UPCSE Biology –
On reading visuals – part 1

Data analysis: Reading tables, charts and graphs

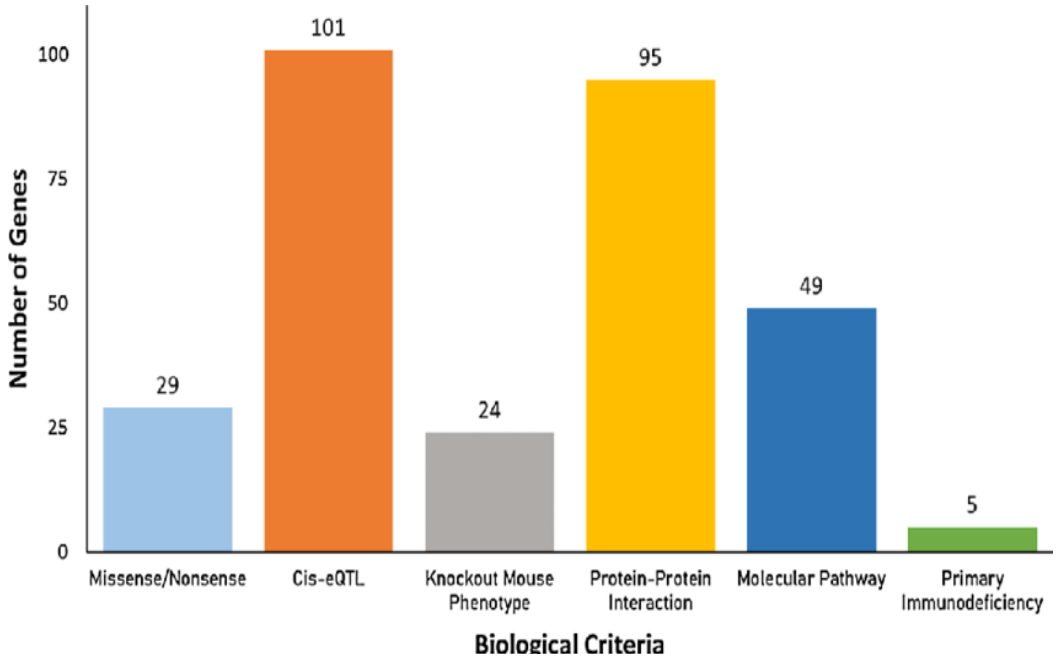
Introduction

- In these slides we look at how to read and analyse different types of data visualisation.
- As a reminder of the lesson 3 notes we have:

  – *comparison visuals*

  such as bar charts.

  Bar charts are used to display data grouped into different categories, such as the biological criteria of the horizontal axis.

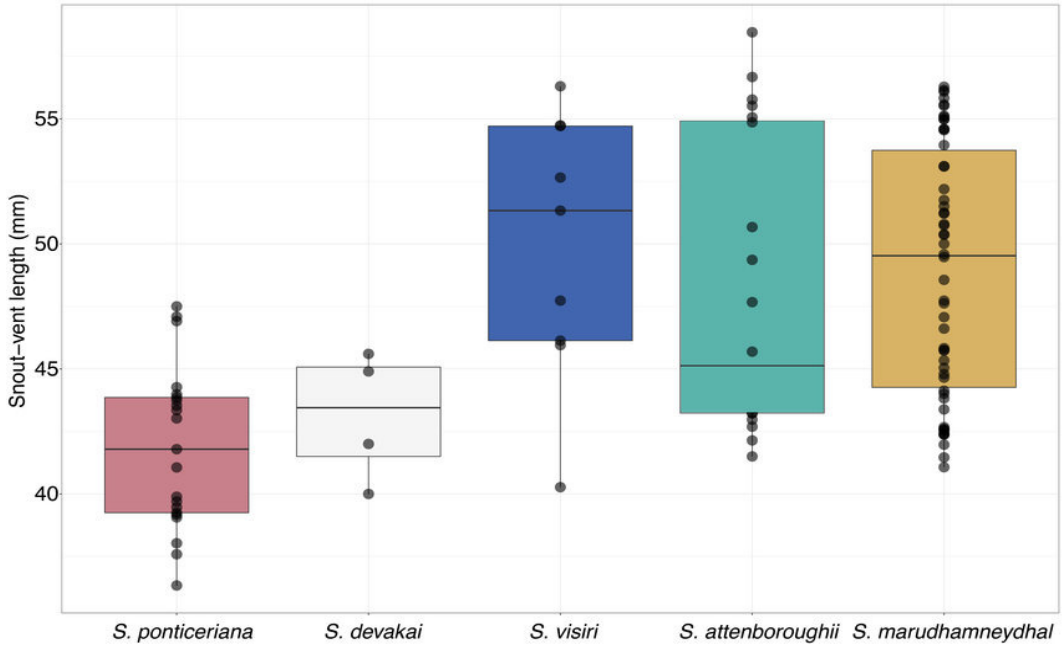


Introduction

- In these slides we look at how to read and analyse different types of data visualisation.
- As a reminder of the lesson 3 notes we have:

  – *distribution visuals* such as box plots and histograms.

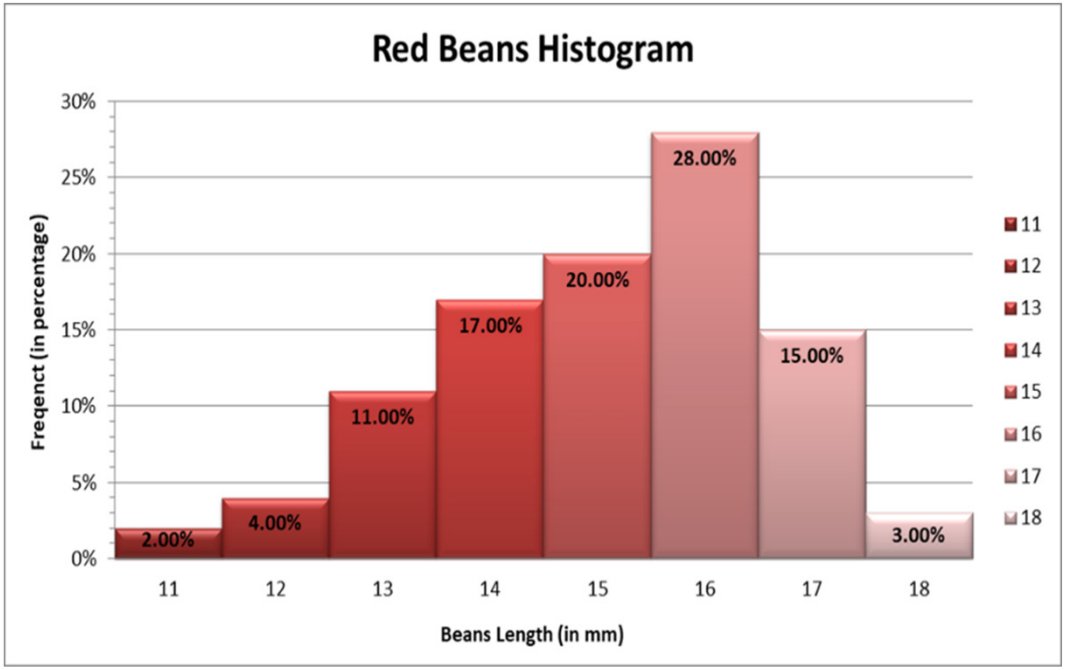
  These are used as bar charts for *continuous numerical data*. They are used to show distribution of data.



