}{12} + \frac{3}{12}$);
- know their multiplication facts and how to multiply unit fractions with whole numbers.

Like a number of the other activities in this book that involve probability, this activity explores the notions of fairness and **variability**.

* For bolded terms, see introductory section.

Whether the game is fair could be tested by playing it a large number of times. But, as question **2** is designed to clarify, three games is definitely not a large number: three losses from three **trials** means nothing. There is, however, only one certain way of showing whether the game is fair, and that is by listing all the possible **outcomes** and seeing whether an equal number favour each player.

All 6 sides of the two dice are equally likely to appear face-up after being thrown, and any one side of the first dice can appear with any of the 6 sides of the second dice, so there are 6 × 6 possible outcomes. These are set out as a table in the Answers.

Page 24: Dylan's Dominoes

Achievement Objectives

Probability

- Investigate simple situations that involve elements of chance by comparing experimental results with expectations from models of all the outcomes, acknowledging that samples vary (Statistics, level 3).
- 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).

Investigation	Literacy	Probability

P	P	D	A	C

Key Competencies

Dylan's Dominoes can be used to develop these key competencies:

- thinking
- participating and contributing.

Activity

In this activity, the **experimental*** → **theoretical** progression that is typical of **probability** activities is reversed: the students start with a theoretical model, use it to make predictions, and then see how experimental results conform to the predictions. They will find that the experimental results vary widely from group to group, providing an excellent example of **variation** at work.

Once the students have determined that cups are needed for each dot total from 2 to 12, they may initially assume that all the **outcomes** are equally likely. Question **2** is designed to challenge any such assumption and provide a rationale for a prediction.

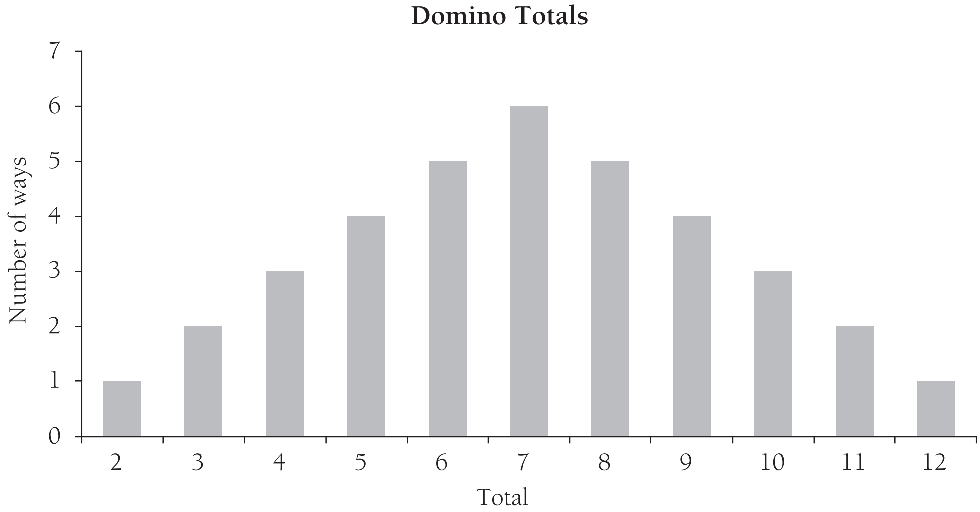
The most convenient way of setting out all possible outcomes for the roll of two dice is in a 6 × 6 table. If your students have done the previous activity (The Unit Fraction Game), they will have encountered a similar table there. If they have not come across this kind of table before, they may need some guidance. The Answers show the completed table.

The table clearly shows, for example, that there are more ways of getting a total of 7 from the roll of two dice than there are of getting a total of 2. Even when students see this, they may fail to realise its significance, mistaking multiple instances of an outcome for repeats (for example, believing that 3 and 4 is the same outcome as 4 and 3). This error is more likely if the dice used

* For bolded terms, see introductory section.

are indistinguishable so, if possible, use dice of different colours: a red 3 and a yellow 4 is likely to be recognised as different from a red 4 and a yellow 3.

It will be helpful to graph the information from the table as a **bar graph**. This will show the pyramid-shaped data distribution that the mathematics would lead us to expect:



For question **3**, a **dot plot** or bar graph are both suitable, but if a number of different groups are doing the activity, creating a sizeable pool of data, the students are likely to find that the best option is to enter the data in a spreadsheet and then create a bar graph. They can then compare the shape of this graph with that of the theoretical model above.

