EXERCISE 11.1 STATISTICAL INVESTIGATIONS

2 (a) This graph has a vertical axis on both sides with a different scale on each because there are two different types of information shown on the graph. One scale is for the rainfall bar graph and the other is for the line (time series) graph for average number of rain days.

(b) Look for the tallest bar. The month with the highest average monthly rainfall is February.

(c) Look for the highest point on the line graph. The month with the most days of rain is February.

(d) False. The average daily rainfall in January must be less than in February because the same number of rainy days produce less rainfall in January than in February.

For January, average daily rainfall is approximately: $95 \div 9 \approx 10.6$ mm

For February, average rainfall is approximately: $110 \div 9 \approx 12.2$ mm

So the average daily rainfall for January is less than for February.

(e) True. August averages about one or two days of rain for the month, so the days of rain in August are rare. Average rainfall is about 2 mm, so the average is about 1 mm each time, so when it rains, very little falls. You could say the rainfall is virtually non-existent.

(f) The production of this graph most likely falls into the analysis phase of the statistical investigation process. Analysing the data involves displaying the data in various ways such as in a graph.

4 D

False. The graph displays the data, results of the investigation, presenting the findings, which is interpreting and communicating the results. You should have already done stages **A**, **B** and **C** before this stage.

6 (a) $\dfrac{221}{8918} \times 100\% = 2.5\%$

(b) Drowning in the bath-tub and in swimming pools (assuming this excludes public swimming pools) would probably include most or all cases of drowning in the home, as there are few if any other ways you can drown at home. This adds up to $22 + 10 = 32$ cases of drowning in the home.

(c) B

Swimming pools: 22

Bath-tubs: 10

Oceans, rivers, lakes, ponds and other water sources (not in swimming pools or in bath-tubs):

$221 - 32 = 189$

 ISBN 978 1 4886 1829 1

(d) Total traffic accident deaths listed $= 66 + 706 + 236 + 39 + 227 = 1274$.

This leaves $1484 - 1274 = 210$ traffic accident deaths not accounted for.

8 **(a)** Answers may vary.

(b) Answers may vary.

(c) Answers may vary.

(d) Answers may vary.

(e) Answers may vary.

(f) Answers may vary, but percentages are unlikely to be the same.

EXERCISE 11.2 TYPES OF DATA

2 **(a)** There is no order to the radio stations, so the data is nominal. This is a nominal variable.

(b) The degree of support for the new jumper design could be a ranking system so there is an order. This data is ordinal.

(c) Starting letters don't have to be ordered, so the data is nominal. This is a nominal variable.

(d) 'Top 10' means the movies are in order, so the data is ordinal. This is an ordinal variable.

(e) Makes of cars do not need to be ordered, so this data is nominal.

(f) Brands of oil do not have an order, so this data is nominal.

4 **(a)** Ordinal

The categories of 'strongly disagree' to 'strongly agree' are ordered, so the categorical data is ordinal. The numbers are not a measure. They are simply used to put the responses in an order, so it is not discrete.

(b) Continuous

The height of the tides is measured, and has an onfinite number of possibilitieds with a given range, so the data is numerical and continuous.

(c) Discrete

The number of siblings can be counted, so the numerical data is discrete.

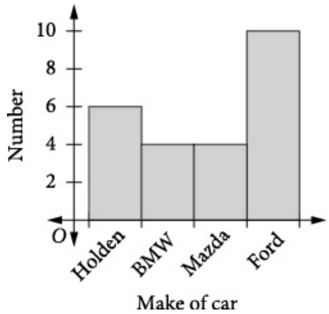
6 Continuous data has an infinite number of possibilities within a particular range, but measuring devices can only be accurate to, for example, hundredths or thousandths of a second. This data can then be counted, and hence appears discrete.

 ISBN 978 1 4886 1829 1

EXERCISE 11.3 DISPLAYING DATA

2 (a) Count the number of each make of car and summarise the results in a table. Then draw bars of suitable heights.

Make of car	Frequency
Holden	6
BMW	4
Mazda	4
Ford	10



(b) You can read from either the graph or the table that Holden has a frequency of 6 and Ford has a frequency of 10.

4 (a) Calculate the percentages of the total complaints (270) as follows.

Punctuality: $\frac{120}{270}\times 100\% \approx 44\%$

Lack of seats: $\frac{80}{270}\times 100\% \approx 30\%$

Cleanliness: $\frac{30}{270}\times 100\% \approx 11\%$

Slow trip: $\frac{20}{270}\times 100\% \approx 7\%$

Cost of trip: $\frac{15}{270}\times 100\% \approx 6\%$

Attitude of the Driver: $\frac{5}{270}\times 100\% \approx 2\%$

Add the cumulative percentages. The result should be 100%.

Cumulative percentages

Punctuality: 44%

Lack of seats: $44\% + 30\% = 74\%$

Cleanliness: $74\% + 11\% = 85\%$

 ISBN 978 1 4886 1829 1

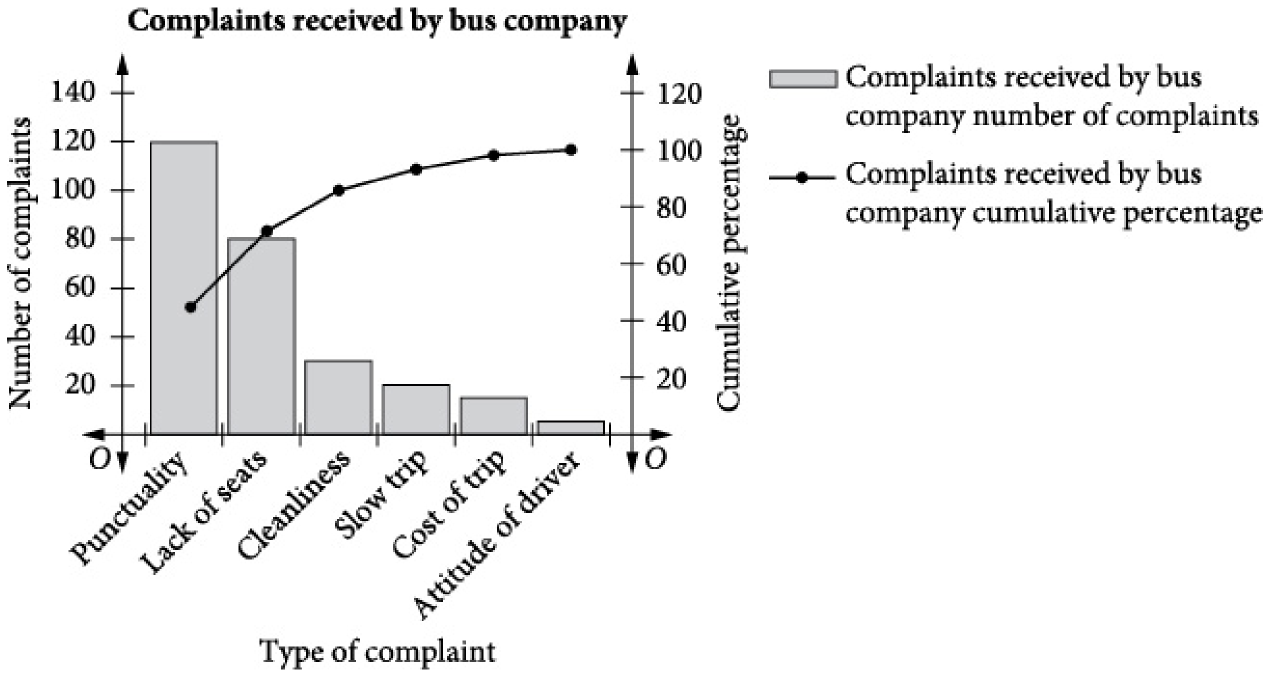
Slow trip: $85\% + 7\% = 92\%$

Cost of trip: $92\% + 6\% = 98\%$

Attitude of the Driver: $98\% + 2\% = 100\%$

Type of Complaint	Number of Complaints	Percentage	Cumulative Percentage
Punctuality	120	44%	44%
Lack of seats	80	30%	74%
Cleanliness	30	11%	85%
Slow trip	20	7%	92%
Cost of trip	15	6%	98%
Attitude of the Driver	5	2%	100%
	270	100%	

(b) Graph the number of complaints as a bar graph and the cumulative percentages as a line graph, using suitable scales for each.



 ISBN 978 1 4886 1829 1

6 **(a)** Look at the total in the 'Apple' column, 35.

(b) Subtract the number having a Samsung (in the total) from total students (100).

$100-22=78$

(c) Write down the fraction of students with an LG phone out of the total number of students.

$\frac{30}{100}=\frac{3}{10}$

This could also be written as $3:10$.

(d) Number of students not Apple, Samsung or LG = number with 'other', 13.

$P(\text{other})=\frac{13}{100}$ or 0.13 or 13%.

8 **(a) (i)** Choose a suitable number to divide 85 by.

$85\div 5=17$ cm

A convenient length for the chart is 17 cm.