Introduction

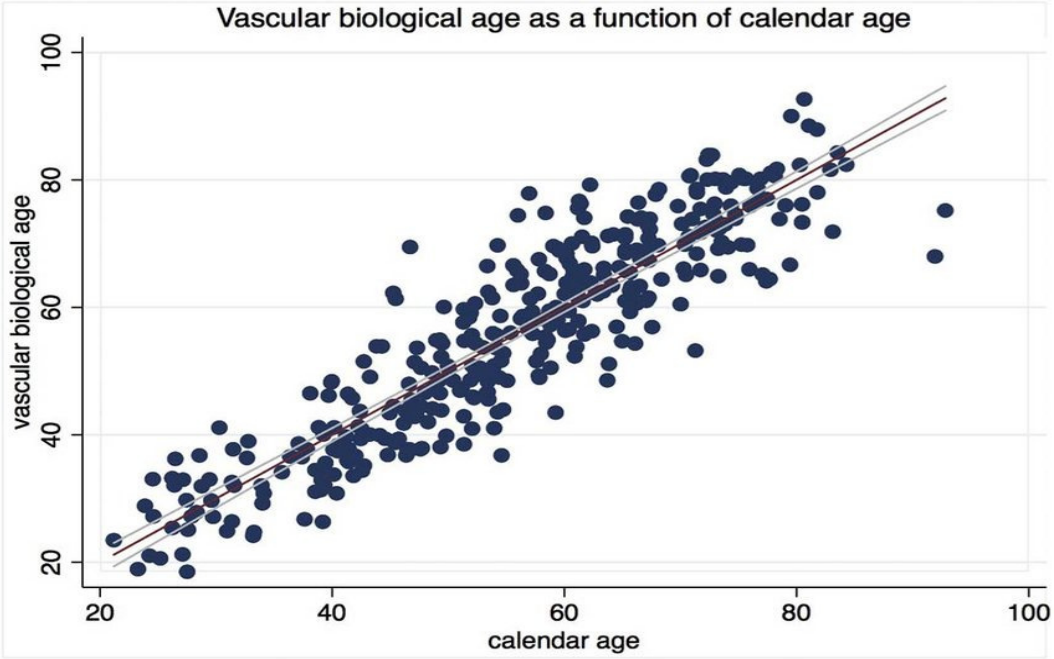
- In these slides we look at how to read and analyse different types of data visualisation.
- As a reminder of the lesson 3 notes we have:
  - *distribution visuals* such as box plots and histograms.

    These are used as bar charts for *continuous numerical data*. They are used to show distribution of data.



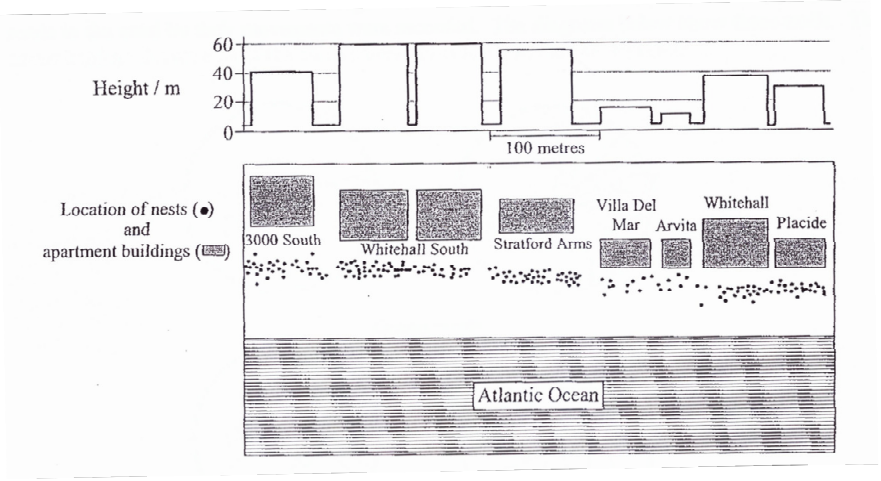
Introduction

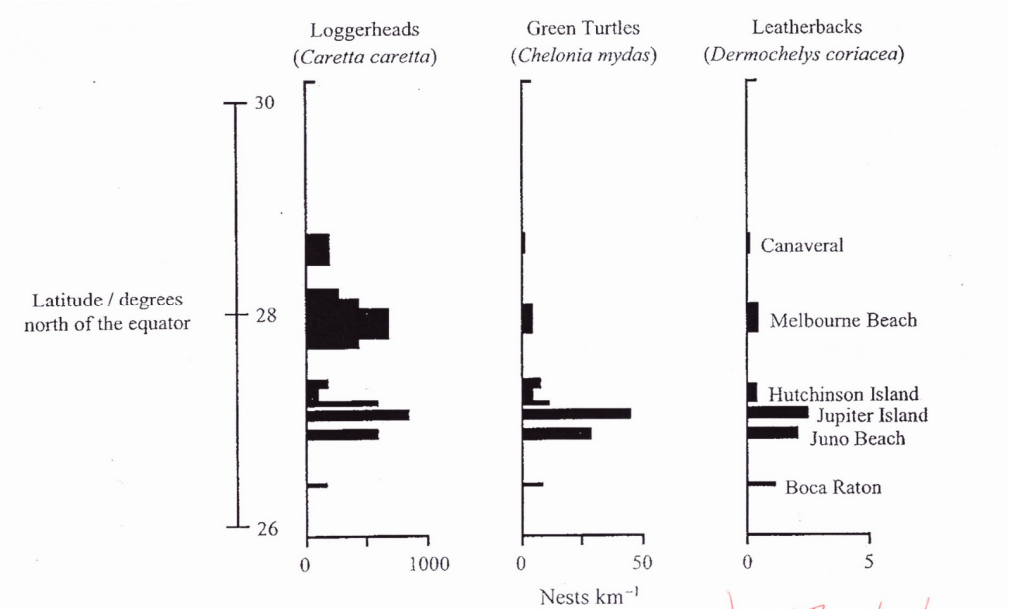
- In these slides we look at how to read and analyse different types of data visualisation.
- As a reminder of the lesson 3 notes we have:
  - *relationship visuals* such as graphs of functions or scatter plots. These are used to represent the relationship between a control variable *x* and an effect variable *y*.



Introduction

- There are many other ways to visualise information, two of which are shown below.





- We will go through some of these next week.

Reading tables

- **Example 1:** Consider the table below (adapted from C.J. Clegg, *Biology for the IB diploma: second edition*):

Experiment	Distance between points A and B (mm)	Time taken for sucrose to travel from A to B (hours)	Rate of transport (mm/h)
1	510	2.1	
2	650		260
3	480	1.6	
4	710	2.3	
5		1.5	300
	Mean distance =	Mean time =	Mean rate =

Reading tables

1) What type of title would you give this table?
2) What is the mean distance travelled by sucrose?
3) What is the mean time taken for sucrose to travel between A and B?
4) What is the rate of transport of sucrose for experiment 3?

Reading tables

4) For which experiment is the time taken for sucrose to travel from A to B, 2.3 hours?

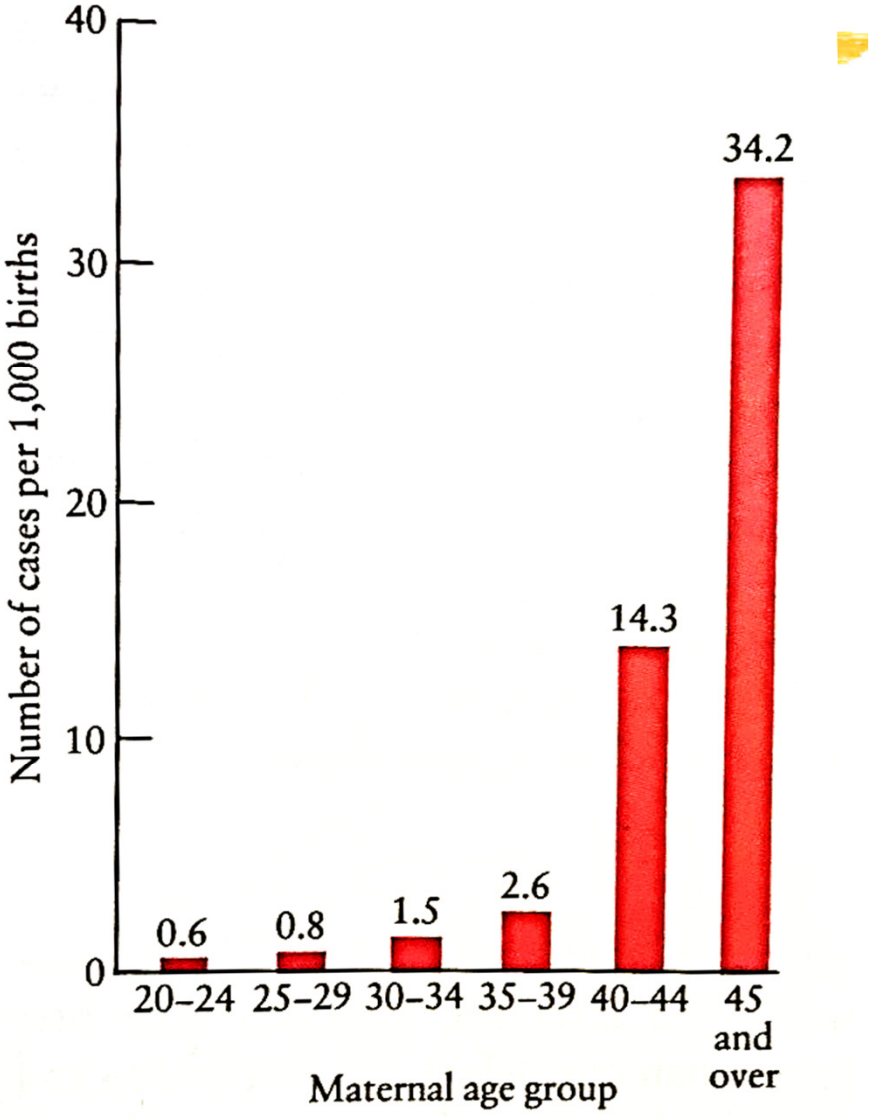
5) If the mean distance between A and B was 500mm, what would the mean time taken for the sucrose to travel have to be for the mean rate of transport to be the same?

