Representations of data, Mixed Exercise 3

1 **a** Q_1: $\frac{31}{4} = 7.75$ so we pick the 8th value: 178

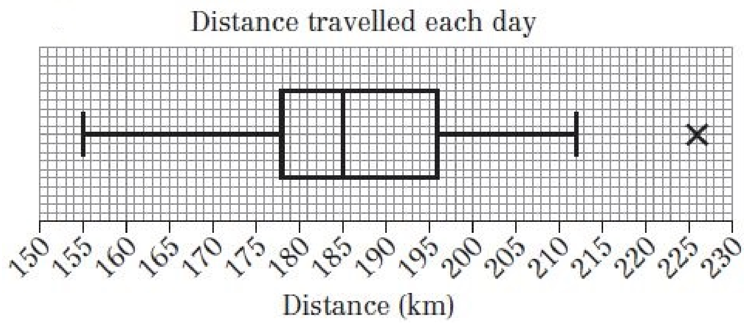
Q_2: $\frac{31+1}{2} = 16$ so we pick the 16th value: 185

Q_3: $\frac{3}{4} \times 31 = 23.25$ so we pick the 24th value: 196

b $Q_1 - 1.5(Q_3 - Q_1) = 178 - 1.5(196 - 178) = 151$
$Q_3 + 1.5(Q_3 - Q_1) = 196 + 1.5(196 - 178) = 223$

So 226 km is an outlier.

c



2 **a** 45 minutes

b 60 minutes

c This is an outlier that does not fit the pattern.

d The Irt club had the highest median, so overall they had the slowest runners.

The IQR ranges were about the same, with the Irt club slightly more spread out.

e With the exception of the outlier, the Esk Club runners were faster in every respect. Their minimum, Q_1, Q_2, Q_3 and maximum times were all lower than the corresponding times for the Irt Club runners.

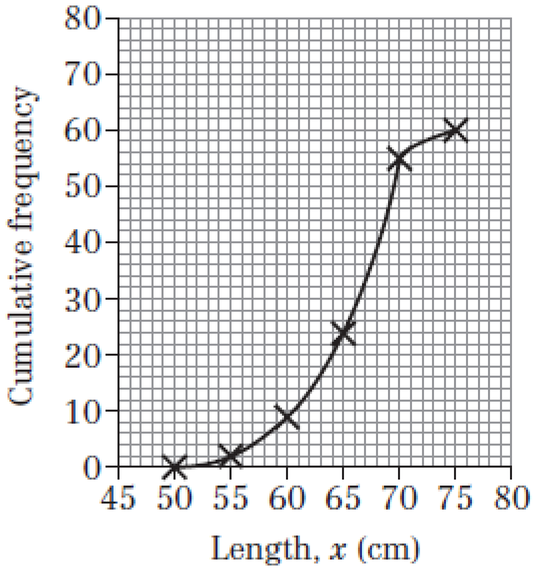
f Advantages, any one from:

It helps us to see the spread of the data easily.
The plot is clear and easy to understand.
It uses the range and the median values.
It is easy to compare the stratified data.

Disadvantages, any one from:

Original data is not clearly shown in the box plot.
Mean and mode cannot be identified using the box plot.
It can be easily misinterpreted.
If large outliers are present, the box plot is more likely to give an incorrect representation.

3 **a** Cumulative frequencies are: 2, 9, 24, 55, 60



b Median = 66 cm

c IQR = 69 − 62.5 = 6.5 cm

d Median = 65.5 cm

IQR = 70 − 63.25 = 6.75

The median length of the honey badgers is slightly higher than the median length for the European badgers, showing that honey badgers are only a little longer on average than European badgers. Comparing interquartiles ranges, the spread of lengths of European badgers is slightly greater than the spread of lengths of honey badgers.

e The cumulative frequency diagram shows us how the lengths of the badgers is spread and enables us to estimate the median and quartiles, but tells us little about individual data points.

4 **a** The areas of the bars is proportional to the frequency represented.

