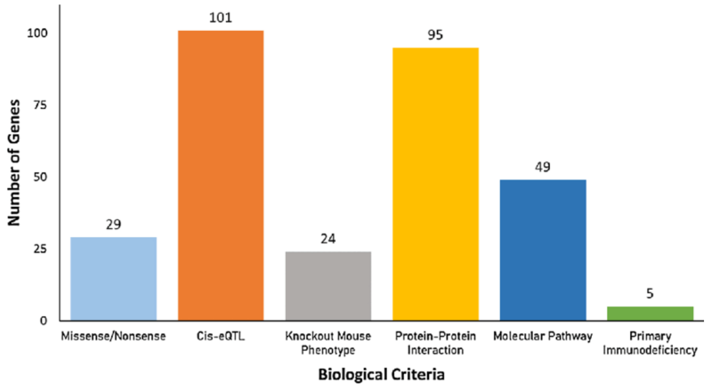
UPCSE Biology – On graphs

Introduction

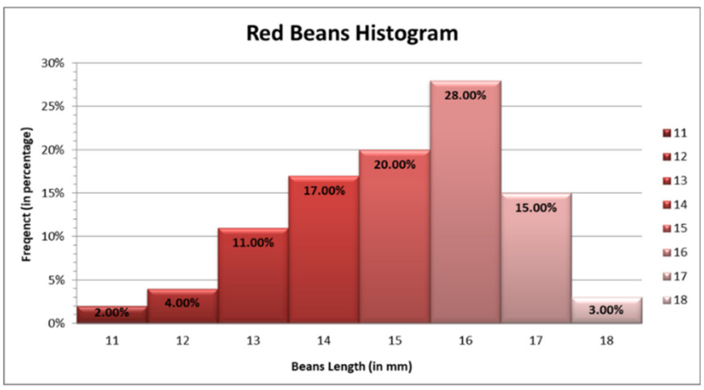
There are three broad categories of visualisations, as shown below.

- *comparison visuals* such as bar charts. These are designed to compare different categories (such as colour, blood groups, number of cilia in kidney tubules, …). We went through this in the last lesson.

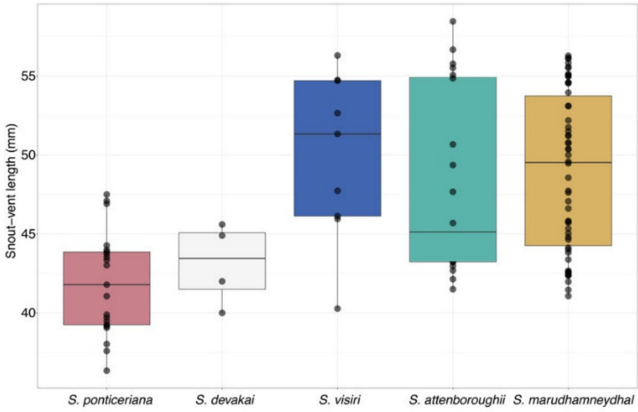


- *distribution visuals* such as box plots and histograms. These are designed to show the pattern of how continuous data is distributed. We went through histograms in the last lesson and will return to this in term 2. We won't be dealing with box plots but I have included an example here just to show you that there exists another way of visualising distribution of data.

  Such type of data could be age range, weight, salary range, etc. Histograms are used when we are dealing with the mean of data, box plots are used when we are dealing with the median of data.

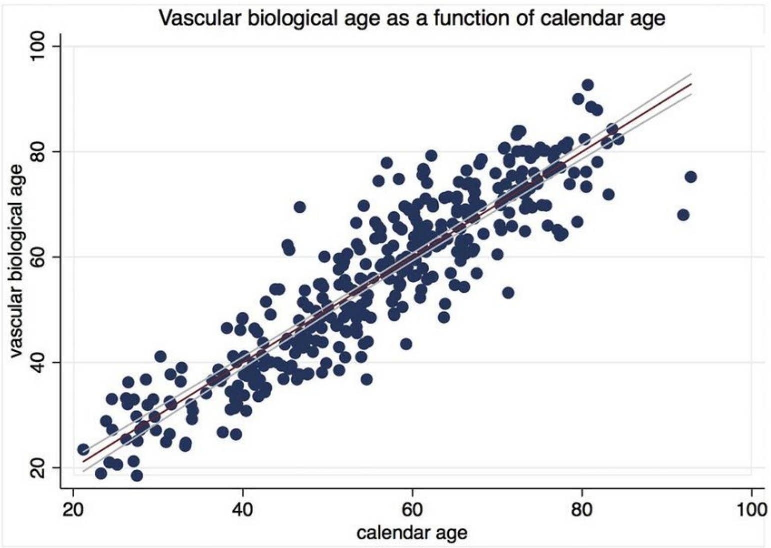


Histogram showing the percentage of beans of different length



Boxplots showing body size distributions in the five currently recognized species of fan-throated lizards in the S. ponticeriana complex