Reading bar charts

Example 2

- This chart represents the frequency of births of babies with Down's syndrome in relation to the age of the mother

- What information can you gather from this bar chart? Describe what you see.



Reading bar charts

1) Relatively speaking, there are significantly fewer instances of babies with Down's syndrome for mothers between 20 and 39 compared to mothers between 40 and above.
2) There is a sharp rise in the number of babies with Down's syndrome born to mothers aged 40 or above. One question we could ask is, Why? What factors might possibly cause this to occur?

Reading bar charts

3) The numbers specified above each bar are listed as decimal numbers. This does not mean that we have fractions of babies. These are numbers per 1000 births. So, if there are 10,000 births then numbers would all be integers.

   What we don't know from the chart is the time period over which this data was collected. Is this average data for a 1-year period, a 2-year period a 5-year period?

Reading bar charts

4) There is an issue with the age ranges listed on the horizontal axis. The bars are shown to be separate, and the first two bars show an age range from 20-24 and then 25-29. But age is a continuous variable, so what happened to the data in the age range of 24-25? Did the study miss out this age range?

Reading bar charts

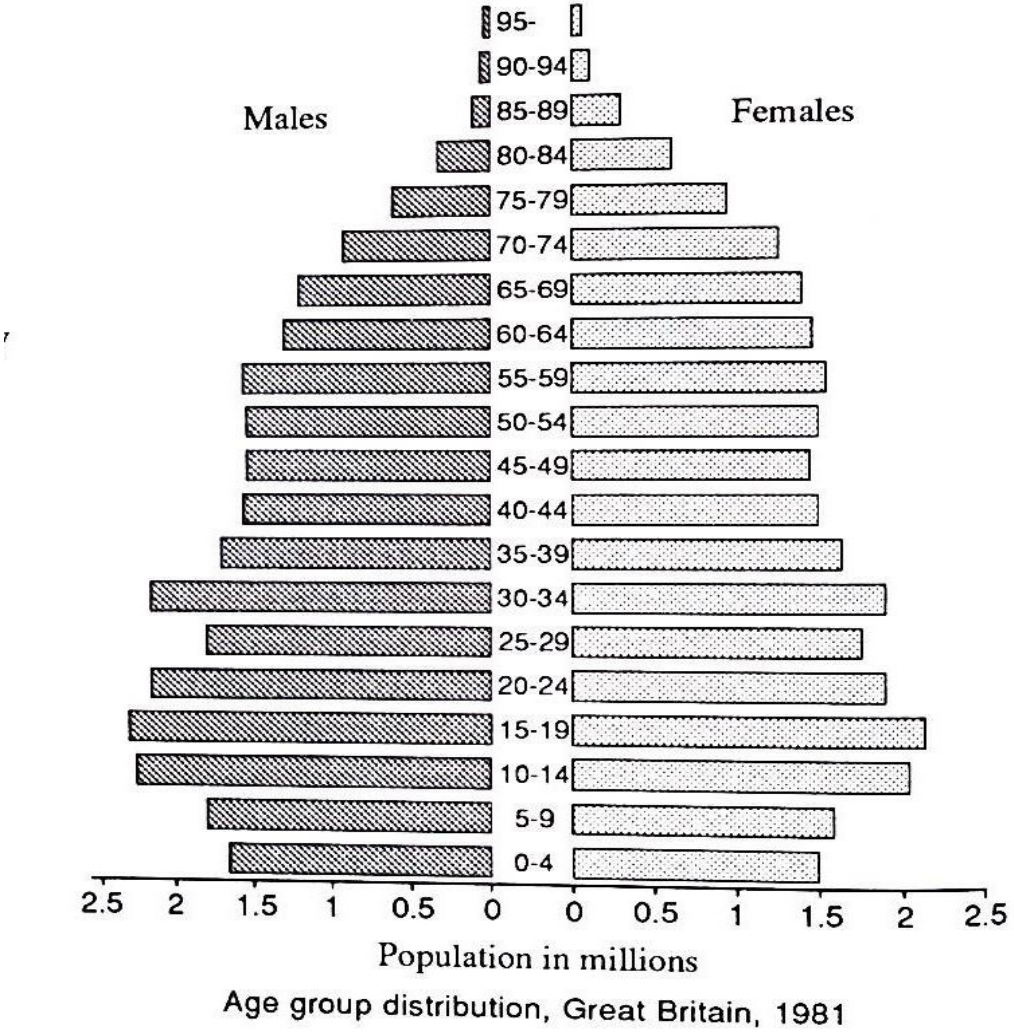
4) Does the first bar stop at 23 years and 365 days? Does the second bar start at 25 years and 0 days? Or does the first bar represent ages rounded down from 24 years, 5 months and 30 days, with ages greater than this rounded up into the second bar?

5) The width of the last bar ("45 and over") is not consistent with the other bars. This may not be problem. Why? Is it relevant to talk about births for women who are 60 years old?

Reading bar charts

Example 3:

Here is another form of bar chart, where the bars are placed side by side.



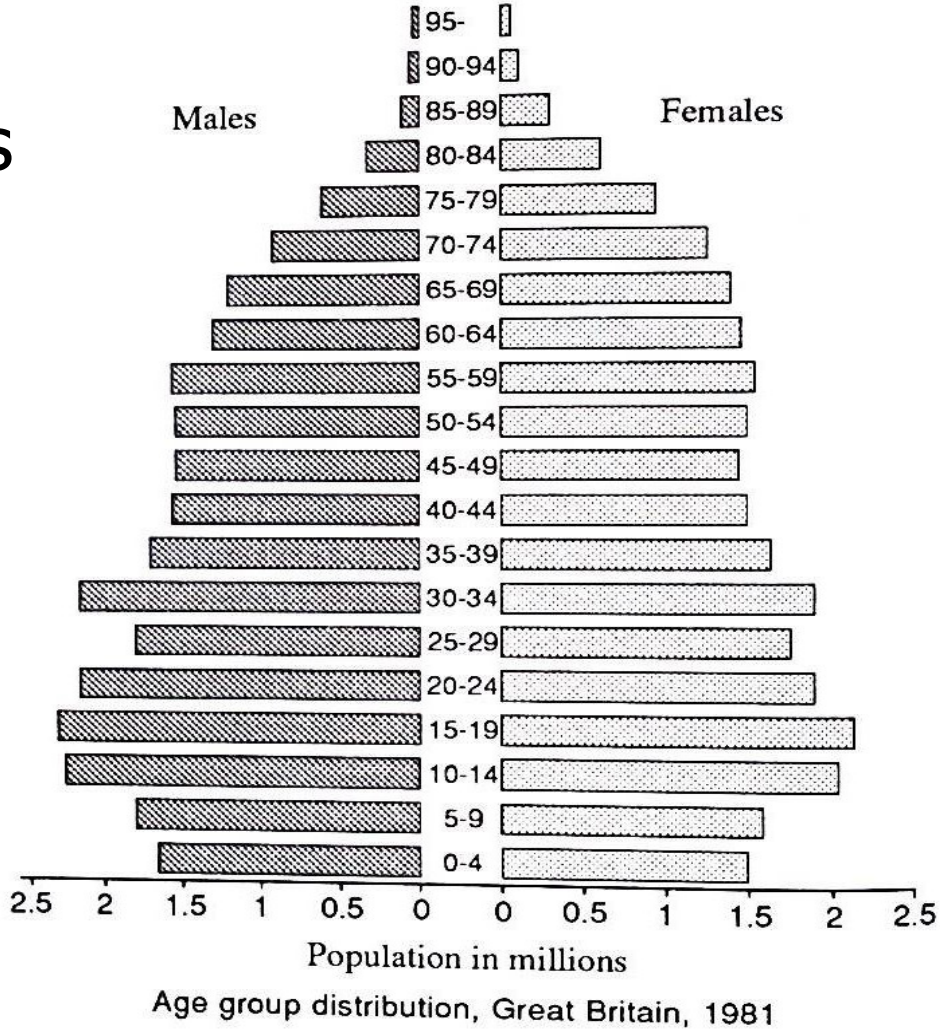
Age group distribution, Great Britain, 1981