Encourage different groups to share and compare their graphs. The students are likely to find that they vary widely. It is important that they think about why this is and that their thinking contributes to a growing appreciation of what is meant by variation.

In question **4**, the different groups pool their data. When graphed, the result is likely to conform more closely to the expected pyramid shape than did the graphs of the data for the individual groups. It is nevertheless entirely possible (as a result of variation) that an individual group has a better pyramid than the combined groups.

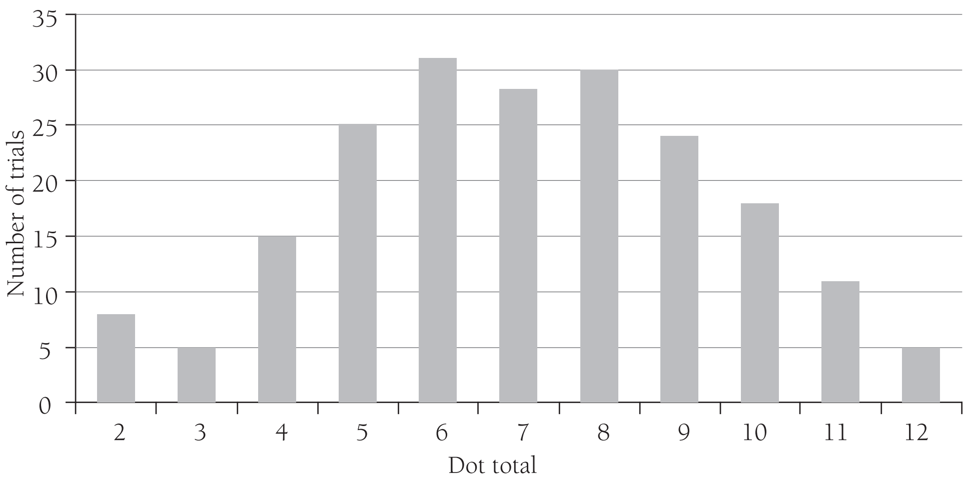
Data from a computer simulation

For your information, the following table contains 10 computer simulated sets of data for this activity, each consisting of 200 **trials**. Each column contains the data for one set of 200 trials. From the table, it can be seen, for example, that the two dice gave a sum of 2 eight times in the first set of trials, four times in the second set, never in the third set, and 61 times in the entire set of 2000 (10 × 200) trials.

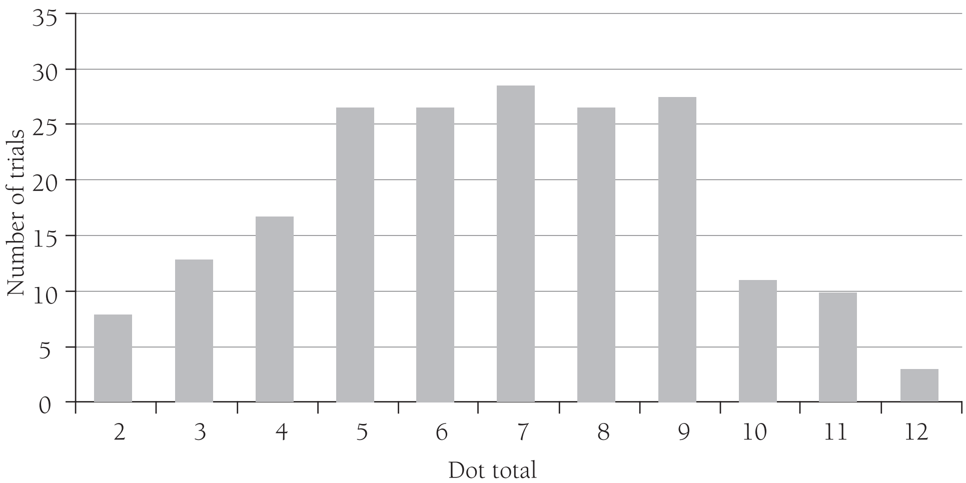
Dice total \ Trial	1	2	3	4	5	6	7	8	9	10	Total (out of 2000 trials)
2	8	4	0	6	8	9	9	8	6	3	61
3	5	14	12	13	13	11	5	10	11	9	103
4	15	19	16	13	17	9	26	21	18	13	167
5	25	19	22	23	27	22	18	21	28	24	229
6	31	26	24	40	27	27	29	20	28	31	283
7	28	36	34	31	29	35	32	34	29	32	320
8	30	23	29	21	27	29	26	32	24	26	267
9	24	19	33	22	28	29	18	23	27	26	249
10	18	23	15	19	11	20	18	16	14	19	173
11	11	12	10	8	10	8	12	11	11	7	100
12	5	5	5	4	3	1	7	4	4	10	48

The following two bar graphs, for the first and fifth sets of 200 trials, show how different the results can be, even with a reasonably large pool of data.