(ii) Protein: $6\div 5=1.20$ cm

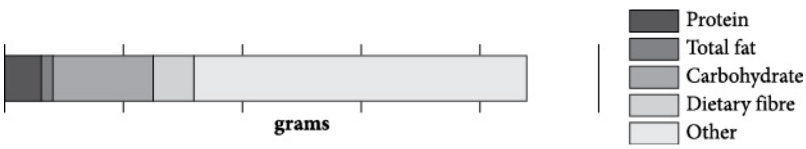
Total fat: $1.9\div 5=0.38$ cm

Carbohydrate: $16.3\div 5=3.26$ cm

Dietary fibre: $6.4\div 5=1.28$ cm

Other: $54.4\div 5=10.88$ cm

Draw the graph.



Note that the vertical black bars each represent 20 cm.

(iii) Find what fraction 68 cm is of 85 cm.

$\frac{68}{85}=\frac{4}{5}$ or 0.8.

Protein: $6\times\frac{4}{5}=4.8$ cm

Carbohydrate: $16.3\times\frac{4}{5}=13.04$ cm

Alternatively, 68 cm is four times 17 cm, so multiply the lengths you calculated earlier by 4.

Protein: $1.20 \text{ cm} \times 4 = 4.8$ cm

Carbohydrate: $3.26 \times 4 = 13.04$ cm

If a 68 cm bar is drawn to represent the total, then protein should take 4.8 cm and carbohydrate should take 13.04 cm.

(b) (i) $125 \div 5 = 25 \text{ cm}$

But best length is between 10 and 20 cm.

$125 \div 10 = 12.5$ cm

A convenient length for the chart is 12.5 cm.

(ii) Protein: $6 \div 10 = 0.60$ cm

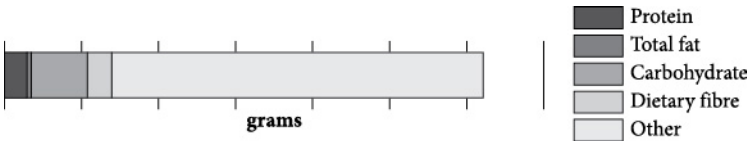
Total fat: $0.9 \div 10 = 0.09$ cm

Carbohydrate: $14.3 \div 10 = 1.43$ cm

Dietary fibre: $6.5 \div 10 = 0.65$ cm

Other: $97.3 \div 10 = 9.73$ cm

Draw the graph.



Note that the vertical black bars each represent 20 cm.

(iii) Protein: $0.6 \text{ cm} \times \dfrac{50}{12.5} = 2.4$ cm

Carbohydrate: $1.43 \text{ cm} \times \dfrac{50}{12.5} = 5.72$ cm

If 50 cm represented the total, then protein should take 2.4 cm and carbohydrate should take 5.72 cm.

(c) (i) A convenient length for the chart is 15 cm, so the measurements in cm will be the same as the components in grams.

(ii) Protein: 0.10 cm

Total fat: 0.10 cm

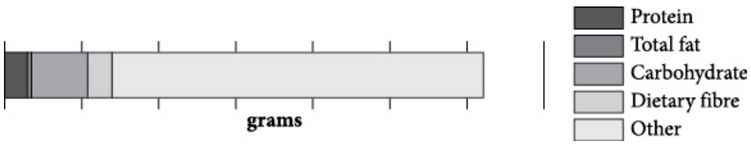
Carbohydrate: 4.80 cm

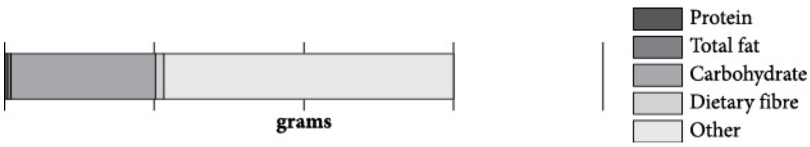
 ISBN 978 1 4886 1829 1

Dietary fibre: 0.30 cm

Other: 9.70 cm

Draw the graph.





Note that the vertical black bars each represent 5 cm.

(iii) Protein: $0.1 \text{ cm} \times \dfrac{60}{15} = 0.4$ cm

Carbohydrate: $4.8 \text{ cm} \times \dfrac{60}{15} = 19.2$ cm

If a 60 cm bar is drawn to represent the total, then protein should take 0.4 cm and carbohydrate should take 19.2 cm.

(d) (i) $50 \div 5 = 10$

A convenient length is 10 cm.

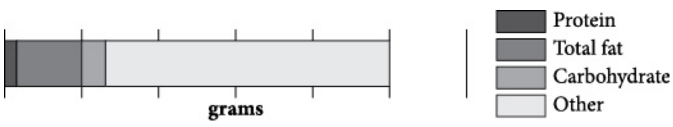
(ii) Protein: $1.6 \div 5 = 0.32$ cm

Total fat: $8.4 \div 5 = 1.68$ cm

Carbohydrate: $3.1 \div 5 = 0.62$ cm

Other: $36.9 \div 5 = 7.38$ cm

Draw the graph.



Note that the vertical black bars each represent 2 cm.

(iii) Protein: $0.32 \text{ cm} \times \dfrac{20}{10} = 0.64$ cm

Carbohydrate: $0.62 \text{ cm} \times \dfrac{20}{10} = 1.24$ cm

If a 20 cm bar is drawn to represent the total, then protein should take 0.64 cm and carbohydrate should take 1.24 cm.

10 B

A histogram is used for numerical continuous (not discrete) data.

A dot plot can be used for discrete data and a bar chart is only used for categorical data.

12 (a) Count how many times each number occurs. 0 occurs 3 times, 1 appears 6 times, 2 appears 7 times, 3 appears 4 times, 4 appears 3 times and 5 appears once. Draw a number line from 0 to 5 and put dots equally spaced apart above each number as follows.

Put 3 dots above 0

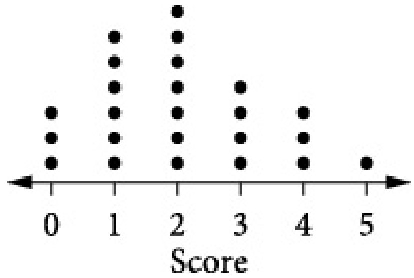
Put 6 dots above 1

Put 7 dots above 2

Put 4 dots above 3

Put 3 dots above 4

Put 1 dot above 5



(b) Count how many times each number occurs. 0 occurs 2 times, 1 appears once, 2 appears 3 times, 3 appears 5 times, 4 appears 4 times, 5 appears 4 times and 6 appears once. Draw a number line from 0 to 6 and put dots equally spaced apart above each number as follows.

2 dots above 0

1 dot above 1

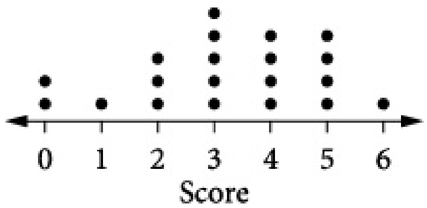
3 dots above 2

5 dots above 3

4 dots above 4

4 dots above 5

1 dot above 6



 ISBN 978 1 4886 1829 1

14 (a) First place the numbers with the tens in the stem and digits in the leaves. An option is to divide the leaves into low $(0\text{-}4)$ and high $(5\text{-}9)$. Initially do not bother about the order. This gives:

Stem	Leaf
1_H	5 8
2_L	2 4 4 2 1
2_H	6 6 5 7
3_L	1 0 0 1
3_H	7 5
4_L	2 0 4 1
4_H	5 6 5

Now rewrite the stem-and-leaf plot, putting the leaves in order. For example, 1_H would be unchanged, but 2_L would change from $2\ 4\ 4\ 2\ 1$ to $1\ 2\ 2\ 4\ 4$. Finally write in the key, in the form tens| units. It does not matter which numbers you use.

Stem	Leaf
1_H	5 8
2_L	1 2 2 4 4
2_H	5 6 6 7
3_L	0 0 1 1
3_H	5 7
4_L	0 1 2 4
4_H	5 5 6

Key: $3|1 = 31$

Note: There is a small enough number of leaves for each stem to just use numbers from 1 to 4 for the stems, giving the following stem-and-leaf plot, which would also be correct. However, it gives a small number of options, so dividing the stems gives more information and a better picture.

 ISBN 978 1 4886 1829 1

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

Key: $3|1 = 31$

(b) First place the numbers with the tens in the stem and digits in the leaves. An option is to divide the leaves into low $(0\text{-}4)$ and high $(5\text{-}9)$. Initially do not bother about the order. This gives:

Stem	Leaf
1_H	7 6 9
2_L	0 2 2
2_H	9 9 7
3_L	2 0 1 3 1 0
3_H	5 6
4_L	1 3 2 1 0
4_H	5 7

 ISBN 978 1 4886 1829 1

Now rewrite the stem-and-leaf plot, putting the leaves in order. For example, 1_H would be unchanged, but 2_L would change from $2\ 4\ 4\ 2\ 1$ to $1\ 2\ 2\ 4\ 4$. Finally write in the key, in the form tens| units. It does not matter which numbers you use.

