

KEY WORDS:

| | |
|-----------------|-----------|
| bias | sample |
| continuous data | secondary |
| discrete data | survey |
| experiment | two-way |
| primary | table |
| questionnaire | |

You can count discrete data, for example the number of eggs in a nest. Continuous data is data you measure, for example weight, height.

- **Primary** data is data that you collect yourself. You can collect data by conducting a **survey** or an **experiment**.
- Secondary data is data taken from an existing source such as newspapers or the internet.
- Data can be **discrete** or **continuous**.
- A **two-way table** shows two sets of data about the same group of people or things, for example, hair colour and eye colour.

If the group you are surveying is small you can collect data from everyone, but if the group is large you need to choose a **sample**. You have to ensure that the sample is not biased. **Bias** can arise if, for example, you only ask your friends or you only ask people in a certain age group. In general the larger the sample size the more reliable is the data.

| | |
|------------|-------------|
| | Hair colour |
| Eye colour | |

- To choose a **random sample**, assign a number to each possible member of the sample and use a random number generator to pick numbers.
- **Questionnaires** should be relevant and useful to your survey.

Use clear language. Cover all options and leave no gaps between the answer boxes.

EXAMPLE

Wendy carried out a survey to find out how much people spent on food each week. She asked 100 people at the supermarket closest to where she lives on a Tuesday morning. Explain why Wendy's survey might be biased and advise her on how to avoid this bias.

- Only asking shoppers at one supermarket - she should ask shoppers at all supermarkets.
- Only asking shoppers who shop at a store - she should include people who shop online.
- Only asking shoppers on one weekday - she should ask on other days including weekends.
- Only asking shoppers at one time of day - she should vary the time of day when she carries out the survey.

EXAMPLE

This question appeared in a survey about time spent using a mobile phone. How much time do you spend on your mobile phone?

Less than 5 minutes Up to 10 minutes Over an hour

Criticise the question, and write a better question.

- The question does not have a time frame.
- It does not specify just phone calls or also using the phone to listen to music or play games.
- The answer choices have gaps and overlaps.
- How long, on average, do you spend making calls on your mobile each day?
- Less than 10 minutes 10–30 minutes Over 30 minutes

Exercise S1

- 1 Terry began drawing this observation sheet to collect data on women's shoe sizes and glove sizes.

| | |
|------------|-----------|
| | Shoe size |
| Glove size | |

One manufacturer makes women's shoes in sizes that range from 3 to 8 and gloves in sizes that range from extra small to large.

- a Copy the observation sheet and put on it the sizes.
 b In Terry's survey 7 people had shoe size 5 and glove size medium
 9 people had shoe size 4 and glove size small
 Add this information to your observation sheet.

- 2 Design an observation sheet to collect data on colour and type of vehicle in a road traffic survey. You will need to choose categories for the different types of vehicle.

you will need to think of your own categories for type of vehicle.

- 3 Answers to a crossword puzzle are completed either across or down. In a crossword puzzle book all the crosswords have across answers 4 to 9 letters long and down answers 5 to 10 letters long.
- a i Design an observation sheet to capture data about the length of answers in this crossword book.
 ii There were 12 crosswords with across answers 7 letters long and down answers 6 letters long. Write this value in the correct position of your observation sheet.
- b Debbie wants to do a survey of the crosswords in the book. She gives each of twelve colleagues a copy of the crossword book and a questionnaire to complete.
 Give two reasons why Debbie's survey may be biased.

- 4 Below are the questions used in a survey on hair products. Some of the questions are not suitable.

Write down what is wrong with them and write better questions.

- a What colour is your hair? Blonde Brown Black
 b Is your hair dyed? Yes No
 c How long is your hair? Short Medium Long
 d How many times per week do you wash your hair?
 Once Twice Every day
 e Do you always use conditioner? Yes No Sometimes

- The range of a data set = largest value – smallest value
- An average is a representative value of a set of data.
- The mode is the data value that occurs most often.
- The median is the middle value when data is arranged in order.

If there are two middle numbers, the median is the middle value of these two numbers, for example, for the data set

2 2 4 5 7 8 10 11

the middle values are 5 and 7, so the median is 6.

- The mean is calculated by adding all the data values then dividing by the number of pieces of data.

KEY WORDS:

| | |
|---------|---------------|
| average | mode |
| mean | stem and leaf |
| range | diagram |
| median | |

A data set can have no mode, one mode or more than one mode.