$2k(1+1.5+5.5+4.5)+4k(1)=58$

$29k=58$ so $k=2$

Number of girls who took longer than 56 seconds $=2\big((4.5\times 2)+(1\times 4)\big)=26$ girls

b Number of girls who took between 52 and 55 seconds $=2\big((1.5\times 2)+(\tfrac{1}{2}\times 2\times 5.5)\big)=17$ girls

5 $1.5\times 5.7\times k=2565$, so $k = 300$

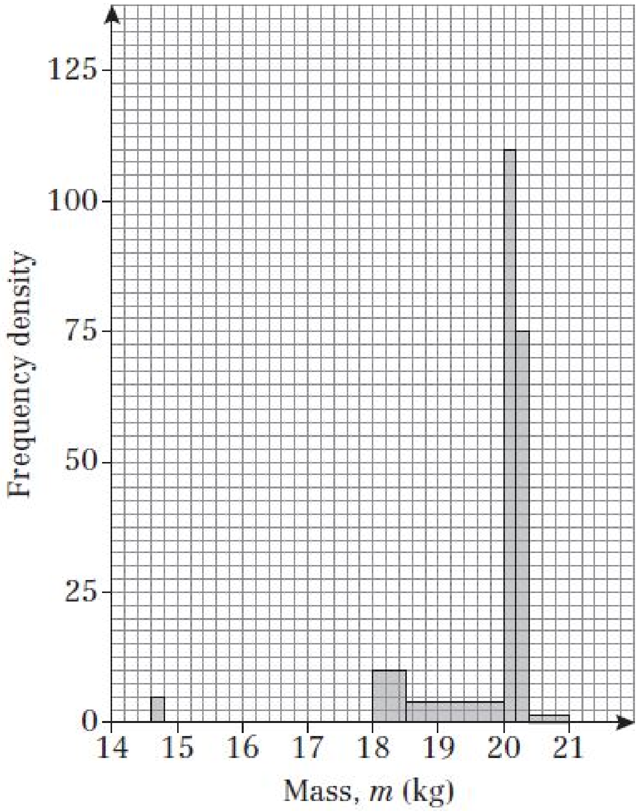
a $\text{width}=1\times 1.5=1.5\,\text{cm}$

$$\text{height}=\frac{\text{frequency}}{k\times\text{width}}=\frac{1170}{300\times 1.5}=2.6\,\text{cm}$$

5 b width $= 5 \times 1.5 = 7.5\,\text{cm}$

height $= \dfrac{630}{300 \times 7.5} = 0.28\,\text{cm}$

6 a



b Mean $= \dfrac{\sum fy}{n} = \dfrac{988.85}{50} = 19.777$ kg

Standard deviation $= \sqrt{\dfrac{\sum fy^2}{n} - \mu^2} = \sqrt{\dfrac{19602.84}{50} - 19.777^2} = \sqrt{0.927} = 0.963$ (to 3 s.f.)

c Median $= 20.0 + \dfrac{13.5}{22} \times 0.2 = 20.123$ (to 3 d.p.)

7 a $\dfrac{312}{14} = 22.286$ (to 3 d.p.)

b Median is the $\dfrac{14+1}{2} = 7.5$th piece of data: 20

Q_1 is the $\dfrac{14}{4} = 3.5$th piece of data, so we choose the 4th: 13

Q_3 is the $\dfrac{3 \times 14}{4} = 10.5$th piece of data, so we choose the 11th: 31

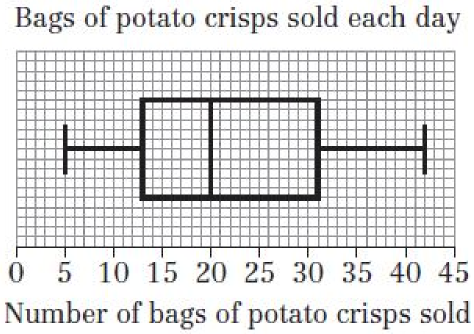
7 c $\text{IQR} = 31 - 13 = 18$ so $1.5 \times \text{IQR} = 27$