Stem	Leaf
1_H	6 7 9
2_L	0 2 2
2_H	7 9 9
3_L	0 0 1 1 2 3
3_H	5 6
4_L	0 1 1 2 3
4_H	5 7

Key: $3|1\ =31$

16 (a) Football total is 50 and there are 32 males, so females favouring football becomes $50-32=18$.

Total for swimming is $30+28=58$.

Basketball total is 20 and there are 12 females, so males favouring basketball becomes $20-12=8$.

Total females $=18+28+12=58$.

Grand total $=50+58+20=128$ or $70+58=128$.

There are other ways to work out the missing numbers. For example, you could start by going across the top row, giving males whose favourite sport is basketball as $70-32-30=8$.

 ISBN 978 1 4886 1829 1

Fill in the table.

	Football	Swimming	Basketball	Total
Male	32	30	8	70
Female	18	28	12	58
Total	50	58	20	128

(b) Read the total from the table, under 'Swimming'. 58

(c) There are 8 males who gave basketball, out of a total of 70 males, so $\frac{8}{70} = \frac{4}{35}$

(d) There is a total of 50 students for football, out of a total of 128 students, so $\frac{50}{128} = \frac{25}{64}$

18 (a) Poor housekeeping: $\frac{45}{106} \times 100\% \approx 42\%$

Slow check-in/checkout: $\frac{25}{106} \times 100\% \approx 24\%$

Poor breakfast: $\frac{15}{106} \times 100\% \approx 14\%$

Not value for money: $\frac{9}{106} \times 100\% \approx 8\%$

Poor WiFi: $\frac{6}{106} \times 100\% \approx 6\%$

Television: $\frac{4}{106} \times 100\% \approx 4\%$

Parking: $\frac{2}{106} \times 100\% \approx 2\%$

Add the cumulative percentages. The result should be 100%.

Cumulative percentages

Poor housekeeping: 42%

Slow check-in/checkout: $42\% + 24\% = 66\%$

Poor breakfast: $66\% + 14\% = 80\%$

Not value for money: $80\% + 8\% = 88\%$

 ISBN 978 1 4886 1829 1

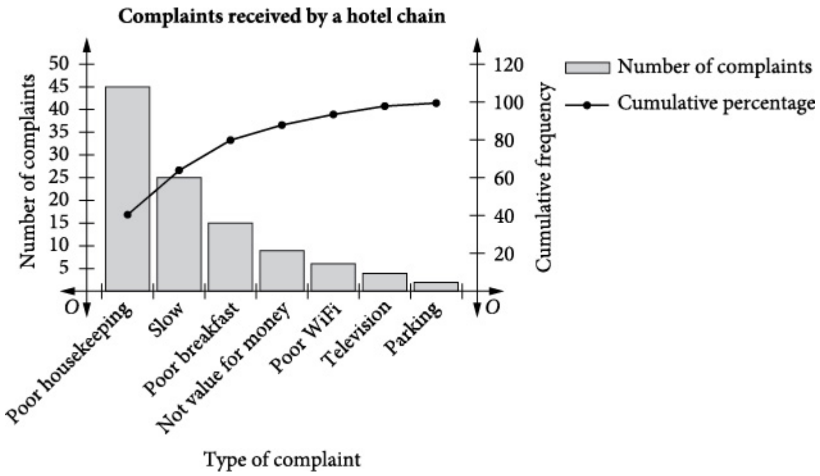
Poor WiFi: $88\% + 6\% = 94\%$

Television: $94\% + 4\% = 98\%$

Parking: $98\% + 2\% = 100\%$

Type of Complaint	Frequency	Percentage	Cumulative Percentage
Poor housekeeping	45	42	42
Slow check-in/checkout	25	24	66
Poor breakfast	15	14	80
Not value for money	9	8	88
Poor WiFi	6	6	94
Television	4	4	98
Parking	2	2	100
	106	100	

(b) Graph the number of complaints as a bar graph and the cumulative percentages as a line graph, using suitable scales for each.



(c) Most of the complaints were the first three areas, so the first priority should be to improve their housekeeping (nearly half of all complaints). After that, they should look at modifying their check-in/checkout procedures and consider ways to improve the quality of the breakfast.

 ISBN 978 1 4886 1829 1

EXERCISE 11.4 MEASURES OF CENTRAL TENDENCY

2 **(a)** Ordered data: 1, 1, 2, 3, 3, 4, 4, 4, 6, 8, 9

There are 11 scores, the median is the 6^{th} score. The 6^{th} score is 4.

The median is 4

(b) Ordered data: 2, 2, 2, 2, 3, 3, 5, 6, 6, 6, 9, 9

There are 12 scores; the median is the average of the 6^{th} and 7^{th} score. The 6^{th} score is 3 and the 7^{th} score is 5.

The median is $\frac{3+5}{2} = 4$

(c)

Score (x)	Frequency (f)	Cumulative Frequency
0	3	3
1	6	9
2	7	16
3	4	20
4	3	23
5	1	24
	$\sum f = 24$	

There are 24 scores, the median is the average of the 12^{th} and 13^{th} score. The 12^{th} score is 2 and the 13^{th} score is 2.

The median is $\frac{2+2}{2} = 2$

4 **(a)** Data in order: 1, 2, 2, 3, 4, 4, 4, 6, 7, 8, 9

The most frequently occurring score is 4.

The mode is 4

(b) Data in order: 1, 2, 3, 3, 4, 5, 5, 6, 7, 8, 9

The most frequently occurring scores are 3 and 5.

The two modes are 3 and 5

(c) Data in order: 1, 1, 1, 2, 2, 2, 3, 4, 4, 4, 5

 ISBN 978 1 4886 1829 1

As there are more than two numbers with the highest frequency, no mode is given.

There is no mode or it is multimodal

6 (a) The data is already in order. There are 11 scores. The median of data is the 6th score. $Q_2 = 11$. This is excluded from the quartile calculation.

Lower half of data: 2, 3, 5, 8, 9

The median of the lower half is the 3rd score. $Q_1 = 5$

Upper half of the data: 15, 16, 18, 25, 36

The median of the upper half is the 3rd score. $Q_3 = 18$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 18 - 5 \\ &= 13 \end{aligned}$$

(b) Data in order: 1, 2, 4, 5, 6, 7, 9, 9, 11, 12, 16, 16, 17, 17, 18, 21

There are 16 scores. The median of the data is the average of the 8th and the 9th scores. The data is divided after the 8th score, and both these scores are used in the quartile calculation.

$$Q_2 = \frac{9+11}{2} = 10$$

Lower half of data: 1, 2, 4, 5, 6, 7, 9, 9

The median of the lower half is the average of the 4th and the 5th score.

$$Q_1 = \frac{5+6}{2} = 5.5$$

Upper half of the data: 11, 12, 16, 16, 17, 17, 18, 21

The median of the upper half is the average of the 4th and the 5th score. Both of these scores are used in the quartile calculation.

$$Q_3 = \frac{16+17}{2} = 16.5$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 16.5 - 5.5 \\ &= 11 \end{aligned}$$

(c) Data in order: 1, 1, 2, 3, 7, 7, 9, 11, 11, 12, 14, 16, 17, 22

There are 14 scores. The median of data is the average of the 7th score and the 8th score.

$$Q_2 = \frac{9+11}{2} = 10$$

Lower half of data: 1, 1, 2, 3, 7, 7, 9

The median of the lower half is the 4th score. $Q_1 = 3$

Upper half of the data: 11, 11, 12, 14, 16, 17, 22

The median of the upper half is the 4th score. $Q_3 = 14$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 14 - 3 \\ &= 11 \end{aligned}$$

8 (a)

Score	Frequency (f)	Cumulative Frequency
0	3	3
1	6	9
2	7	16
3	4	20
4	3	23
5	1	24
	$\sum f = 24$	

There are 24 scores. The median of data is the average of the 12th score and the 13th score, both of which are 2. $Q_2 = 2$.

The lower quartile is the median of the lower 12 scores. The median of these 12 scores is the average of the 6th and 7th score, both of which are 1. $Q_1 = 1$

The upper quartile is the median of the upper 12 scores. The median of these 12 scores is the average of the 6th and 7th score, or between the 18th and 19th score for the whole dataset. These are both 3. $Q_3 = 3$

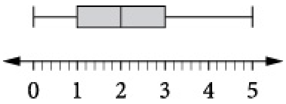
The minimum is 0 and the maximum is 5.

Five-figure summary: 0, 1, 2, 3, 5

 ISBN 978 1 4886 1829 1

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 3 - 1 \\ &= 2 \end{aligned}$$

The range is $5 - 0 = 5$.