Reading bar charts

Example 3:

Q: How many 0-4 year olds are there?

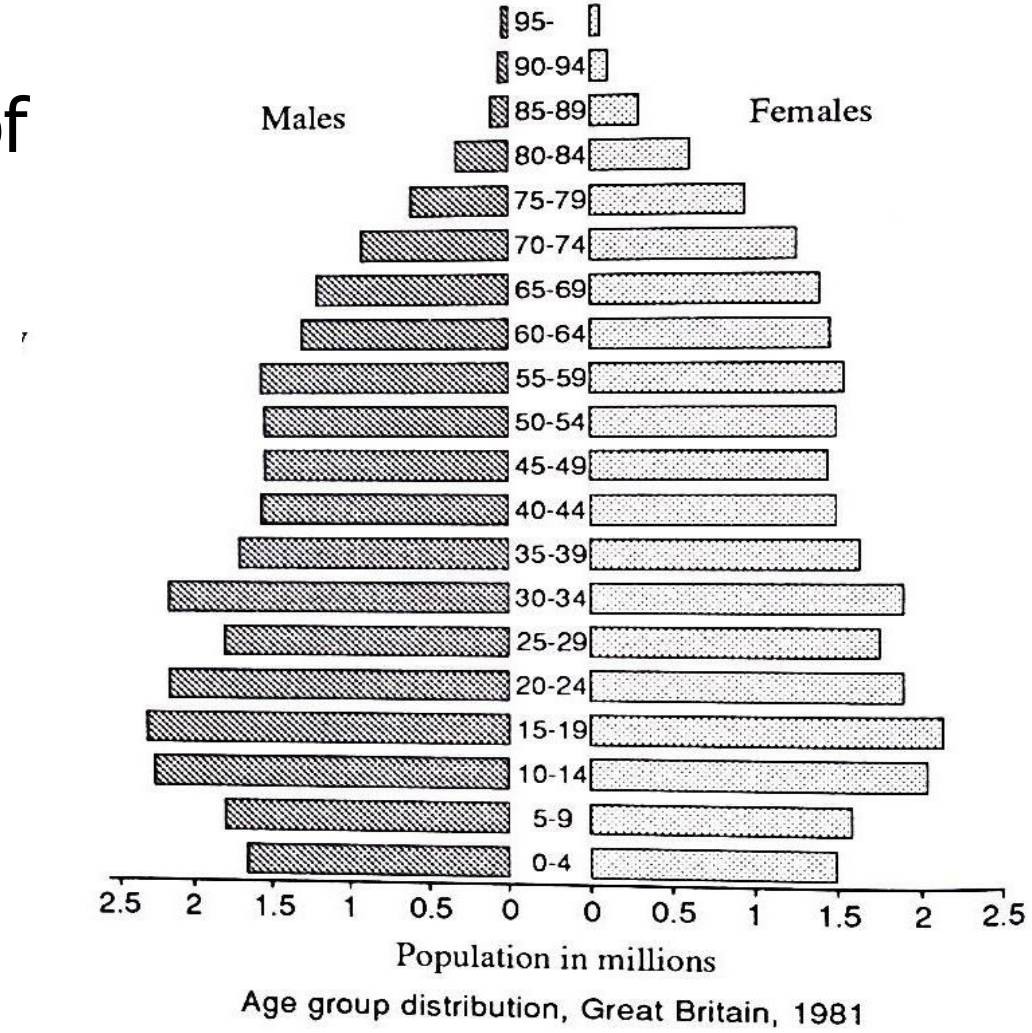
Q: How many women are there over 95?



Age group distribution, Great Britain, 1981

Reading bar charts

Q: How many men are younger than 20 years of age?

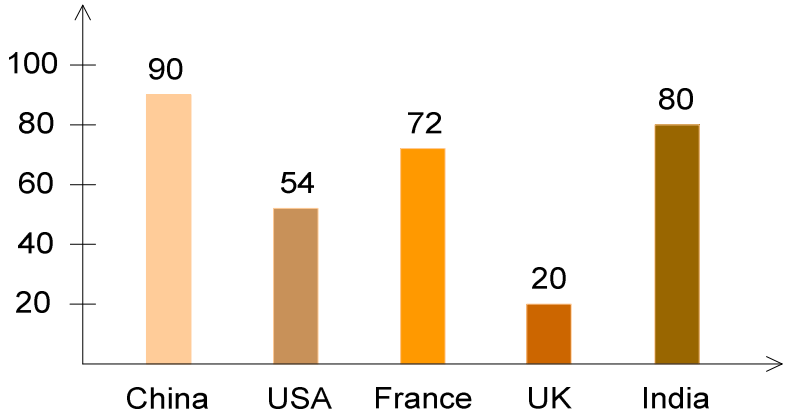


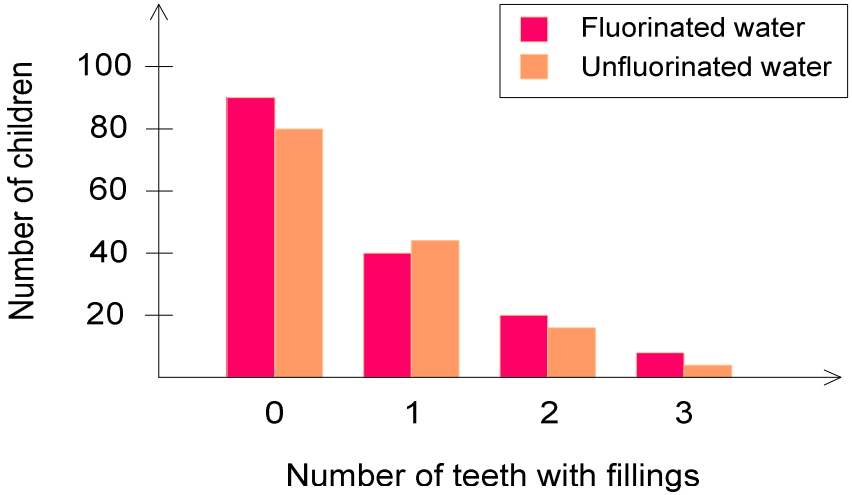
Age group distribution, Great Britain, 1981

Reading bar charts

Example 4:

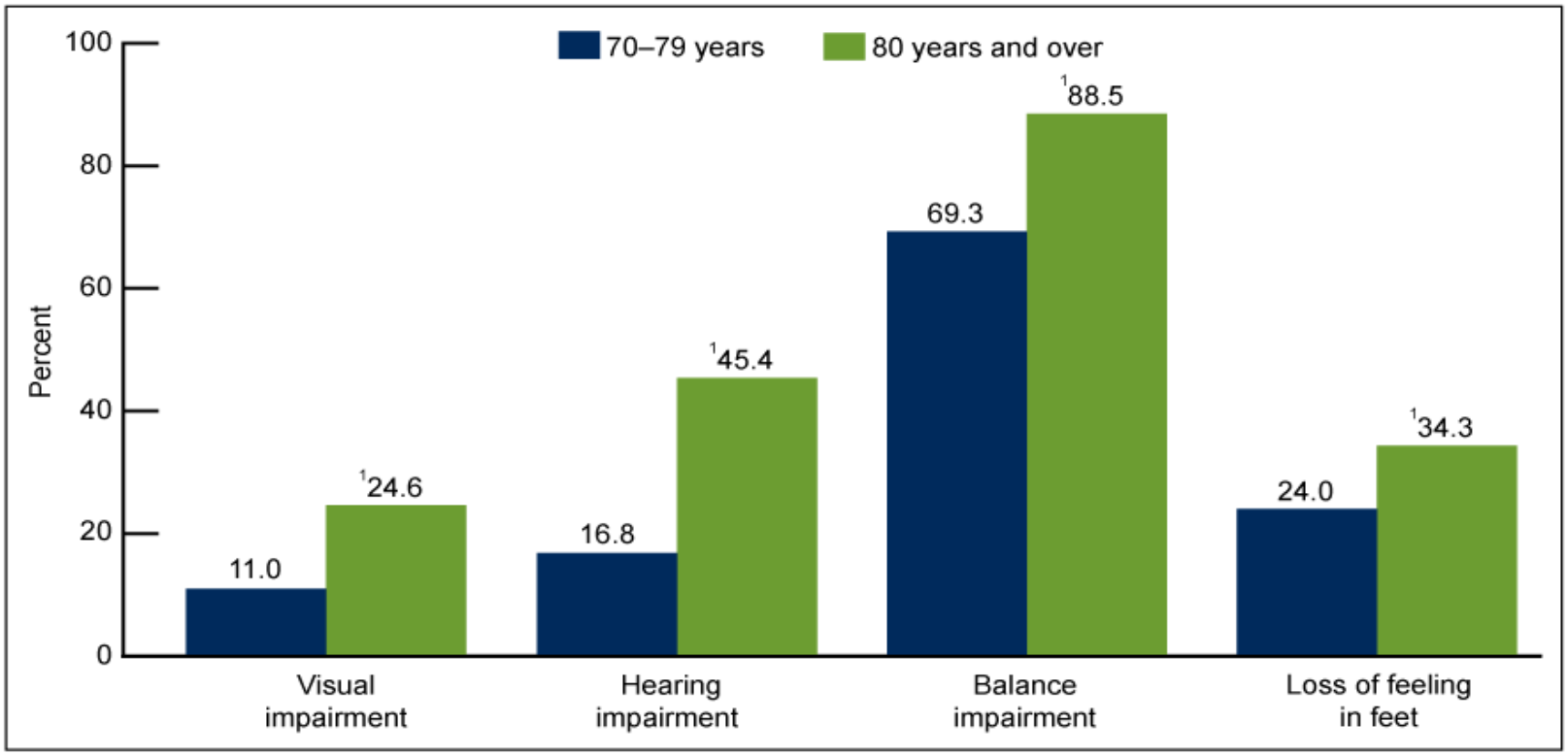
1. What titles would you give the following tables?
2. What information is missing from the charts?
3. What information can you read from these charts? What do the two bar charts tell us?





Reading bar charts

Exercise: The bar chart below represents the occurrence of sensory impairments amongst people aged 70-79 years compared with people aged 80 years or over.



Reading bar charts

- How would you describe the *x*-axis? What information is it describing?

- How would you describe the *y*-axis? What information is it describing?

Reading bar charts

- At what value does the *y*-axis start? Why is it important to know this?

- What do the blue bars represent? What do the green bars represent?

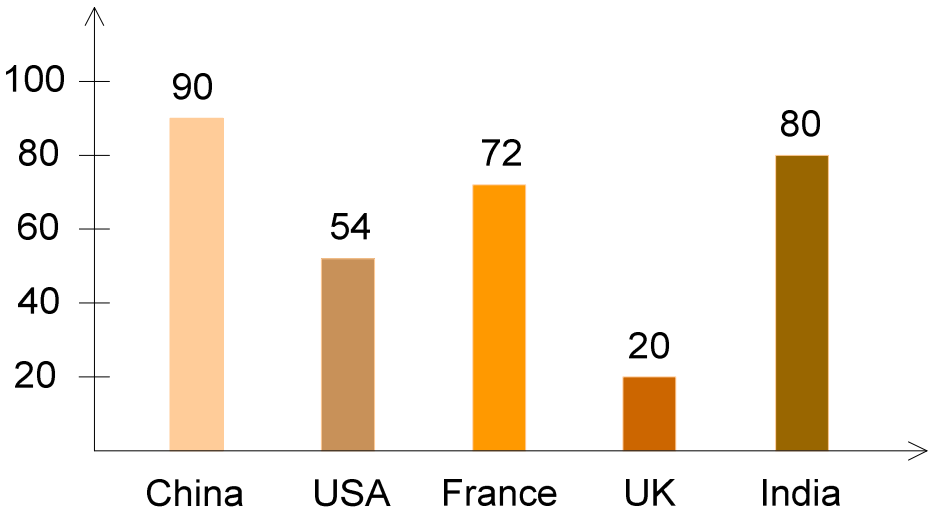
Reading bar charts

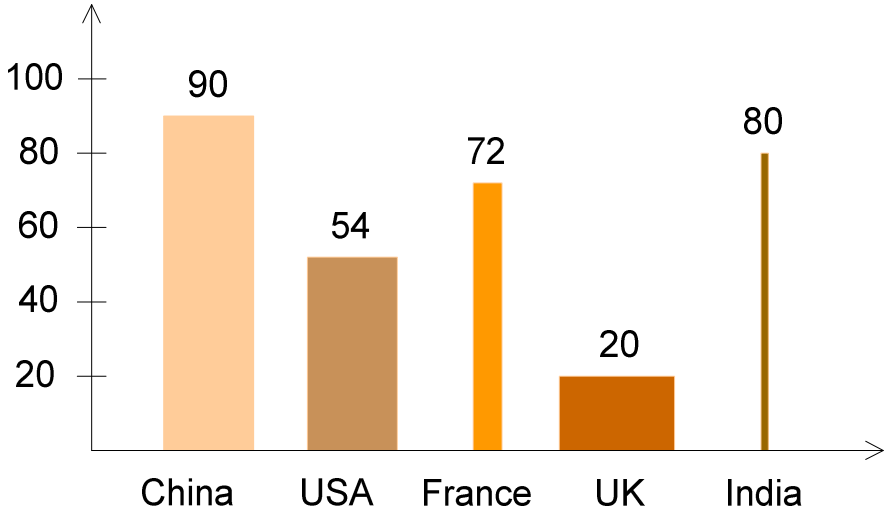
- What percentage of people have visual impairment?

- 34.3% of people have loss of feeling in their feet. What age group does this relate to?

Other aspects of bar charts

- The following two bar charts are the same. Why?