$13 - 27 = -14$

$31 + 27 = 58$

So there are no outliers.

d



8 a The maximum gust is a continuous variable and the data is given in a grouped frequency table.

b $10 \leqslant g < 15$ bar:

Width = 2.5, class width = 5, so width = class width × 0.5
Area = 2.5 × 1.8 = 4.5 cm^2, frequency = 3, so area = frequency × 1.5

$18 \leqslant g < 20$ bar:

Width = class width × 0.5 (from above) = 2 × 0.5 = 1 cm
Area = frequency × 1.5 (from above) = 9 × 1.5 = 13.5 cm^2

$$\text{Height} = \frac{\text{Area}}{\text{Width}} = 13.5 \text{ cm}$$

c $$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{1334.5}{57} = 23.4 \text{ kn (to 3 s.f.)}$$

$$\text{Standard deviation} = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\frac{34299.25}{57} - \left(\frac{1334.5}{57}\right)^2} = 7.32 \text{ (to 3 s.f.)}$$

d Using exact figures throughout:

lower boundary = mean – standard deviation = 16.1 (to 3 s.f.)
higher boundary = mean + standard deviation = 30.7 (to 3 s.f.)

$$\frac{\text{lower boundary} - 15}{18 - 15} = \frac{g - 3}{12 - 3} \Rightarrow g = 6.27 \text{ (to 3 s.f.)}$$

$$\frac{30.72 - 30}{50 - 30} = \frac{h - 50}{57 - 50}$$

$h - g = 44.0$ (to 3 s.f.)

9 a Mean for 1987 = $\frac{\sum x}{n} = \frac{356.1}{30}$ = 11.9 °C (to 1 d.p.)

Mean for 2015 = $\frac{\sum x}{n} = \frac{364.1}{30}$ = 12.1 °C (to 1 d.p.)

b Standard deviation for 1987 = $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{4408.9}{30} - \left(\frac{356.1}{30}\right)^2}$ = 2.46 °C (to 3 s.f.)

The mean temperature was slightly higher in 2015 than in 1987. The standard deviation of temperatures was higher in 1987 than in 2015, so the temperatures were more spread out.

c In all calculations, exact figures are used.
critical temperature = mean − 1.02 = 11.08 °C

Let a = the number of the day with this temperature.

$$\frac{\text{critical temperature} - 10.1}{14.1 - 10.1} = \frac{a - 0}{30 - 0} \text{ so } a = 7.35$$

So there are 7 days where the temperature $\leqslant$ 11.08 °C.

critical temperature = mean + 1.02 = 13.12 °C

Let b = the number of the day with this temperature.

$$\frac{\text{critical temperature} - 10.1}{14.1 - 10.1} = \frac{b - 0}{30 - 0} \text{ so } b = 22.65$$

So there are 22 days where the temperature ≤ 13.12 °C.

30 − (b − a) = 14.7 (to 3 s.f.)

So there are 15 'abnormal' days.

Assumption: the temperatures are equally distributed throughout the range.

Challenge

Length (mins)	Frequency	Area of bar	Class width	Bar height
70–89	4	$4k$	20	x
90–99	17	$17k$	10	$x + 3$
100–109	20	$20k$	10	
110–139	9	$9k$	30	
140–179	2	$2k$	40	

Area of 70–89 bar = $20x = 4k$, so $x = \frac{k}{5}$

Area of 90–99 bar = $10(x + 3) = 10x + 30 = 17k$

Using substitution

$10 \times \frac{k}{5} + 30 = 17k$, so $2k + 30 = 17k$, i.e. $k = 2$

$x = \frac{k}{5} = 0.4$

Area of 110–139 class = $9k = 9 \times 2 = 18$ cm^2

Height = $\frac{\text{Area}}{\text{Class width}} = \frac{18}{30} = 0.6$ cm