(b)

Score	Frequency (f)	Cumulative Frequency
4	15	15
5	23	38
6	14	52
7	23	75
8	17	92
9	19	111
10	20	131
	$\sum f = 131$	

There are 131 scores. The median of data is the 66th score. $Q_2 = 7$

The lower quartile is the median of the lower 65 scores. The median of these 65 scores is the 33rd score. $Q_1 = 5$

The upper quartile is the median of the upper 65 scores. The median of these 65 scores is the 33rd score or the 99th score for the whole dataset. $Q_3 = 9$

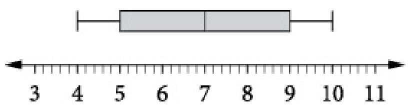
The minimum is 4 and the maximum is 10.

Five-figure summary: 4, 5, 7, 9, 10

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 9 - 5 \\ &= 4 \end{aligned}$$

The range is $10 - 4 = 6$

 ISBN 978 1 4886 1829 1



(c)

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

There are 26 scores. The median of the data is the average of the 13th and 14th score.

$$Q_2 = \frac{30+31}{2} = 30.5$$

The lower quartile is the median of the lower 13 scores. The median of these 13 scores is the 7th score. $Q_1 = 15$

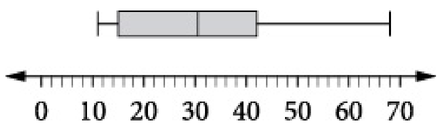
The upper quartile is the median of the upper 13 scores. The median of these 13 scores is the 7th score or the 20th score for the whole dataset. $Q_3 = 42$

The minimum is 11 and the maximum is 68.

Five-figure summary: 11, 15, 30.5, 42, 68

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 42 - 15 \\ &= 27 \end{aligned}$$

The range is $68 - 11 = 57$



(d)

Stem	Leaf
15	2 5 8
16	0 1 3 4 5
17	1 8 9 9
18	9 9
19	8 9
20	0 2

There are 18 scores. The median of the data is the average of the 9th and 10th scores.

$$Q_2 = \frac{171+178}{2} = 174.5$$

The lower quartile is the median of the lower 9 scores. The median of these 9 scores is the 5th score. $Q_1 = 161$

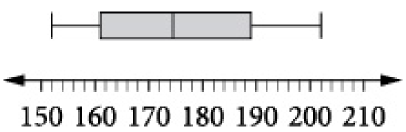
Upper quartile is the median of the upper 9 scores. The median of these 9 scores is the 5th score or the 14th score for the whole dataset. $Q_3 = 189$

The minimum is 152 and the maximum is 202.

Five-figure summary: 152, 161, 174.5, 189, 202

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 189 - 161 \\ &= 28 \end{aligned}$$

The range is $202 - 152 = 50$.



10 B

x	f	fx	Cumulative Frequency
1	5	5	5
2	8	16	13
3	3	9	16
4	2	8	18
5	4	20	22
6	5	30	27
	$\sum f = 27$	$\sum fx = 88$	

Mean: $\bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{88}{27} \approx 3.26$

There are 27 scores, so the median is the 14th score. The 14th score is 3.

The median is 3.

12 The numbers are already in order, in all parts of this question.

(a) The minimum is 2.

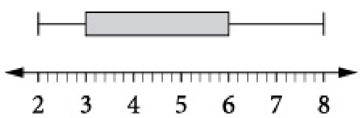
There are 7 scores, so the median is the 4th score. $Q_2 = 6$.

The first quartile is the middle of the first three scores. $Q_1 = 3$

The third quartile is the middle of the last three scores. $Q_3 = 6$

The maximum is 8.

The median is equal to the upper quartile.



(b) The minimum is 2.

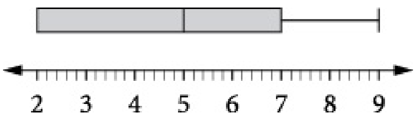
There are 7 scores, so the median is the 4th score. $Q_2 = 5$

The first quartile is the middle of the first three scores. $Q_1 = 2$

The third quartile is the middle of the last three scores. $Q_3 = 7$

The maximum is 9.

The minimum is equal to the lower quartile.



(c) The minimum is 2.

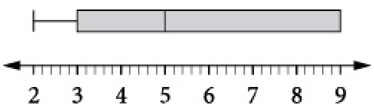
There are 7 scores, so the median is the 4th score. $Q_2 = 5$

The first quartile is the middle of the first three scores. $Q_1 = 3$

The third quartile is the middle of the last three scores. $Q_3 = 9$

The maximum is 9.

The upper quartile is equal to the maximum.



(d) The minimum is 2.

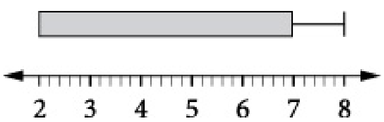
There are 7 scores, so the median is the 4th score. $Q_2 = 2$

The first quartile is the middle of the first three scores. $Q_1 = 2$

The third quartile is the middle of the last three scores. $Q_3 = 7$

The maximum is 8.

The minimum is equal to the lower quartile and the median.



14 The minimum must be the same as the first quartile, on the left end of the box plot.

The minimum is 2.

The maximum is 8 (the right-hand end).17

The range is $8 - 2 = 6$

16 (a) Data in order: $16,\ 16,\ 17,\ 18,\ 19,\ 19,\ 21,\ 22,\ 22,\ 23,\ 24,\ 25,\ 25,\ 25,\ 28,\ 45,\ 73$

There are scores. The median of data is the 9th score.

$Q_2 = 22$

Lower half of data (excluding Q_2): $16,\ 16,\ 17,\ 18,\ 19,\ 19,\ 21,\ 22$

The median of the lower half is the average of the 4th score and the 5th score.

$$Q_1 = \frac{18+19}{2} = 18.5$$

Upper half of the data (excluding Q_2): $23,\ 24,\ 25,\ 25,\ 25,\ 28,\ 45,\ 73$

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{25+25}{2} = 25$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 25 - 18.5 \\ &= 6.5 \end{aligned}$$

Test for outliers.

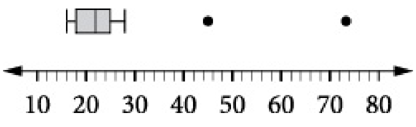
$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 18.5 - 1.5 \times 6.5 \\ &= 8.75 \end{aligned}$$

All scores lie above the lower fence.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 18.5 + 1.5 \times 6.5 \\ &= 28.25 \end{aligned}$$

The scores 45 and 73 are both well above the upper fence and are therefore outliers.

The minimum is 16 and the maximum excluding outliers is 28.



 ISBN 978 1 4886 1829 1

(b) Data in order:

$67,\ 97,\ 101,\ 108,\ 113,\ 113,\ 116,\ 118,\ 119,\ 119,\ 121,\ 121,\ 123,\ 123,\ 132,\ 134,\ 135,\ 141$

There are 18 scores. The median of data is the average of the 9th score and the 10th score.

$$Q_2 = \frac{119+119}{2} = 119$$

Lower half of the data: $67,\ 97,\ 101,\ 108,\ 113,\ 113,\ 116,\ 118,\ 119$

The median of the lower half is the 5th score.

$$Q_1 = 113$$

Upper half of the data (excluding Q_2): $119,\ 121,\ 121,\ 123,\ 123,\ 132,\ 134,\ 135,\ 141$

The median of the upper half is the 5th score in the upper half.

$$Q_3 = 123$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 123 - 113 \\ &= 10 \end{aligned}$$

Test for outliers.

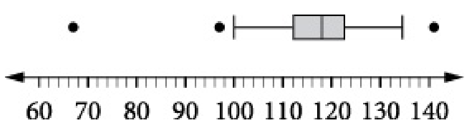
$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 113 - 1.5 \times 10 \\ &= 98 \end{aligned}$$

The scores 67 and 97 are both below the lower fence and are therefore outliers.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 123 + 1.5 \times 10 \\ &= 138 \end{aligned}$$

141 is above the upper fence and is therefore an outlier.

Excluding outliers, the minimum is 101 and the maximum is 135.



 ISBN 978 1 4886 1829 1

(c) Data in order:

18, 32, 37, 38, 38, 39, 40, 41, 41, 41, 42, 42, 42, 42, 42, 43, 43, 44, 45, 45, 45, 46, 47, 72

There are 24 scores. The median of data is the average of the 12th score and the 13th score.

$Q_2 = 42$

Lower half of data: 18, 32, 37, 38, 38, 39, 40, 41, 41, 41, 42, 42

The median of the lower half is the average of the 6th score and the 7th score.

$$Q_1 = \frac{39+40}{2} = 39.5$$

Upper half of the data: 42, 42, 42, 43, 43, 44, 45, 45, 45, 46, 47, 72

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{44+45}{2} = 44.5$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 44.5 - 39.5 \\ &= 5 \end{aligned}$$

Test for outliers.

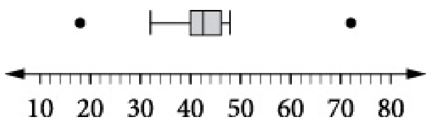
$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 39.5 - 1.5 \times 5 \\ &= 32 \end{aligned}$$

18 is below the lower fence and is therefore an outlier.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 44.5 + 1.5 \times 5 \\ &= 52 \end{aligned}$$

72 is above the upper fence and is therefore an outlier.