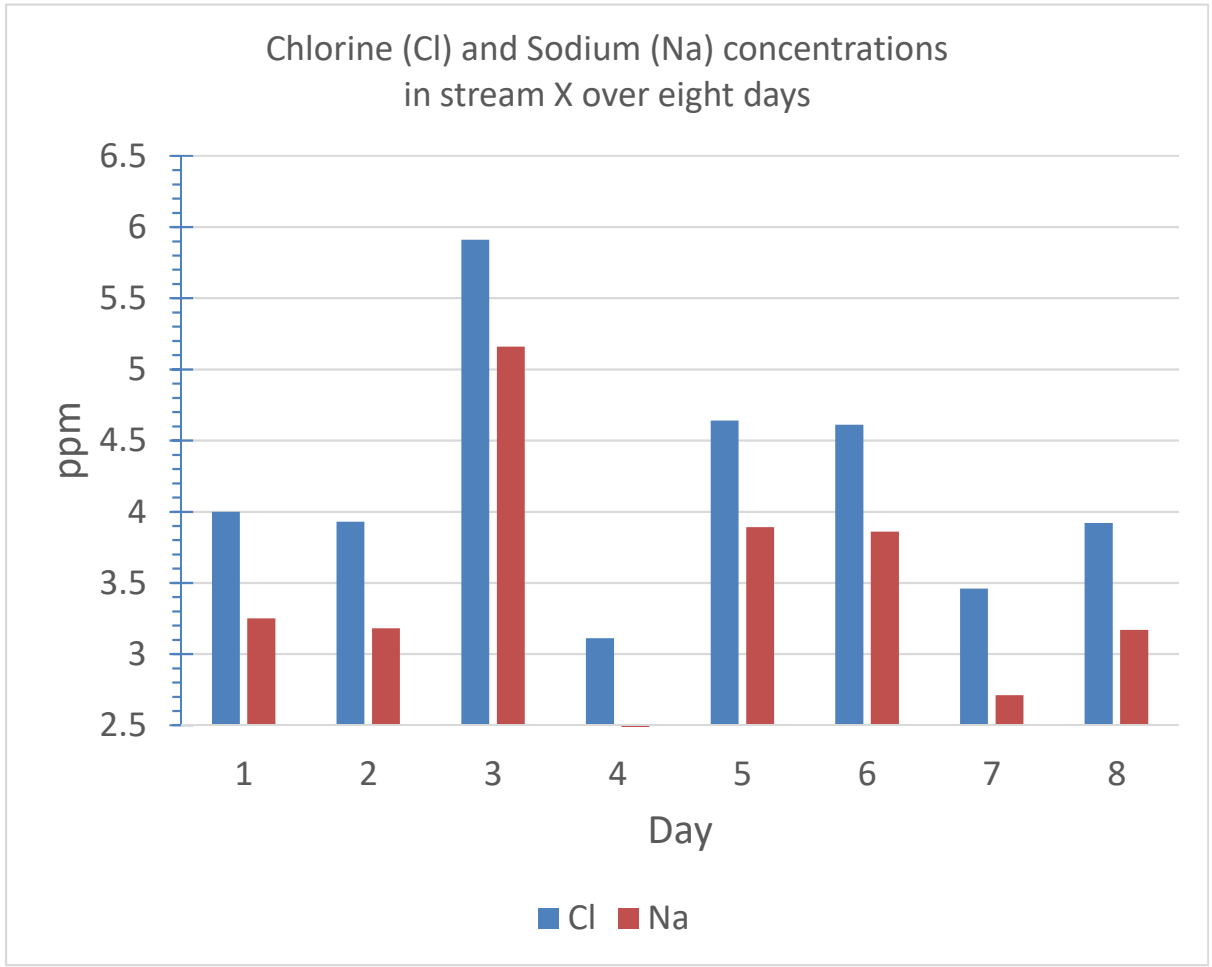


- But we would never draw the bar chart on the right. Why? What makes the left chart better?

Other aspects of bar charts

Issues of scale

Consider the following bar chart relating to Cl and Na concentrations (in ppm) in a stream over eight days:



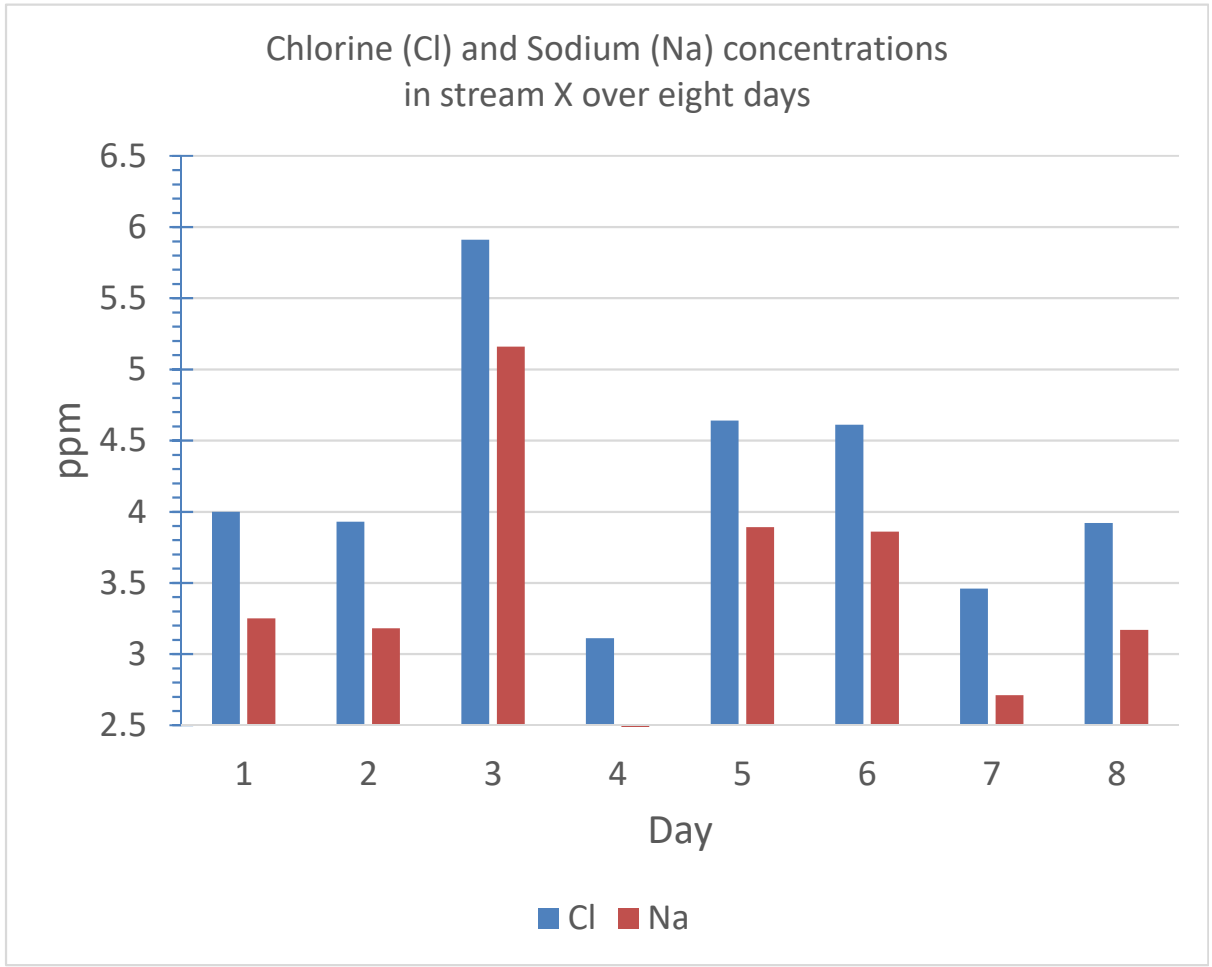
Other aspects of bar charts

Issues of scale

So:

1) No sodium was found in the stream on day 4.

Is this correct?