- *relationship visuals* such as *x-y* graphs of functions or scatter plots. We will go through this in today's lesson. These are designed to show the relationship between two variables *x* and *y* (such as time, length, weight, tumour volume, etc., anything which is measured continuously).



On drawing an *x*-*y* scatter graph

In these notes we will look at how to draw *x*-*y* graph, sometimes misleadingly known as line graphs (*x*-*y* graphs are used to plot curves as well as lines). We will go through an example of how to draw a graph by hand using data you already have from a biology experiment, or using data which will be provided to you. We can then use the checklist on the next page to see if we have a good, clean graph.

But before we draw the final graph it is worthwhile drawing a sketch first. A sketch is not a graph. A sketch is a rough draft of graph. We draw sketches to get a sense of the size and scale the final graph will have. From this we can then draw an accurate, precise version which is our final graph.

Before we draw a graph it is always good to draw an initial draft in order to see the scale of things:

- draw your axes with a ruler (not freehand);
- Put the *y*-axis on the left-hand side of the graph paper. Decide where to put the *x*-axis: do you want to put it at the bottom of the graph paper, or in the middle of the graph paper? This will depend on whether or not your have any negative *y* data;
- Mark the maximum *x* value at the right end of the *x*-axis, the maximum *y* value at the top of the *y*-axis, and the minimum *y* value at the bottom of the *y*-axis;
- Place tick marks at equal intervals along each axis;
- Now plot your data as best as possible;
- Identify the trend, and draw a line through your data.
- **Do not connect the dots:** Do not draw lines connecting your data points. The only line to draw will be your best-fit line.

Once you have done this you will get a sense of the scale your graph will need, as well as the position the data within the graph.

You can now move onto drawn the precise version of the sketch, namely the graph itself. The detailed elements to drawing a good, clear, clean graph are shown in the checklist box on the next page. Check your graph against the items in this box.

Checklist for drawing a graph

It is extremely important that your graph be clean (not dirty, no eraser marks, no typex marks, etc.), well laid out (take at least ½ of an A4 page if not the whole page), with axes and best fit line which are straight (use a ruler), etc.

Criteria	**Done? (Yes/No)**
My graph takes up at least half of the A4 page	
My axes are drawn straight (using a ruler)	
My axes are perpendicular to each other,	
My x- and y-axes are in their correct locations,	
My x and y axes are slightly longer than my max x and y values (you don't want to plot your max data point at the top right corner of the graph. Always leave some space for visual clarity, and for practical reasons in case you need to extrapolate values during any data analysis)	
I have major tick marks, equally spaced from 0 to the max x value and from 0 to the max y value (this is for linear scales only. For a log scale we do something different. Also, note that the spacing of your divisions on the *x*-axis does not need to be the same as the spacing of your divisions on the *y*-axis).	
I have numbered my major tick marks	
I have decided that I do/don't need minor tick marks. It is preferrable not to number the minor tick marks (minor tick marks give greater accuracy when reading the graphs)	
I have labelled my axes, and included units.	
I have accurately plotted each data point.	
I have given my graph a descriptive title.	

On drawing a line of best fit through your data and interpolating results

On drawing the line of best fit

Draw a line of best fit through your data. See the checklist on p5 as a way of checking the quality of your best-fit line. But the general summary of the check lit is

- Draw your line with a ruler;
- Make sure that your line ends up with as many data points above it as below it. There is a caveat to this. See the next bullet point). We want (roughly) an equal number of data points above the line as below it (there is mathematical theory behind finding the correct line of best fit, but we won't go into this in our maths-for-biology course);
- *Caveat*: Only draw your best-fit-line through the major trend of the data. If you have any outliers (data which lies far away from the main trend) do not take these into account when placing your line of best fit.