Excluding outliers, the minimum is 32 and the maximum is 47.



(d) Data in order:

72, 76, 112, 113, 123, 123, 127, 129, 129, 130, 132, 132, 134, 136, 137, 137, 139, 140, 141, 142, 143

There are 21 scores. The median of data is the 11th score.

$Q_2 = 132$

Lower half of data (excluding Q_2): 72, 76, 112, 113, 123, 123, 127, 129, 129, 130

The median of the lower half is the average of the 5th score and the 6th score.

$Q_1 = 123$

Upper half of the data (excluding Q_2): 132, 134, 136, 137, 137, 139, 140, 141, 142, 143

The median of the upper half is the average of the 5th score and the 6th score in the upper half.

$$Q_3 = \frac{137+139}{2} = 138$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 138 - 123 \\ &= 15 \end{aligned}$$

Test for outliers.

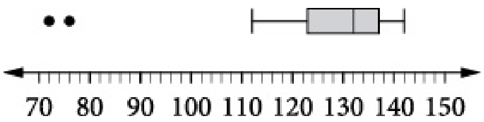
$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 123 - 1.5 \times 15 \\ &= 100.5 \end{aligned}$$

The scores 72 and 76 both lie below the lower fence and are outliers.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 138 + 1.5 \times 15 \\ &= 160.5 \end{aligned}$$

All scores are below the upper fence so there are no more outliers.

The minimum, excluding the two outliers, is 112 and the maximum is 143.



18 (a) There are 21 temperatures in the data.

$$\bar{x} = \frac{3+18+9+3+1+34+13+0-4+18+5+14+6+1+13-3+0+33+27+10-1}{21}$$

$$= \frac{200}{21}$$

$$\approx 9.52$$

(b) Data in order: $-4,\ -3,\ -1,\ 0,\ 0,\ 1,\ 1,\ 3,\ 3,\ 5,\ 6,\ 9,\ 10,\ 13,\ 13,\ 14,\ 18,\ 18,\ 27,\ 33,\ 34$

The median is the 11^{th} temperature $= 6°C$

(c) Listing the temperatures in order in part (b) is useful here.

Temperature	Frequency
$-5-<0$	3
$0-<5$	6
$5-<10$	3
$10-<15$	4
$15-<20$	2
$20-<25$	0
$25-<30$	1
$30-<35$	2
	$\sum f = 21$

 ISBN 978 1 4886 1829 1

(d)

Temperature	Frequency	Mid-point x_m	fx_m
−5–<0	3	−2.5	−7.5
0–<5	6	2.5	15
5–<10	3	7.5	22.5
10–<15	4	12.5	50
15–<20	2	17.5	35
20–<25	0	22.5	0
25–<30	1	27.5	27.5
30–<35	2	32.5	65
	$\sum f = 21$		$\sum f x_m = 207.5$

$$\begin{aligned}\bar{x} &= \frac{\sum fx_m}{\sum f} \\ &= \frac{207.5}{21} \\ &\approx 9.88°C\end{aligned}$$

(e)

Temperature range	Frequency	Cumulative Frequency
$-5\text{–}<0$	3	3
$0\text{–}<5$	6	9
$5\text{–}<10$	3	12
$10\text{–}<15$	4	16
$15\text{–}<20$	2	18
$20\text{–}<25$	0	18
$25\text{–}<30$	1	19
$30\text{–}<35$	2	21
	$\sum f = 21$	

$n = 21$, so the median is the 11^{th} value. The median class interval is $5\text{–}<10$.

 ISBN 978 1 4886 1829 1

EXERCISE 11.5 STANDARD DEVIATION

2 (a)

Length of fish	x_m	Number of fish
20 – 29	24.5	8
30 – 39	34.5	12
40 – 49	44.5	28
50 – 59	54.5	38
60 – 69	64.5	54
70 – 79	74.5	49
80 – 89	84.5	21
90 – 99	94.5	7

$$\begin{aligned}\bar{x} &= 62.195... \\ &\approx 62.20 \text{ cm}\end{aligned}$$

(b) It is a sample, so use sample standard deviation.

$$\begin{aligned}s_x &= 16.251... \\ &\approx 16.25\end{aligned}$$

4 C

Mean will decrease as score is added at lower end of data values. Since the score is almost equal to the lowest score, it must be further away from the mean than the standard deviation, and so the standard deviation would increase.

6 The following working shows how a table could be used to calculate the standard deviations. If the spreadsheet method was used, the table would be identical. Alternatively, the data may be entered into a scientific or graphic calculator.

In all cases, population standard deviation uses n, rather than $n-1$.

The tables contain precise data, obtained from a spreadsheet, so if you have rounded the mean, your answers along the way may vary very slightly from these, but the final answer should be the same.

 ISBN 978 1 4886 1829 1

(a)

Scores (x)	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
1	2	2	9.765625	19.53125
2	2	4	4.515625	9.03125
3	4	12	1.265625	5.0625
4	2	8	0.015625	0.03125
5	1	5	0.765625	0.765625
6	2	12	3.515625	7.03125
7	2	14	8.265625	16.53125
8	0	0	15.015625	0
9	1	9	23.765625	23.765625
Totals:	16	66		81.75

$\bar{x} = \dfrac{66}{16} = 4.125 \qquad Var(X) = \dfrac{81.75}{16} = 5.109\ 375$

$\sigma_x = \sqrt{5.109\ 375} = 2.260... \approx 2.3\ \text{(1 d.p.)}$

(b)

Scores (x)	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
21	1	21	19.22485207	19.22485207
22	2	44	11.4556213	22.9112426
23	0	0	5.686390533	0
24	1	24	1.917159763	1.917159763
25	3	75	0.147928994	0.443786982
26	2	52	0.378698225	0.75739645
27	1	27	2.609467456	2.609467456
28	1	28	6.840236686	6.840236686
29	1	29	13.07100592	13.07100592
30	1	30	21.30177515	21.30177515
Totals:	13	330		89.07692308

$\bar{x} = \dfrac{330}{13} = 25.384\ 615... \approx 25.3846$ Use a much higher accuracy than the question requires.

$Var(X) = \dfrac{89.0769...}{13} = 6.852...$

$\sigma_x = \sqrt{6.852...} = 2.617... \approx 2.6\ \text{(1 d.p.)}$

 ISBN 978 1 4886 1829 1

(c)

Scores (x)	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
2	1	2	79.87890625	79.87890625
3	0	0	63.00390625	0
4	1	4	48.12890625	48.12890625
5	1	5	35.25390625	35.25390625
6	0	0	24.37890625	0
7	2	14	15.50390625	31.0078125
8	1	8	8.62890625	8.62890625
9	0	0	3.75390625	0
10	0	0	0.87890625	0
11	2	22	0.00390625	0.0078125
12	1	12	1.12890625	1.12890625
13	2	26	4.25390625	8.5078125
14	0	0	9.37890625	0
15	1	15	16.50390625	16.50390625
16	2	32	25.62890625	51.2578125
17	1	17	36.75390625	36.75390625
18	1	18	49.87890625	49.87890625
Totals:	16	175		366.9375

$$\bar{x} = \frac{175}{16} = 10.9375 \qquad Var(X) = \frac{366.9375}{16} = 22.933...$$

$$\sigma_x = \sqrt{22.933...} = 4.788... \approx 4.8 \text{ (1 d.p.)}$$

(d)

Scores (x)	Frequency (f)	Cumulative frequency	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
13	1	1	13	56.25	56.25
14	0	1	0	42.25	0
15	1	2	15	30.25	30.25
16	0	2	0	20.25	0
17	2	4	34	12.25	24.5
18	1	5	18	6.25	6.25
19	1	6	19	2.25	2.25
20	1	7	20	0.25	0.25
21	2	9	42	0.25	0.5
22	1	10	22	2.25	2.25
23	1	11	23	6.25	6.25
24	0	11	0	12.25	0
25	0	11	0	20.25	0
26	1	12	26	30.25	30.25
27	1	13	27	42.25	42.25
28	1	14	28	56.25	56.25
Totals:	14		287		257.5

$$\bar{x} = \frac{287}{14} = 20.5$$

 ISBN 978 1 4886 1829 1

$$Var(X) = \frac{257.5}{14} = 18.392...$$

$$\sigma_x = \sqrt{18.392...} = 4.288... \approx 4.3 \text{ (1 d.p.)}$$

(e)

Scores (x)	Frequency (f)	fx	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
15	1	15	61.56213018	61.56213018
16	1	16	46.86982249	46.86982249
17	1	17	34.17751479	34.17751479
18	0	0	23.4852071	0
19	1	19	14.79289941	14.79289941
20	1	20	8.100591716	8.100591716
21	2	42	3.408284024	6.816568047
22	0	0	0.715976331	0
23	0	0	0.023668639	0
24	0	0	1.331360947	0
25	1	25	4.639053254	4.639053254
26	1	26	9.946745562	9.946745562
27	1	27	17.25443787	17.25443787
28	1	28	26.56213018	26.56213018
29	1	29	37.86982249	37.86982249
30	0	0	51.17751479	0
31	0	0	66.4852071	0
32	0	0	83.79289941	0
33	1	33	103.1005917	103.1005917
Totals:	13	297		371.6923077

$$\bar{x} = \frac{297}{13} = 22.846\,153... \approx 22.8462$$

Use a much higher accuracy than the question requires.