A data set can have no mode or more than one mode

Discrete data is countable eg number of eggs
Continuous data is measurable eg weight of eggs

You can use the measures, range, mode, median and mean to compare data sets

When comparing data sets always compare the same type of measure, for example mean with mean, not mean with mode.

EXAMPLE

These are the monthly allowance for a group of students

£27 £34 £24 £10 £15 £12 £30 £16

- a** Work out the
- i** mode **ii** median **iii** mean **iv** range
- b** Another student's pocket money is £27. If this amount is included how would it affect the measures calculated in part **a**?

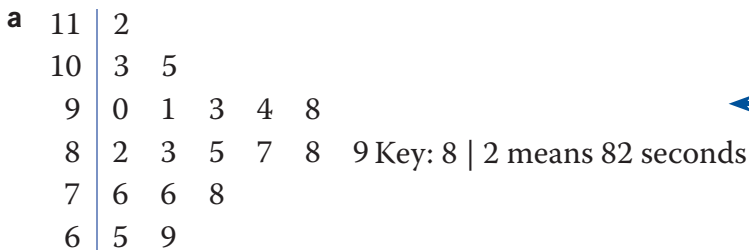
- a i** There is no mode.
- ii** In order £10 £12 £15 £16 £24 £27 £30 £34
Two middle values £16 and £20, so median = £18
- iii** Total $27 + 34 + 24 + 10 + 15 + 12 + 30 + 16 = £168$
 $168 \div 8 = 21$ Mean = £21
- iv** Range $34 - 10 = £24$
- b** £27 is greater than mean and median so both of these will increase.
There will be two amounts of £27, so £27 is now the mode.
The range is unchanged as £27 lies between the highest and lowest amounts.

- A stem and leaf diagram displays all the data in a small data set.
 - The stem is written on one side of a vertical line, with the leaves on the other side.
 - Leaves are written in order, smallest next to the stem.
 - Always write a key.

The times taken, in seconds, for 19 children to complete a jigsaw are

69 103 94 65 88 76 78 93 105 112 83 98 85 89 91 76 87 90 82

- a** Draw a stem and leaf diagram to display these data.
b Write down **i** the median **ii** the range.



Use tens as the stem and units as the leaves.
 112 is 11 tens and 2 units.
 Choose any value in the diagram for the key.

- b** **i** Median $\frac{1}{2}(19 + 1)$ th = 10th value, median = 88 seconds
ii Range = 112 – 65 = 47 seconds

Exercise S2

- 1** Gemma has a spelling test each week. The spelling tests are always scored out of the same total. Her scores for fifteen consecutive weeks are
- 18 13 15 16 17 16 9 12 7 11 20 19 6 12 14
- a** One week Gemma scored full marks on her spelling test. What is the total possible score for the spelling tests each week?
b What is the range of Gemma's scores?
c Work out the score that is
i the mode **ii** the median **iii** the mean
d In week sixteen Gemma scored 19. When this score is included what is her new
i mode **ii** median **iii** mean?
- *2** The table shows the scores of a multiplication test and a division test taken by a group of ten students.

| | | | | | | | | | | |
|-----------------------|----|----|----|----|----|----|----|----|----|----|
| Multiplication | 25 | 18 | 20 | 20 | 16 | 12 | 14 | 20 | 23 | 9 |
| Division | 22 | 19 | 13 | 22 | 14 | 15 | 15 | 19 | 19 | 12 |

Were the students better at multiplication or division?
 You must justify your answer with calculations, explaining how they help you decide.

- 3** These are the times taken, to the nearest minute, to deliver leaflets to 15 roads.
- 22 15 28 34 29 9 11 24 23 16 20 30 31 32 41
- a** Work out the mean delivery time.
b Represent these data on a stem and leaf diagram.

Overmatter exercise S2

- 4 The average attendance, to the nearest 100, at premier league football games in 2009/10 were as follows.