Dylan's Dominoes (first 200 trials)

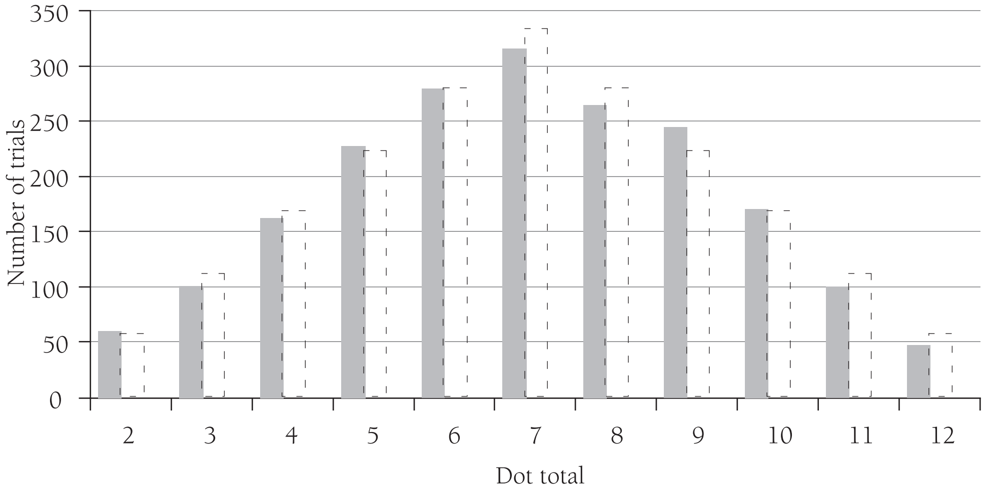


Dylan's Dominoes (fifth 200 trials)



This final bar graph shows the data for all 2000 trials, together with (outlined bars) the **theoretical** expectation for ease of comparison. It can be seen that the data now conforms very closely to the expected pyramid shape.

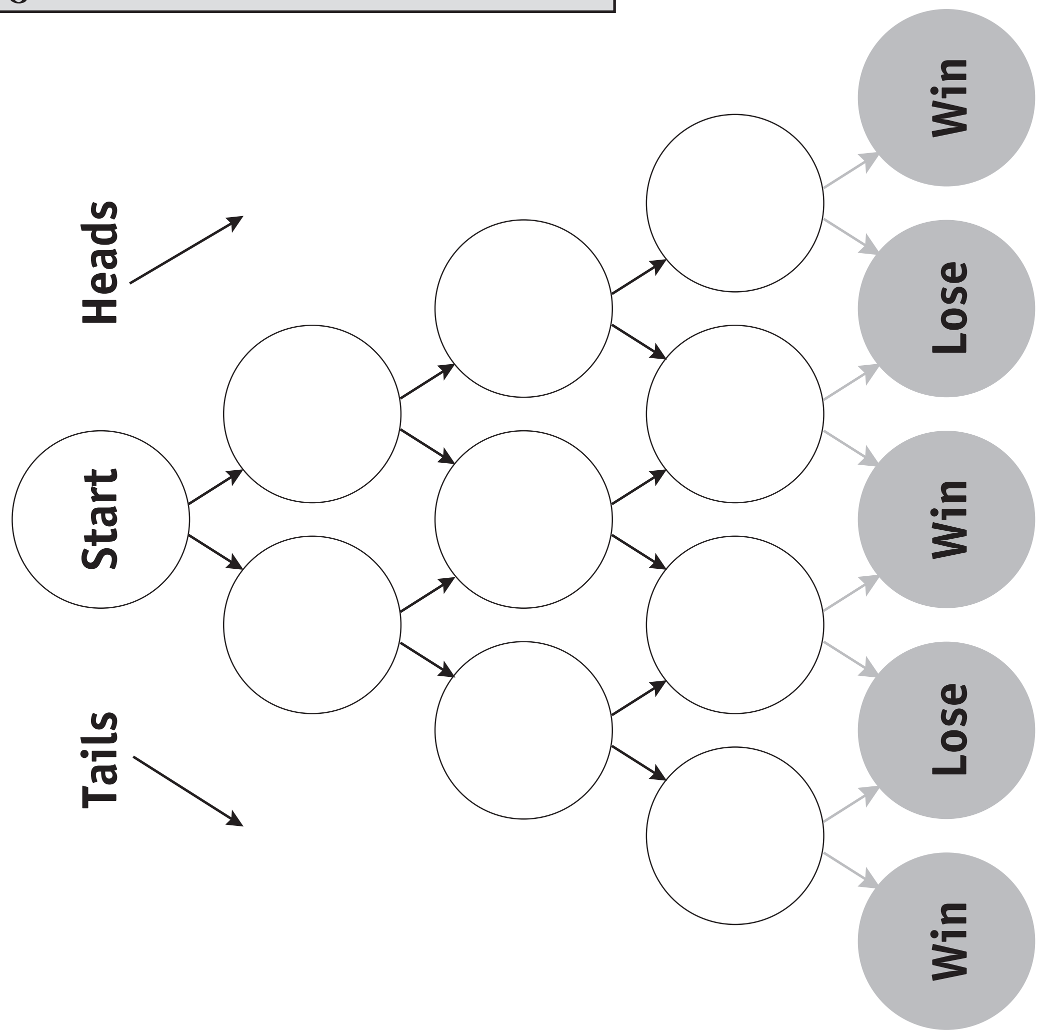
Dylan's Dominoes (2000 trials)