$$Var(X) = \frac{371.6923...}{13} = 28.591...$$

$$\sigma_x = \sqrt{28.591...} = 5.347... \approx 5.3 \text{ (1 d.p.)}$$

 ISBN 978 1 4886 1829 1

(f)

Scores (x)	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
52	1	52	68.33777778	68.33777778
53	1	53	52.80444444	52.80444444
54	1	54	39.27111111	39.27111111
55	0	0	27.73777778	0
56	1	56	18.20444444	18.20444444
57	1	57	10.67111111	10.67111111
58	0	0	5.137777778	0
59	1	59	1.604444444	1.604444444
60	2	120	0.071111111	0.142222222
61	2	122	0.537777778	1.075555556
62	0	0	3.004444444	0
63	0	0	7.471111111	0
64	0	0	13.93777778	0
65	2	130	22.40444444	44.80888889
66	1	66	32.87111111	32.87111111
67	1	67	45.33777778	45.33777778
68	1	68	59.80444444	59.80444444
Totals:	15	904		374.9333333

$$\bar{x} = \frac{904}{15} = 60.266\ 666... \approx 60.2667$$

Use a much higher accuracy than the question requires.

$$Var(X) = \frac{374.9333...}{15} = 24.995...$$

$$\sigma_x = \sqrt{24.995...} = 4.9995... \approx 5.0 \text{ (1 d.p.)}$$

(g)

Score	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
44	8	352	5.121883657	40.97506925
45	12	540	1.595567867	19.1468144
46	13	598	0.069252078	0.900277008
47	10	470	0.542936288	5.429362881
48	9	432	3.016620499	27.14958449
49	5	245	7.490304709	37.45152355
Totals:	57	2637		131.0526316

$$\bar{x} = \frac{2637}{57} = 46.263\ 157... \approx 46.2632$$

Use a much higher accuracy than the question requires.

$$Var(X) = \frac{131.0526...}{57} = 2.299...$$

 ISBN 978 1 4886 1829 1

$\sigma_x = \sqrt{2.299...} = 1.516... \approx 1.5$ (1 d.p.)

(h)

Score	Frequency (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
101	11	1111	6.336504162	69.70154578
102	14	1428	2.302021403	32.22829964
103	18	1854	0.267538644	4.8156956
104	19	1976	0.233055886	4.428061831
105	13	1365	2.198573127	28.58145065
106	12	1272	6.164090369	73.96908442
Totals:	87	9006		213.7241379

$\bar{x} = \dfrac{9006}{87} = 103.517\ 241... \approx 103.5172$ Use a much higher accuracy than the question requires.

$$Var(X) = \frac{213.724...}{87} = 2.456...$$

$\sigma_x = \sqrt{2.456...} = 1.567... \approx 1.6$ (1 d.p.)

(i)

Score	x_m	Frequency	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10 − 14	12	6	72	105.478 451	632.870 709
15 − 19	17	7	119	27.775 749	194.430 241
20 − 24	22	9	198	0.073 046	0.657 414
25 − 29	27	10	270	22.370 343	223.703 433
30 − 34	32	4	128	94.667 641	378.670 563
35 − 39	37	1	37	216.964 938	216.964 938
Totals		37	824		1647.297 298

$$\bar{x} = \frac{824}{37} = 22.270\ 270... \approx 22.2703$$

Use a much higher accuracy than the question requires.

$$Var(X) = \frac{1647.297...}{37} = 44.521...$$

$\sigma_x = \sqrt{44.521...} = 6.672... \approx 6.7$ (1 d.p.)

(j)

Score	x_m	Frequency	fx	$(x-\overline{x})^2$	$f(x-\overline{x})^2$
$80 -< 90$	85	16	1360	539.271 605	8628.345 679
$90 -< 100$	95	15	1425	174.827 161	2622.407 407
$100 -< 110$	105	17	1785	10.382 716	176.506 173
$110 -< 120$	115	19	2185	45.938 271	872.827 161
$120 -< 130$	125	12	1500	281.493 827	3377.925 926
$130 -< 140$	135	11	1485	717.049 383	7887.543 210
Totals		90	9740		23 565.555 556

$\overline{x} = \dfrac{9740}{90} = 108.222\ 222... \approx 108.2222$

Use a much higher accuracy than the question requires.

$Var(X) \approx \dfrac{23\ 565.555\ 556}{90} = 261.839...$

$\sigma_x = \sqrt{261.839...} = 16.181... \approx 16.2$ (1 d.p.)

8 **(a)** There are 15 scores, so the median will be the eighth score.

$Q_2 = 22$

The first quartile will be the middle of the first 7 scores, i.e. the fourth score.

$Q_1 = 17$

The third quartile will be the middle of the last 7 scores, or the fourth from the end.

$Q_3 = 26$

$$\begin{aligned} IQR &= 26 - 17 \\ &= 9 \end{aligned}$$

$$\begin{aligned} \sigma_x &= 5.29... \\ &\approx 5.3 \end{aligned}$$

(b) There are 15 scores, so the median will be the eighth score.

 ISBN 978 1 4886 1829 1

$Q_2 = 22$

The first quartile will be the middle of the first 7 scores, i.e. the fourth score.

$Q_1 = 17$

The third quartile will be the middle of the last 7 scores, or the fourth from the end.

$Q_3 = 26$

$$\begin{aligned} IQR &= 26 - 17 \\ &= 9 \end{aligned}$$

$$\begin{aligned} \sigma_x &= 6.69... \\ &\approx 6.7 \end{aligned}$$

(c) There are 15 scores, so the median will be the eighth score.

$Q_2 = 22$

The first quartile will be the middle of the first 7 scores, i.e. the fourth score.

$Q_1 = 17$

The third quartile will be the middle of the last 7 scores, or the fourth from the end.

$Q_3 = 26$

$$\begin{aligned} Q_1 &= 17 \\ Q_3 &= 26 \end{aligned}$$

$$\begin{aligned} IQR &= 26 - 17 \\ &= 9 \end{aligned}$$

$$\begin{aligned} \sigma_x &= 8.08... \\ &\approx 8.1 \end{aligned}$$

(d) Answers may vary. Sample answer points may include:

Order of data is unaltered.

Quartiles are unchanged.

IQR is constant.

The standard deviation is changed more greatly by values which are farthest from the mean, the standard deviation of data sets 2 and 3 are greater than data set 1, despite a constant IQR.

 ISBN 978 1 4886 1829 1

EXERCISE 11.6 ANALYSIS OF DATA

2 C

‘About 50% of the paddocks for treatment B had a yield less than 2.2 tonnes/hectare.’ is incorrect.

The interquartile range is the length of the ‘box’, and treatment A has the shortest box, so option A is correct.

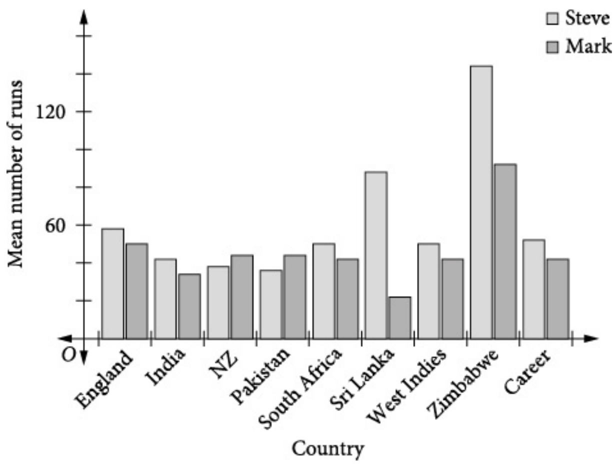
The treatment with the highest yield is the largest maximum, which is in treatment B, so option B is correct.

For treatment B, 2.2 tonnes is almost up to the third quartile, so nearly 75% of paddocks for treatment B yielded less than 2.2 tonnes. Option B is incorrect.

‘About 25% of the paddocks for treatment B had a yield between 1.875 and 1.925 tonnes/hectare’ is correct as 1.875 is the first quartile and 1.925 is the median, comprising one quarter, or 75% of the results.

‘Treatment A produced the most consistent results.’ is correct as its box, and the whole box plot, is the shortest.