4600 4300 4100 2800 3900 3700 6000 3600 2400 4000
1800 2500 2700 3400 2400 1800 2100 2500 7500 2200

- a Represent these data on a stem and leaf diagram.
b Work out
i the mode ii the range of the average attendance.
- 5 The back to back stem and leaf diagram shows the lengths of samples of two different species of caterpillar.

| | Species A | | | | Species B | | | | | |
|---|-----------|---|---|---|-----------|---|---|---|---|---|
| | 5 | 4 | 7 | 2 | | | | | | |
| | 9 | 4 | 3 | 6 | 0 | 1 | 5 | | | |
| | 7 | 7 | 5 | 5 | 2 | 3 | 4 | 6 | 7 | 8 |
| 9 | 8 | 5 | 3 | 4 | 0 | 1 | 9 | 9 | | |
| 7 | 5 | 4 | 3 | 3 | 1 | 4 | 5 | 6 | | |
| | 6 | 4 | 1 | 2 | 9 | | | | | |

Key: 1 | 2 | 9 means 21 mm for species A and 29 mm for species B

Make three comparisons between these samples of species of caterpillars.

You can summarise large amounts of raw data in a table to show the frequencies.

| | | | | |
|------------------|---|----|----|----|
| Shoe size | 3 | 4 | 5 | 6 |
| Frequency | 2 | 39 | 36 | 13 |

Mode is 4 Median is 5

You use the frequencies to find the various averages. Do not use frequencies as averages.

KEY WORDS:

estimated mean
median
modal class
mode

EXAMPLE

On a box of drawing pins the average contents are stated as 50. The table shows the number of drawing pins in 35 boxes.

| | | | | | | | |
|---------------------|----|----|----|----|----|----|----|
| Drawing pins | 48 | 49 | 50 | 51 | 52 | 53 | 54 |
| Frequency | 3 | 6 | 21 | 1 | 2 | 1 | 1 |

Joni says that the average is 50, whatever measure of average is used. Explain why Joni is correct.

Mode is 50 as 50 has the highest frequency (21).

35 boxes, middle value is $\frac{1}{2}(35 + 1) = 18$ th value, $3 + 6 = 9$, $9 + 21 = 30$ so the 18th value is 50, so median = 50

Total number of pins = $48 \times 3 + 49 \times 6 + 50 \times 21 + 51 + 52 \times 2 + 53 + 54$
= 1750

mean = $1750 \div 35 = 50$

All three averages are 50, Joni is correct.

To find the median in a small data set find $\frac{1}{2}(n + 1)$ th value.

Imagine listing the number of pins in order, 50 would be the 18th number in the list.

You can group data when there are lots of different values.

- In grouped data
 - The **modal class** is the class with the highest frequency.
 - You can find the class in which the **median** lies, not an actual median value.
 - You calculate an **estimated mean** using the midpoint and the frequency of each class.

EXAMPLE

The table summarises the number of miles Reuben cycled each day for the first 29 days in April.

| | | | | |
|--------------------------------------|-----------------|------------------|------------------|------------------|
| Miles cycled (m) | $0 \leq m < 20$ | $20 \leq m < 40$ | $40 \leq m < 60$ | $60 \leq m < 80$ |
| Frequency | 5 | 9 | 11 | 4 |

Estimate does not mean guess - use the midpoint and the frequency of each class. Remember to divide by the total frequency not the number of classes.

- Work out an estimate of the mean number of miles Reuben cycled.
- How many miles could Reuben cycle on April 30th so that the class interval in which the median lies does not change? Explain your answer.

a Total miles $10 \times 5 + 30 \times 9 + 50 \times 11 + 70 \times 4 = 1150$
 $1150 \div 29 = 39.655\dots$ mean = 39.7 miles

b $\frac{1}{2}(29 + 1) = 15$ th value $5 + 9 = 14$, so the median (15th value) is in the class $40 \leq m < 60$

With 30 values median will be $\frac{1}{2}(30 + 1) = 15.5$ th value, so you average 15th and 16th values, Both must be in the $40 \leq m < 60$ class for the median to stay where it is.

