3.2 Presenting Data Graphically

- One advantage of frequency polygons over bar charts is that several sets of data can be displayed on the same diagram. (Another solution is to use "compound bar charts", with more than one bar for each category.) One way to illustrate this is to collect some data for boys and for girls separately and then present them on the same graph (see section 3.2.1).

3.2.1 Estimate on scrap paper how long in minutes you spent watching TV last night. If you can't remember, give me your best guess. We're counting between leaving school yesterday and going to bed (not breakfast TV this morning).

Choose sensible groups (e.g., 0-29, 30-59, 60-89 min, etc.).
Tally boys and girls separately and draw two separate diagrams.

Possible hypothesis: "On average, girls watch less TV than boys" (or the opposite).

You can do the same thing with amount of time spent on homework. How does it compare with school expectations?!
Is there any correlation between how much TV pupils watch and how long they spend on their homework? This would require a scatter-graph. (The conclusion is usually that everyone watches far too much TV!)

Some may watch as much as 300 min.

3.2.2 What type of diagram would be suitable for what set of data?

Pupils often find this hard because they are rarely asked such questions.

When do we draw pie charts/scatter-graphs, etc.?

A pupil may "like" pie charts but not see that they are appropriate only for proportions – when a set of values adds up to a "whole".

What are the advantages/disadvantages of presenting this data in this or that form?

This may be an on-going discussion, raised whenever we encounter a new type of diagram/data.

3.2.3 **NEED** "A Day in the Life" sheets.

This can make good display work.

A task like this makes drawing pie charts accessible to pupils with no knowledge of angles.

Pupils can fill in the data for homework. Bear in mind that it takes from midnight until the next midnight; i.e., two *nights!*

3.2.4 **NEED** chocolate wrappers or similar with nutritional information.
Draw a pie chart to illustrate the content of common chocolate bars.

To make it harder you can assemble a collection of wrappers onto an A4 sheet, removing the "per 100 g" columns so that pupils have to calculate this first.

(Note that fat and carbohydrate quantities are often subdivided into saturated/unsaturated and sugars/starch, so you have to be careful not to add these subdivisions to the total and end up with more than 100 g.)

Multi-pack bags of different items (e.g., chocolate bars) are the best because they often contain full information on the outside of the outer bag for all the items.

Also be careful not to add the amount of energy in kJ or kcal to the mass in grams!

Which bar is "healthiest"?
Is there much difference between them?
How do their prices compare with what's in them?
If you spend more, do you get a healthier product or just a tastier one, or neither?

The total normally comes to less than 100 g, so you need to include an "other" sector on the pie chart.

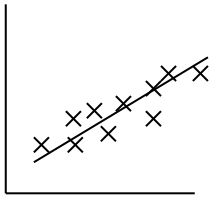
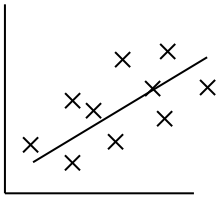
Correlation

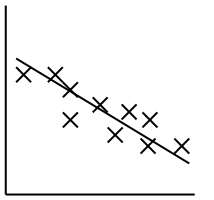
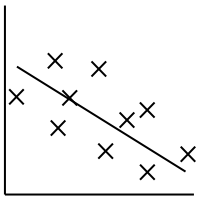
Scatter diagrams are a way of comparing two quantities.

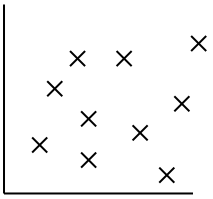
- If large values of one tend to go with large values of the other (and small with small) then we have **positive correlation**. The **line of best fit** is the best straight line you can draw through the points. If the points mostly lie close to the line, we say it is ***strong*** positive correlation. Otherwise it is ***weak*** positive correlation.

- If large values of one quantity tend to go with small values of the other, then we have **negative correlation**. Depending on how close to the *line of best fit* the points mostly are, we say it is either ***strong*** or ***weak*** negative correlation.

- **No correlation** happens when there seems to be no pattern to the arrangement of the points. The two quantities have nothing to do with each other.