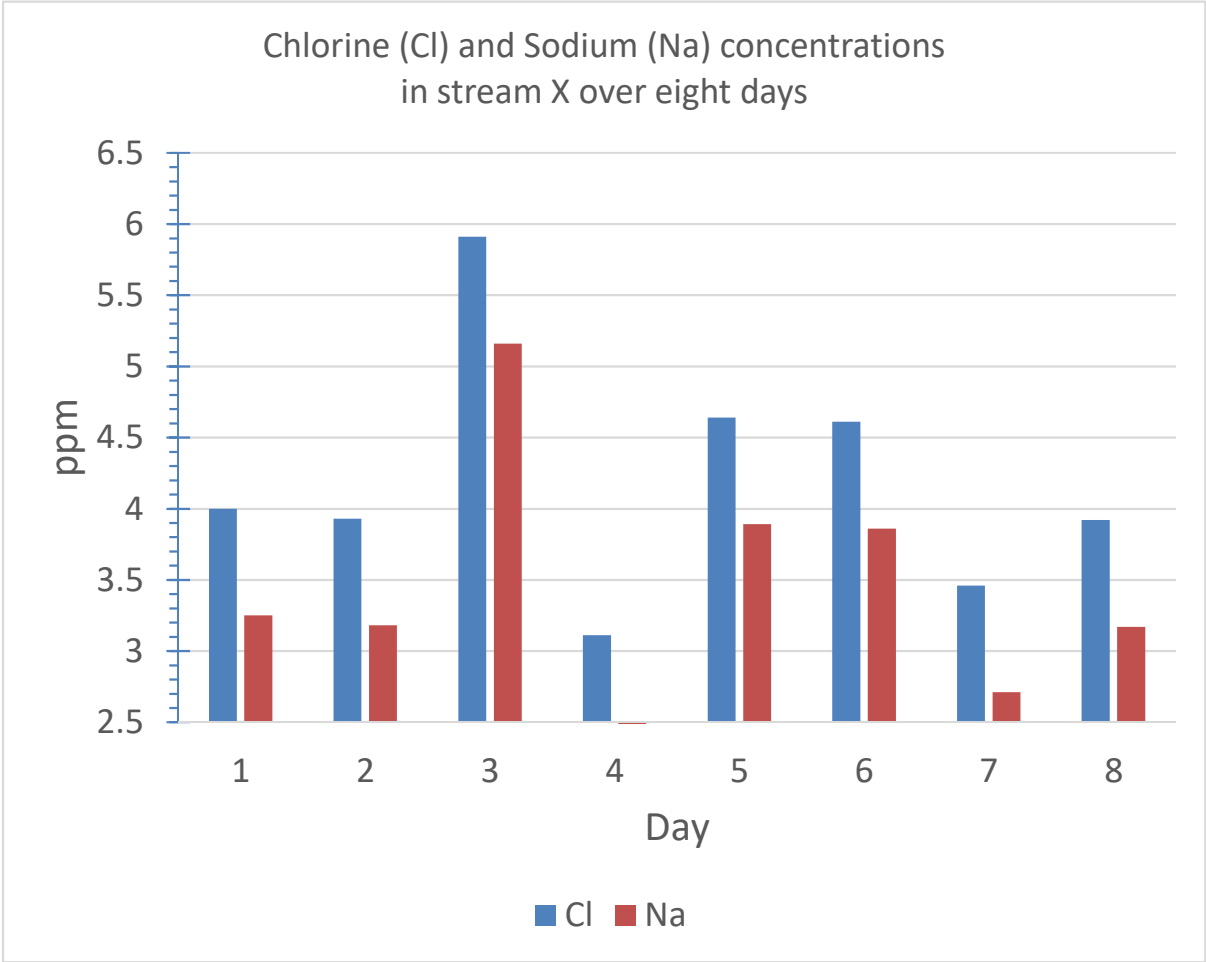
Other aspects of bar charts

Issues of scale

So:

2) Concentration of sodium on days 1, 2 and 8 was approximately half of that of Chlorine.

Is this correct?



Graphs and their presentation

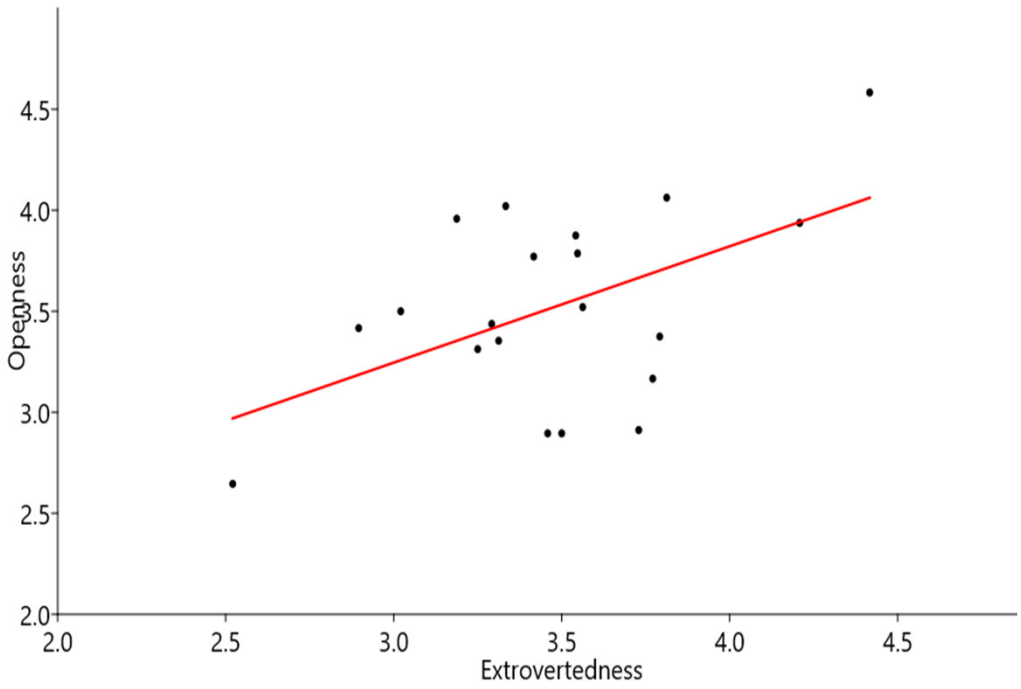
- Last week we went through the topic of line graphs, how to read off information, and how to interpolate data from these graphs.
- These line graphs related to scatter plots and best fit lines.
- We will revise some of this topic now.

Graphs and their presentation

Example 5

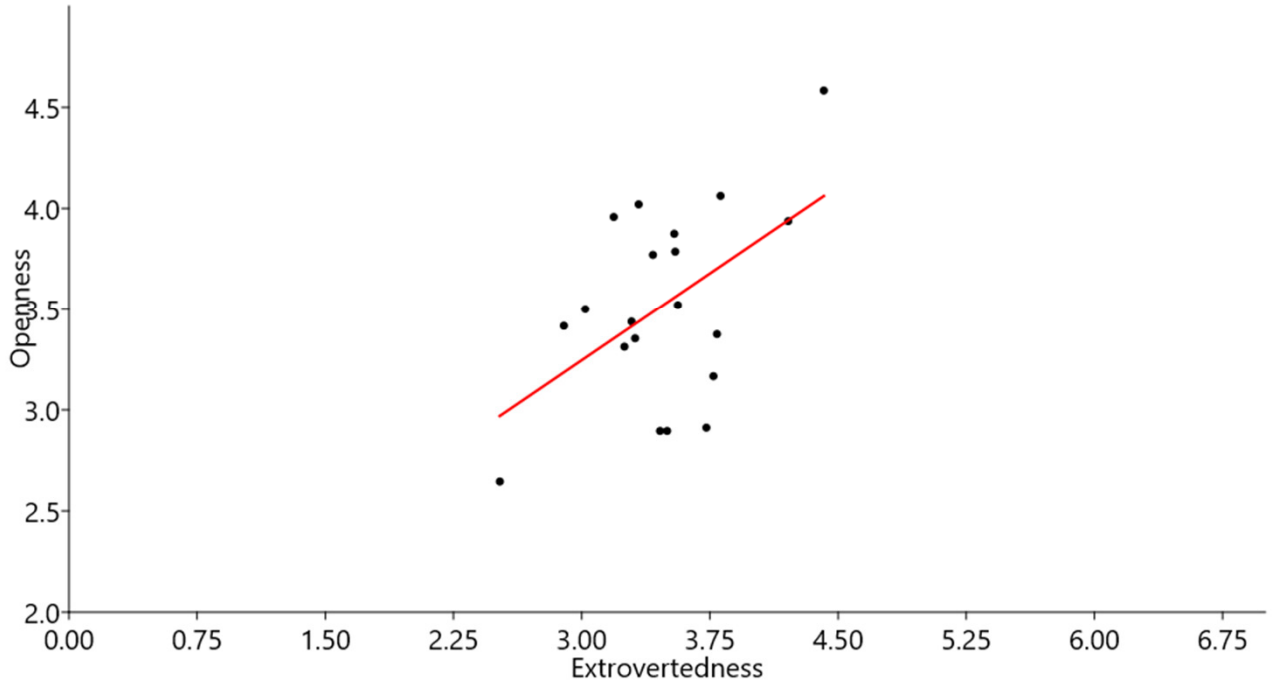
- The following is a graph of people's openness compared to their extrovertedness:

1. What titles would you give the following graph?
2. What information is missing from the graph?
3. What information can you read from this graph?