Exercise S3

- 1 Henry drew this table to show the number of tracks on each of the CDs he owned.

| | | | | | |
|------------------|---|---|----|----|----|
| Number of tracks | 7 | 8 | 9 | 10 | 11 |
| Number of CDs | 4 | 6 | 12 | 5 | 2 |

- a Write down
 i the median ii the modal number of tracks.
- b Work out the mean number of tracks.
- 2 The numbers and prices of theatre tickets available for a performance are shown in the table.

| | | | | |
|------------------|-----|-----|-----|-----|
| Ticket price, £ | £50 | £40 | £25 | £15 |
| Number available | 140 | 50 | 200 | 10 |

- a Calculate the mean price of a theatre ticket.
- b For a special performance one evening all ticket prices were reduced by £5.
 What was the mean price of a theatre ticket for that performance?
- 3 The number of hats of different hat sizes sold by a department store is given in the table.

| | | | | | | | | | | | |
|-------------|----------------|----------------|----------------|----------------|----------------|----|----------------|----------------|----------------|----------------|----------------|
| Hat size | $6\frac{3}{8}$ | $6\frac{1}{2}$ | $6\frac{5}{8}$ | $6\frac{3}{4}$ | $6\frac{7}{8}$ | 7 | $7\frac{1}{8}$ | $7\frac{1}{4}$ | $7\frac{3}{8}$ | $7\frac{1}{2}$ | $7\frac{5}{8}$ |
| Number sold | 1 | 0 | 2 | 3 | 8 | 15 | 12 | 3 | 0 | 0 | 0 |

When the manager of the store orders new hats to sell which type of average should he use? Give reasons for your answer and show your calculations.

- 4 The table shows the times a sample of 84 students spent doing homework one evening.

| | | | | | | |
|-------------------|-----------------|------------------|------------------|------------------|-------------------|--------------------|
| Time, m minutes | $0 < m \leq 20$ | $20 < m \leq 40$ | $40 < m \leq 60$ | $60 < m \leq 80$ | $80 < m \leq 100$ | $100 < m \leq 120$ |
| Frequency | 1 | 21 | 11 | 22 | 8 | 21 |

- a Work out an estimate of the mean time.
- b Which class contains the median?
- c Explain why the modal class may not be a good measure of average to use with these data.
- 5 Ahmed recorded the number of hours of sunshine in May and June.

May

| Number of hours sunshine h | Number of days |
|------------------------------|----------------|
| $0 \leq h < 2$ | 5 |
| $2 \leq h < 4$ | 14 |
| $4 \leq h < 6$ | 7 |
| $6 \leq h < 8$ | 5 |

June

| Number of hours sunshine h | Number of days |
|------------------------------|----------------|
| $0 \leq h < 2$ | 5 |
| $2 \leq h < 4$ | 7 |
| $4 \leq h < 6$ | 10 |
| $6 \leq h < 8$ | 8 |

- a Calculate estimates of the mean number of hours of sunshine in May and in June.
- b Write down i the class which contains the median ii the modal class for both May and June.

KEY WORDS:

- bar chart
- discrete
- mode
- pie chart

You can use a pie chart to represent non-numerical data (like flavours and colours).

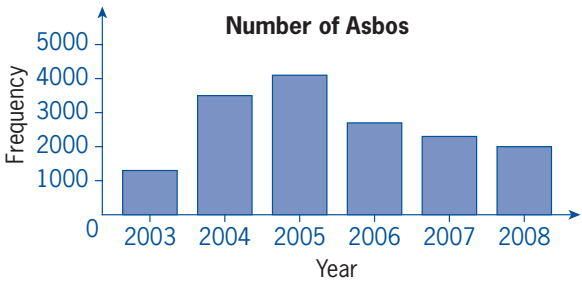
- ▶ A pie chart is a circle shared out between the Frequencies.
 - the bigger the frequency the larger the share.
 - The **mode** has the largest share.
- ▶ To draw a pie chart work out $360^\circ \div$ total frequency, multiply this answer by the frequency of each class to find its share of the pie.
- A **bar chart** represents **discrete** data and can be used for data given in categories.
 - ▶ All bars are rectangles of the same width with gaps between the bars.
 - ▶ The height of each bar represents the frequency
 - ▶ The mode is the highest bar

EXAMPLE

The number of Asbos, to the nearest 100, issued each year from 2003 to 2008 are given in the table.

| Year | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|-------|------|------|------|------|------|------|
| Asbos | 1300 | 3500 | 4100 | 2700 | 2300 | 2000 |

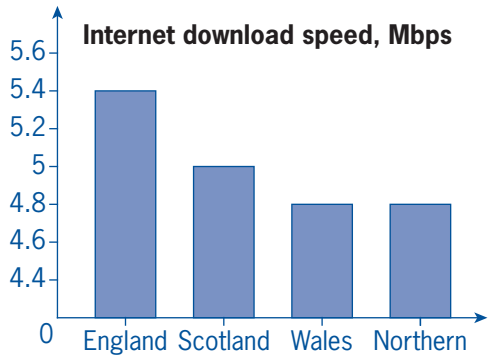
Draw a bar chart to represent these data.