<table>
<tr><td>positive correlation</td><td> 
strong positive correlation</td><td> 
weak positive correlation</td><td>For example,

ice-cream sales against temperature</td></tr>
<tr><td>negative correlation</td><td> 
strong negative correlation</td><td> 
weak negative correlation</td><td>For example,

hot chocolate sales against temperature</td></tr>
<tr><td>no correlation</td><td colspan="2"> 
can’t be weak or strong – the points are just scattered randomly (no line of best fit)</td><td>For example,

newspaper sales against temperature</td></tr>
</table>

Because two quantities correlate, it doesn’t mean that they directly affect one another. It could be that they correlate because they both depend on a third factor that we haven’t thought about.

Presenting Data Graphically

- **Bar charts**
  - ✓ useful for categorical data (qualitative, non-numerical data)
  - ✓ can have horizontal or vertical bars
  - ✓ bars don't touch and must have equal width
  - ✓ *length* of bars is proportional to the frequency

- **Vertical line graphs**
  - ✓ useful for discrete data (data that can take only certain values)
  - ✓ vertical axis labelled *frequency*

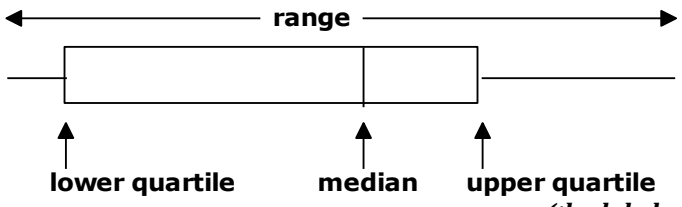
- **Histograms**
  - ✓ useful for grouped continuous data (or sometimes grouped discrete data)
  - ✓ vertical bars which don't necessarily have equal width
  - ✓ bars touch and are labelled on the horizontal axis where they join
  - ✓ *area* of bars is proportional to the frequency
  - ✓ vertical axis labelled *frequency density* ($= \dfrac{\text{frequency}}{\text{width}}$)

- **Pie charts**
  - ✓ area (and angle) of each sector is proportional to the frequency
  - ✓ total area is proportional to the total frequency (when comparing different pie charts)

- **Cumulative frequency graphs**
  - ✓ useful when you have a total frequency n more than about 30
  - ✓ vertical axis is labelled *cumulative frequency*
  - ✓ values are plotted against the *upper boundary* of each class
  - ✓ points are joined by a smooth curve
  - ✓ the **median (Q_2)** is estimated by reading off the $\frac{1}{2}n^{th}$ term
  - ✓ the **upper quartile (Q_3)** is estimated from the $\frac{3}{4}n^{th}$ term
  - ✓ the **lower quartile (Q_1)** is estimated from the $\frac{1}{4}n^{th}$ term
  - ✓ the **inter-quartile range (IQR)** is $\mathbf{Q_3 - Q_1}$
    (any data further than 1.5 × IQR beyond the upper or lower quartiles are generally regarded as **outliers**)

- **Median and upper/lower quartiles**
  - ✓ For a smaller set of data with n items,
    median = $\frac{1}{2}(n+1)^{th}$ value, and if $\frac{1}{2}(n+1)$ is an integer m then
    upper quartile = $\left[m + \frac{1}{2}(m+1)\right]^{th}$ value and **lower quartile** = $\frac{1}{2}(m+1)^{th}$ value.
    If m is not an integer, round it *down* and then use these formulas.

- **Box and whisker plot (box plot)**
  - ✓ useful when drawn underneath the horizontal scale of a cumulative frequency graph or separately with its own scale
  - ✓ shows Q_1, Q_2, Q_3 and the maximum and minimum values



(the labels are not normally needed)

A Day in the Life of _____

time	main activity
0000 – 0100	
0100 – 0200	
0200 – 0300	
0300 – 0400	
0400 – 0500	
0500 – 0600	
0600 – 0700	
0700 - 0800	
0800 – 0900	
0900 – 1000	
1000 – 1100	
1100 – 1200	
1200 – 1300	
1300 – 1400	
1400 – 1500	
1500 – 1600	
1600 – 1700	
1700 – 1800	
1800 – 1900	
1900 – 2000	
2000 – 2100	
2100 – 2200	
2200 – 2300	
2300 – 0000	

In the main activity column, write down the main thing you did during that 1 hour period;
e.g., “sleep”, “watch TV”, “eat lunch”, etc.

Then present your activities in the pie chart below. Put the same activity all together, even if you did it at more than one time.

Either write the name of the activity in each sector, or make a key in the space below.

A Pie Chart to Show How I Spent 24 Hours