4 The heights of the bars correspond to the batting averages.

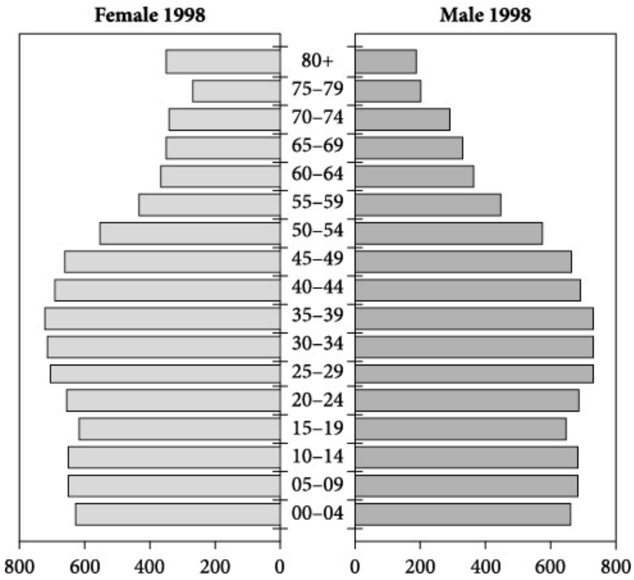


Sample answer: From the graph it can be seen that the batting averages for the brothers was similar for several countries. However, Steve had much better averages against Sri Lanka and Zimbabwe (87.63 and 145 respectively) compared with Mark’s (24.64 and 90). This helped Steve maintain a higher overall career average. Steve’s batting averages are also higher than Mark’s in all but two countries,both of which were close, and a lot higher in two cases.

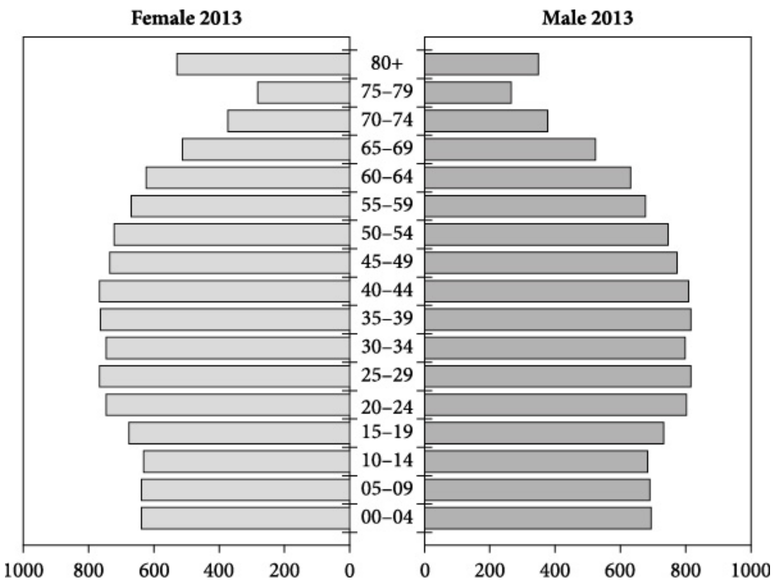
6 Sample answer: The average rainfall per month is roughly constant, as the accumulated rainfall increases by about the same amount each month. The graph shows the rainfall for this particular year was well below average, nearly half the annual rainfall. In this particular year, some months (e.g. May, September) had roughly the average rainfall, while other months (e.g. February, April and December) had little or none, and the rest below average.

 ISBN 978 1 4886 1829 1

8 **(a)** Draw back-to-back horizontal bar graphs, using a suitable scale.



(b) Draw back-to-back horizontal bar graphs, using a suitable scale.



(c) Sample answer: From the graphs, the male and female populations for 1998 are very similar, except for the age groups over 70, where there are a greater number of females. A similar trend is seen in the 2013 data.

There are differences between the two years.

From 0 to 20 years the distribution of populations was similar.

From 20 to 50 years the population is greater for 2013 (approximately 800 000) than for 1998 (approximately 700 000).

For 50 to 70 years the population decreases, but is still greater for 2013 (from 750 000 to 500 000) than for 1998 (from 600 000 to 300 000).

(d)

Male 1988

Data in order:
188, 198, 290, 328, 360, 445, 572, 645, 661, 665, 680, 682, 686, 688, 728, 729, 733

There are 17 scores. The median of data is the 9th score.

$Q_2 = 661$

Lower half of data (excluding Q_2): 188, 198, 290, 328, 360, 445, 572, 645

The median of the lower half is the average of the 4th score and the 5th score.

$$Q_1 = \frac{328+360}{2} = 344$$

Upper half of the data (excluding Q_2): 665, 680, 682, 686, 688, 728, 729, 733

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{686+688}{2} = 687$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 687 - 344 \\ &= 343 \end{aligned}$$

Test for outliers.

$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 344 - 1.5 \times 343 \\ &< 0 \end{aligned}$$

All scores lie above the lower fence, so no low outliers.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 687 + 1.5 \times 343 \\ &= 1201.5 \end{aligned}$$

All scores are below the upper fence so there are no outliers.

The minimum is 188 and the maximum is 733.

The box plot is drawn below, after Female 2012.

Female 1988

Data in order:

$269,\ 338,\ 347,\ 348,\ 365,\ 433,\ 554,\ 613,\ 628,\ 648,\ 650,\ 654,\ 658,\ 689,\ 703,\ 713,\ 724$

There are 17 scores. The median of data is the 9th score.

$Q_2 = 628$

Lower half of data (excluding Q_2): $269,\ 338,\ 347,\ 348,\ 365,\ 433,\ 554,\ 613$

The median of the lower half is the average of the 4th score and the 5th score.

$$Q_1 = \frac{348+365}{2} = 356.5$$

Upper half of the data (excluding Q_2): $648,\ 650,\ 654,\ 658,\ 689,\ 703,\ 713,\ 724$

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{658+689}{2} = 673.5$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 673.5 - 356.5 \\ &= 317 \end{aligned}$$

Test for outliers.

$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 356.5 - 1.5 \times 317 \\ &< 0 \end{aligned}$$

All scores lie above the lower fence, so no low outliers.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 673.5 + 1.5 \times 317 \\ &= 1149 \end{aligned}$$

All scores are below the upper fence so there are no outliers.

The minimum is 269 and the maximum is 724.

The box plot is drawn below, after Female 2012.

Male 2012

Data in order:

$265,\ 348,\ 375,\ 521,\ 633,\ 676,\ 681,\ 689,\ 692,\ 731,\ 745,\ 771,\ 796,\ 800,\ 808,\ 814,\ 815$

There are 17 scores. The median of data is the 9th score.

$Q_2 = 692$

Lower half of data (excluding Q_2): $265,\ 348,\ 375,\ 521,\ 633,\ 676,\ 681,\ 689$

The median of the lower half is the average of the 4th score and the 5th score.

$$Q_1 = \frac{521+633}{2} = 577$$

Upper half of the data (excluding Q_2): $648,\ 650,\ 654,\ 658,\ 689,\ 703,\ 713,\ 724$

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{796+800}{2} = 798$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 798 - 577 \\ &= 221 \end{aligned}$$

Test for outliers.

$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 577 - 1.5 \times 221 \\ &= 245.5 \end{aligned}$$

All scores lie above the lower fence, so no low outliers.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 798 + 1.5 \times 221 \\ &= 1129.5 \end{aligned}$$

All scores are below the upper fence so there are no outliers.

The minimum is 265 and the maximum is 815.

The box plot is drawn below, after Female 2012.

Female 2012

Data in order:
$301,\ 392,\ 527,\ 546,\ 642,\ 648,\ 653,\ 656,\ 684,\ 692,\ 737,\ 753,\ 762,\ 762,\ 781,\ 782,\ 783$

There are 17 scores. The median of data is the 9th score.

$Q_2 = 684$

Lower half of data (excluding Q_2): $301,\ 392,\ 527,\ 546,\ 642,\ 648,\ 653,\ 656$

The median of the lower half is the average of the 4th score and the 5th score.

$$Q_1 = \frac{546+642}{2} = 594$$

Upper half of the data (excluding Q_2): $692,\ 737,\ 753,\ 762,\ 762,\ 781,\ 782,\ 783$

The median of the upper half is the average of the 4th score and the 5th score in the upper half.

$$Q_3 = \frac{762+781}{2} = 771.5$$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 771.5 - 594 \\ &= 177.5 \end{aligned}$$

Test for outliers.

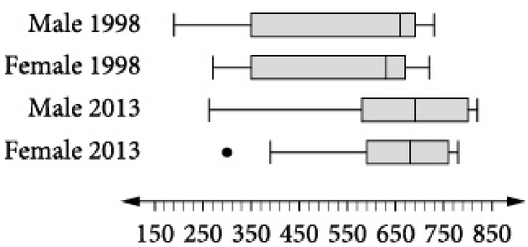
$$\begin{aligned} \text{Lower fence } &= Q_1 - 1.5 \times IQR \\ &= 594 - 1.5 \times 177.5 \\ &= 327.75 \end{aligned}$$

301 lies below the lower fence and is therefore an outlier.

$$\begin{aligned} \text{Upper fence } &= Q_3 + 1.5 \times IQR \\ &= 771.5 + 1.5 \times 177.5 \\ &= 1037.75 \end{aligned}$$

All scores are below the upper fence so there are no more outliers.

The minimum, excluding the outlier, is $301$and the maximum is 783.