On interpolating results using the line of best fit

Here we use the graph itself to estimate the values of unknown data points (i.e. data values we were not able to collect during the experiment). *Interpolation* is based on choosing an unknown data points *within the range* of data we have collected. As such it allows us to estimate what our y-value could have been (within reasonable error) if we had actually collected the x data value during our experiment.

To perform interpolation:

1) Simply locate (on the x-axis) an x-value you want to find the answer to;
2) From your x value go up to your line of best fit and then across to the y-axis;
3) Read of the y value. This will be your approximate answer your x data value would have given you.

Extrapolation is the opposite of interpolation in the sense that we estimate the values of unknown data points that lie *outside the range* of data we have collected. Extrapolation is less recommended since we have no idea how the best fit line will continue outside our data range. Collecting more data (i.e. data beyond the range we have already collected) will most likely change the height and slope of the line.

Checklist for drawing line of best fit, finding its equation, dealing with outliers, and interpolating results

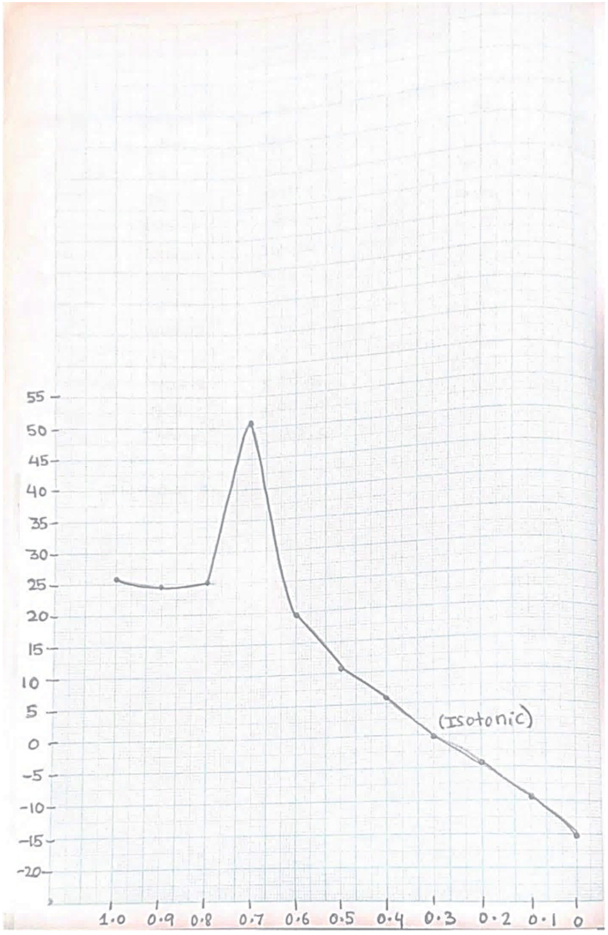
Criteria for drawing a line of best fit	Done? (Yes/No)
Equal number of data points above and below: Try to put as many data points above the line as below it. This means that your line is the "average" line through all the points. **On outliers** ○ Visually identify any outliers. These are data values/points which lie well beyond the main trend (i.e. the region where nearly all your point congregate). ○ "Reading" outliers: Do not automatically ignore outliers. Further investigation is required. They can be a sign of a human error or a sign of a biological anomaly or hidden effect: – Human error: Did you record the data correctly? Is your data collection method appropriate for what you are doing? – Biological anomaly/effect: I can't comment on this. Speak to Karen or Alice.	
Calculate the gradient of your line: If you are finding the gradient directly from the graph then choose the data points to be far apart. This increases the chance of you measuring your gradient more accurately. The gradient m is found as follows: for two data points (x_1, y_1) and (x_2, y_2) we have $m = (y_2 - y_1)/(x_2 - x_1)$. The steepness of best fit line = the rate of change of x with respect to y. A steep line = things changing fast; a shallow line = things changing slowly. What is the biological meaning of the gradient of your line?	
Find the equation of the line: The equation of the line is given by $y - y_1 = m(x - x_1)$. Then, rewrite your equation in the form $y = mx + c$.	

Examples of good and bad graphs

The following are two examples of hand drawn graphs from previous UPCSE biology students, and relate to an experiment on osmosis. You should look at these graphs only in terms of the quality of their presentation.

An example of a poor graph:

Why is this a poor graph? Compare with the checklist.



An example of a good graph

Why is this a good graph? Compare with the checklist. Note that there are still some problems with this graph. Can you find them?