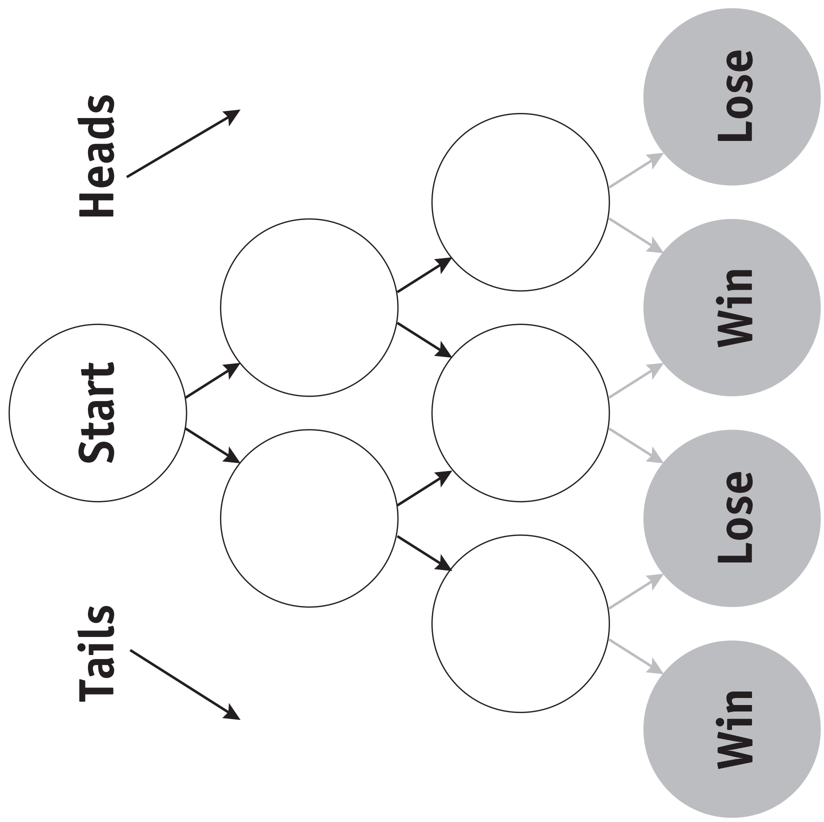
If wished, you could share some or all of this data with your students for their own explorations.

Copymaster: Left to Chance

The 5-Row Game



The Green Game



Acknowledgments

The Ministry of Education and Learning Media would like to thank Sandra Cathcart (The University of Auckland Faculty of Education) and Karen Gibbs (Taradale High School) for their contribution to the development of these teachers' notes.

The illustration on the cover is by Ben Galbraith. All diagrams and figures are by Rachel Street.
All illustrations, except for the poster noted below, are copyright © Crown 2008.
The Data Detective poster on page 15 is copyright © CensusAtSchool New Zealand, www.censusatschool.org.nz

Editorial team: Susan Slaughter, Ian Reid, June Hannah
Designer: Rachel Street

Published 2008 for the Ministry of Education by
Learning Media Limited, Box 3293, Wellington, New Zealand.
www.learningmedia.co.nz

The *Answers and Teachers' Notes* are available online at
http://www.tki.org.nz/r/maths/curriculum/figure/level_3_4_e.php

Dewey number 510
ISBN 978 0 7903 3351 9
Item number 33351
Students' book: item number 33354

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