Sample answer:

There is an outlier in females for 2013.

The median population age is greater in 2013 than 1998.

The maximum is also greater for 2013.

1998 has the least of any age group, with 188 for the 80+ age group.

The middle 50% of ages are more varied in 1998 than for 2013.

CHAPTER REVIEW 11

2 **D** Discrete

Data is counted, so it is (numerical) discrete.

4 **C** 2.8

x	*f*	fx
1	3	3
2	5	10
3	2	6
4	4	16
5	2	10
	$\sum f = 16$	$\sum fx = 45$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{45}{16} = 2.8125 \approx 2.8$$

6 **D** 2

The most frequently occurring number, the number with the highest frequency, is 2.

The mode is 2.

8 **D** 66

There are 37 data values, so the median is the 19th data value
The median is 66.

10 **B** 4

First put the data in order.

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

There are 39 scores, so the median is the 20th data value.

$Q_2 = 4$

Not including the median there are 19 scores to the left of the median.

The first quartile will be the tenth score, which is 2.

$Q_1 = 2$

Not including the median there are 19 scores to the right of the median.

 ISBN 978 1 4886 1829 1

The third quartile will be the tenth of these, or ten from the right, which is 6.

$Q_3 = 6$

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 6 - 2 \\ &= 4 \end{aligned}$$

12 B

The mean will increase as one of the scores has increased, and the rest are unchanged.

The data value has moved further away from the mean, and is now nearly the maximum score, so the standard deviation would increase.

14 A

There is a lot of time in these calculations. Students may alternatively use a spreadsheet (giving the same table) or a calculating device.

Height of student (cm)	Mid-point	Frequency	fx	$\left(x - \overline{x}\right)^2$	$f\left(x - \overline{x}\right)^2$
140–<145	142.5	5	712.5	637.484…	3187.420…
145–<150	147.5	11	1622.5	409.999…	4509.995…
150–<155	152.5	18	2745	232.515…	4185.272…
155–<160	157.5	26	4095	105.030…	2730.797…
160–<165	162.5	44	7150	27.546…	1212.032…
165–<170	167.5	92	15 410	0.061…	5.678…
170–<175	172.5	64	11 040	22.577…	1444.944…
175–<180	177.5	32	5680	95.092…	3042.969…
180–<185	182.5	17	3102.5	217.608…	3699.3412…
185–<190	187.5	9	1687.5	390.123…	3511.114…
190–<195	192.5	4	770	612.639…	2450.557…
Totals		322	54 015		29980.12422

 ISBN 978 1 4886 1829 1

First fill in the first four columns to find the mean.

$$\overline{x} = \frac{\sum fx}{\sum f} = \frac{54\ 015}{322} \approx 167.75$$

This is a population, not a sample, as every height is included. Divide by $n = 322$.

$$Var(x) = \frac{29\ 980.124\ 22}{322} = 93.105...$$

$$\begin{aligned}\sigma_x &= \sqrt{93.105...} \\ &= 9.649... \\ &\approx 9.65\end{aligned}$$

16 (a) On the face of it, the data in the table does not seem to support the sub-headline because the figure of 16.3 is less than those for the majority of other states.

(b) This is true as the table reads 16.3 next to Victoria for 2012 and 'per million of population' is the same as 'for every million residents'.

(c) $\dfrac{92}{16.3/1\ 000\ 000} = 92 \times \dfrac{1\ 000\ 000}{16.3} = 5\ 644\ 171.779...$

Round to the nearest thousand: $5\ 644\ 000$

(d) $\dfrac{15.6}{1\ 000\ 000} = \dfrac{x}{22\ 692\ 308}$

$$\begin{aligned}x &= \frac{15.6 \times 22\ 682\ 308}{1\ 000\ 000} \\ &= 353.8... \\ &\approx 354\end{aligned}$$

(e) If 1052 Australians received transplants, then at least 1052 organs must have been donated. (some may have received more than one organ.) If there were only 354 organ donors, some donors must have donated more than one organ.

(f) It would be a lot quicker, cheaper and reliable to use a census by accessing all hospital records than to do a survey, particularly as such a low proportion of the population donated or received

 ISBN 978 1 4886 1829 1

organs, and many of the donors would be deceased. Therefore it is most likely that the data is from a census of the population rather than from a sample.

(g) The article has analysed the data and is communicating results to the community via the press. This is the interpretation and communication of statistical information.

18

x	Frequency	Cumulative Frequency
0–4	2	2
5–9	5	7
10–14	7	14
15–19	3	17
20–24	1	18
	$\sum f = 18$	

There are 18 results, so the median is the average of the 9th and 10th data values. Both fall in the interval 10–14.

The median is in the interval 10–14.

20 (a) The method will depend on the kind of device used.

$$\begin{aligned}\sigma_x &= 2.983...\\ &\approx 2.98\end{aligned}$$

(b) $$\begin{aligned}s_x &= 3.065...\\ &\approx 3.07\end{aligned}$$

22 First count the number of times each number occurs. It may be useful to first list this in a table.

0 pets: 3 1 pet: 3

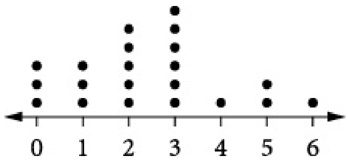
2 pets: 5 3 pets: 6

4 pets: 1 5 pets: 2

 ISBN 978 1 4886 1829 1

6 pets: 1

Draw a number line from 0 to 6. Place 3 dots equally spaced above 0, 3 dots equally spaced above 0, 5 dots equally spaced above 2, etc.



24 Sample response: By far the most deaths are to drivers (other than motorcyles) but this has slightly declined from 2010 to 2012.Pedestrian fatalities have increased from 2010 to 2012; they have about doubled. The number of drivers killed decreased between 2010 and 2011 but increased a little in 2012, though not back to the 2010 level. Passenger deaths have fallen each year. There has not been much movement between 2010 to 2012 in the other figures. There have been very few bicyclists or pillion passenger deaths, but this probably reflects the smaller numbers of these on the roads. A more meaningful graph may be of the proportion of each category killed.

26 (a) Dent in door: $\frac{16}{45}\times 100\% \approx 36\%$

Door seal: $\frac{12}{45}\times 100\% \approx 27\%$

Damaged box: $\frac{8}{45}\times 100\% \approx 18\%$

Faulty light: $\frac{6}{45}\times 100\% \approx 13\%$

Wiring defect: $\frac{2}{45}\times 100\% \approx 4\%$

Refrigerant leak: $\frac{1}{45}\times 100\% \approx 2\%$

Add the percentages as you go.

Dent in door: 36%

Door seal: $36\% + 27\% = 63\%$

Damaged box: $63\% + 18\% = 81\%$

Faulty light: $81\% + 13\% = 94\%$

Wiring defect: $94\% + 4\% = 98\%$

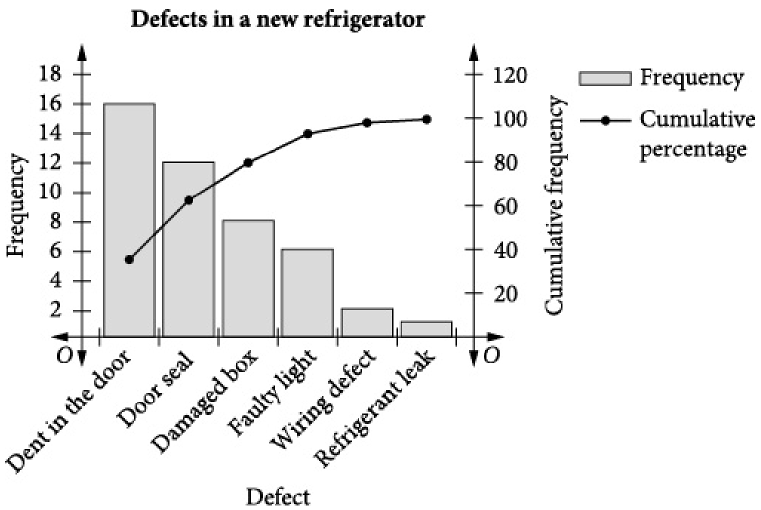
Refrigerant leak: $98\% + 2\% = 100\%$

Fill these numbers in two extra columns in the table.

Defect	Frequency	Percentage	Cumulative percentage
Dent in door	16	36%	36%
Door seal	12	27%	63%
Damaged box	8	18%	81%
Faulty light	6	13%	94%
Wiring defect	2	4%	98%
Refrigerant leak	1	2%	100%
	45	100%	

Using a suitable scale, draw a bar graph of the frequencies and a line graph of the cumulative percentages on the same grid.

(b)



(c) They should take more care with the doors, or possibly inspect them more carefully before the refrigerators are sold, as door problems make up nearly two thirds of all complaints. They may wish to look at their packing and freighting, as $36\% + 18\% = 54\%$ (dents and damage to the box, over half) of problems could occur as the refrigerators are moved from the factory to the shops.

 ISBN 978 1 4886 1829 1