EXAMPLE

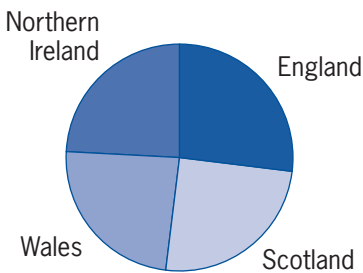
This bar chart shows the estimated maximum internet download speeds in megabytes per second in Great Britain in May 2010.

Draw a pie chart for these data.



Total download speeds $5.4 + 5.0 + 4.8 + 4.8 = 20$
 $360 \div 20 = 18^\circ$

- England $5.4 \times 18 = 97.2^\circ$
- Scotland $5.0 \times 18 = 90.0^\circ$
- Wales $4.8 \times 18 = 86.4^\circ$
- NI $4.8 \times 18 = 86.4^\circ$
- Check Total 360°



Label each sector or use a key.
 The pie chart must cover the whole circle.
 Check that your angles add up to 360° .

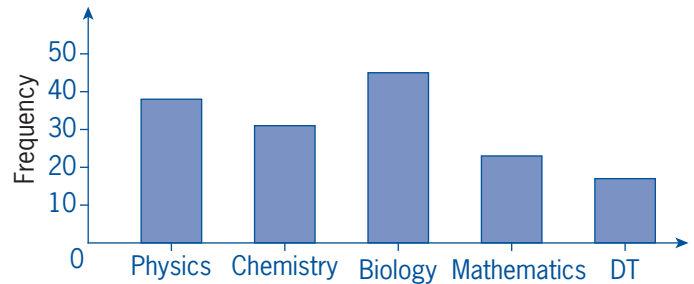
Exercise S4

- 1 The owner of a beach café drew this table to show the number and type of drinks he sold one day in August.

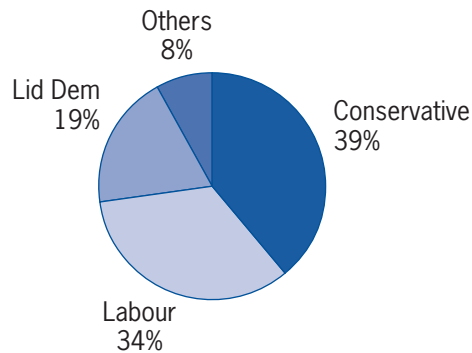
| Tea | Coffee | Water | Fizzy drink | Still drink |
|-----|--------|-------|-------------|-------------|
| 80 | 64 | 160 | 40 | 16 |

Draw a bar chart to display the information.

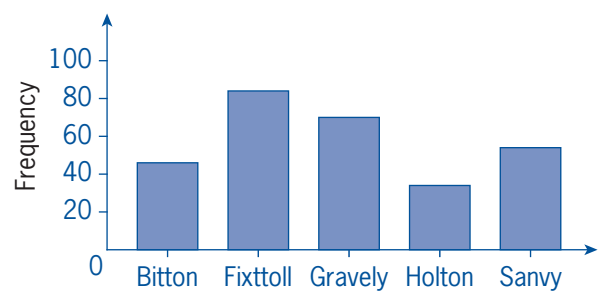
- 2 The bar chart shows the number of books in a school library for some subjects.
- How many more books are there for physics than DT?
 - How many books are there altogether?
 - Which subject is the mode?



- 3 The results of a poll in June 2010 on voting intentions are given in this pie chart.
- What angle is represented by 'Others'?
 - If 1000 people voted according to the results of this poll, how many would vote Labour?



- 4 Avenue School students live in five local villages. This bar chart shows the number of students attending the school from each village in January 2010.
- How many students came from Sanvy village?
 - How many students are there at Avenue School?
 - Draw a pie chart to show the same information as the bar chart.



KEY WORDS:

frequency diagram
frequency polygon
grouped
histogram

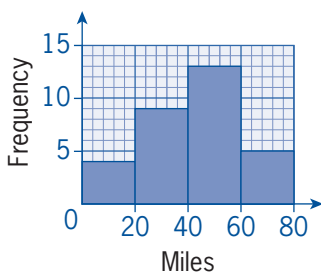
- Equal interval histograms and frequency polygons are both **frequency diagrams** that are used to display **grouped** continuous data.
- An equal interval **histogram** has bars to represent the frequencies.
 - There are no gaps between the bars.
 - Both axes have a scale.

EXAMPLE

The table summarises the number of miles Reuben cycled each day in July.

| Miles cycled (m) | $0 \leq m < 20$ | $20 \leq m < 40$ | $40 \leq m < 60$ | $60 \leq m < 80$ |
|----------------------|-----------------|------------------|------------------|------------------|
| Frequency | 4 | 9 | 13 | 5 |

Draw a histogram for these data.



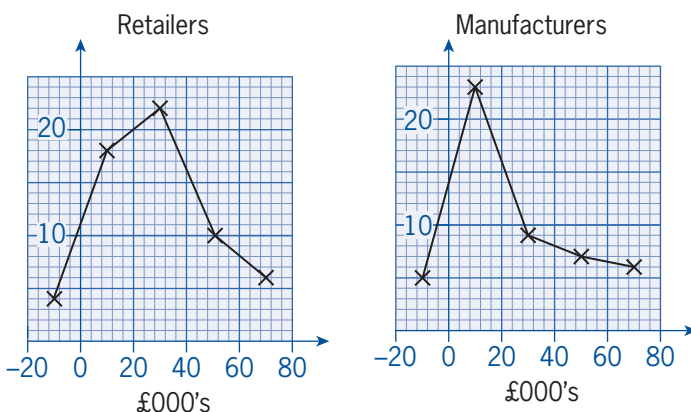
- A **frequency polygon** has straight lines joining the midpoints of each class interval.

You can use a histogram to draw a frequency polygon. Just join the midpoints of the tops of each bar.

When you use graphs to compare data sets, you should always make comparisons between the same type of average (compare means or medians or modes) and the range.

EXAMPLE

Compare the profits (or losses) for these samples of retailers and manufacturers.



Don't be tempted to compare individual values such as the last class of both graphs.

Don't be tempted to compare the height of the peak, compare the classes of the highest peak from each graph.

On average retailers made more profit as their modal class (£20 000-£40 000) is higher than the modal class (£0-£20 000) for manufacturers.

Exercise S5

- 1 These are the times, in minutes, that a group of students spent using a computer one evening.

| Time, m minutes | $0 < m \leq 20$ | $20 < m \leq 40$ | $40 < m \leq 60$ | $60 < m \leq 80$ | $80 < m \leq 100$ | $100 < m \leq 120$ |
|-------------------|-----------------|------------------|------------------|------------------|-------------------|--------------------|
| Frequency | 5 | 11 | 17 | 15 | 13 | 8 |

- Draw a histogram to display these data.
 - Draw a frequency polygon to display these data.
 - What is the modal class interval?
- 2 Todd measured the lengths of leaves, in mm, dropped from trees in his garden one day in autumn.

44 85 72 42 37 29 78 43 79 91 43
 45 28 42 79 34 92 87 41 43 43 78
 82 47 85 43 92 32 86 76

- a Copy and complete the tally chart for these data.

| Leaf length, l mm | Tally | Frequency |
|---------------------|-------|-----------|
| $20 < l \leq 30$ | | |
| $30 < l \leq 40$ | | |
| $40 < l \leq 50$ | | |
| $50 < l \leq 60$ | | |
| $60 < l \leq 70$ | | |
| $70 < l \leq 80$ | | |
| $80 < l \leq 90$ | | |
| $90 < l \leq 100$ | | |

Remember:
 A tally chart collects data in groups of 5 like this ||||

- Draw a histogram shows the length of the leaves.
 - Explain how the data suggests that there may be more than one type of tree in Todd's garden.
- 3 In a triathlon there are transition times when competitors change sport from swim to bike and then from bike to run. Paolo collected data on the transition times, in seconds, for 80 competitors. These are the transition times for swim to bike.

| Time, s seconds | $120 < s \leq 150$ | $150 < s \leq 180$ | $180 < s \leq 210$ | $210 < s \leq 240$ | $240 < s \leq 270$ |
|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Frequency | 12 | 21 | 26 | 15 | 6 |

- Draw a frequency polygon to represent these data.
- Paolo drew this frequency polygon for the transition times for bike to run.
 Compare the transition times taken for swim to bike and for bike to run