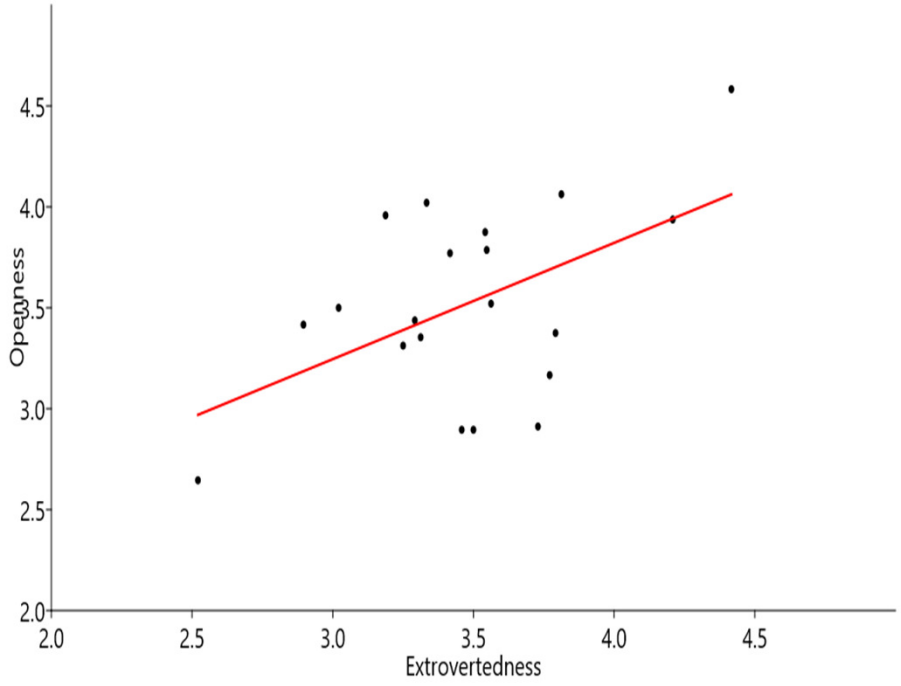
Graphs and their presentation

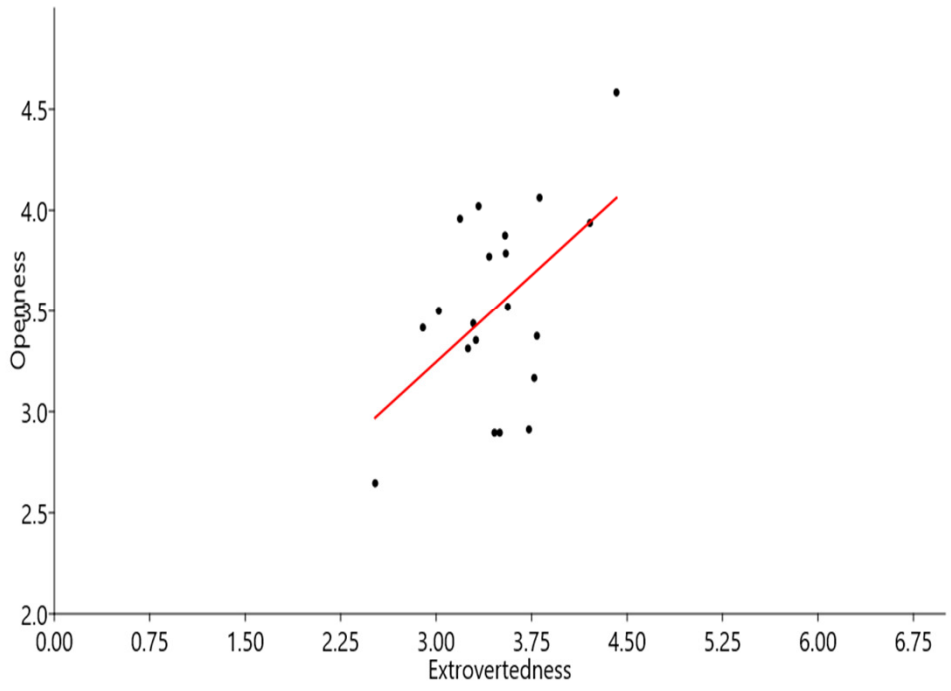
- The following is another graph of people's openness compared to their extrovertedness, using the same data as above.



Graphs and their presentation

- If the same data has been used for both graphs, why do they look different? Which one is correct?



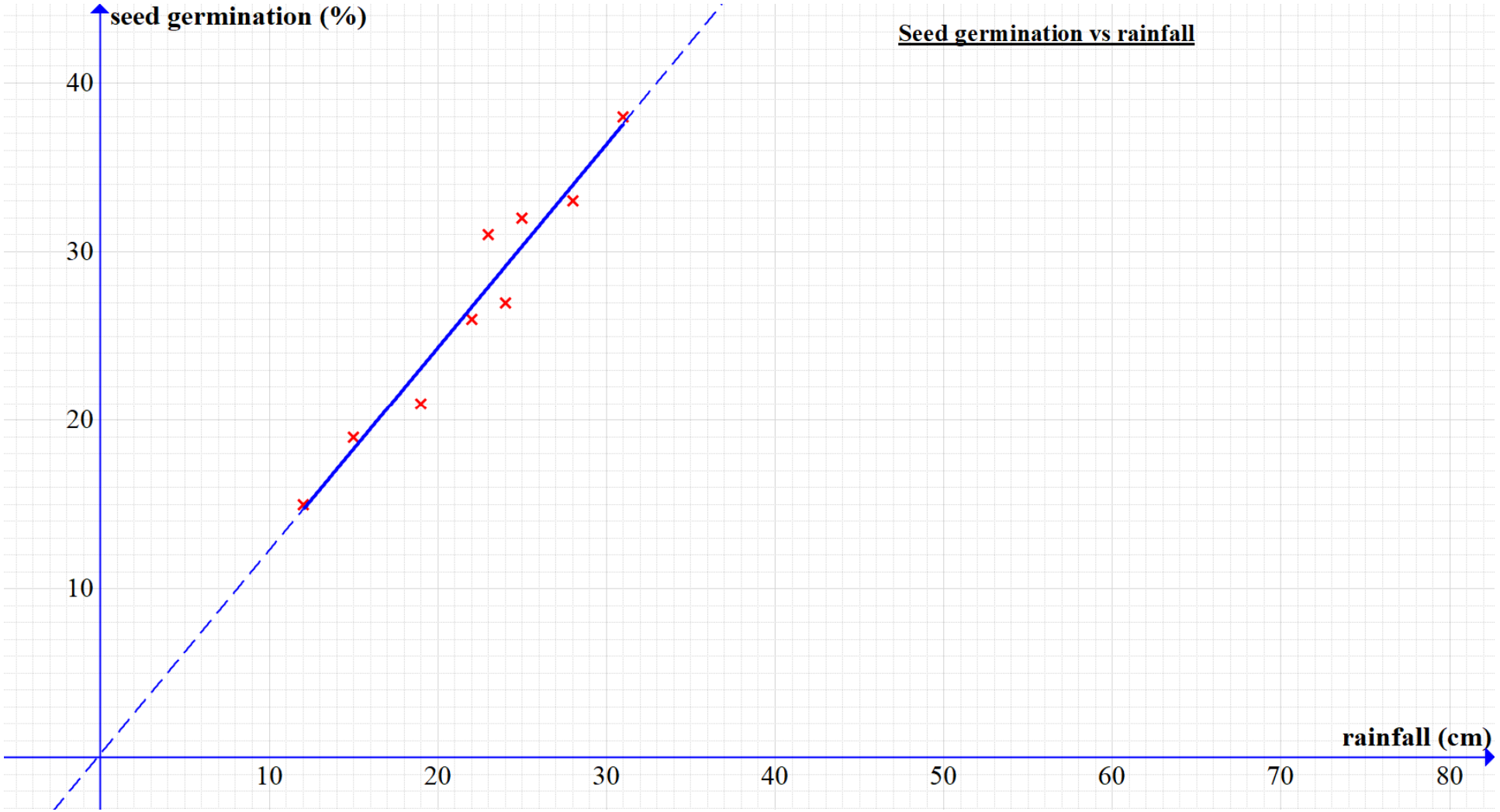


Graphs and their presentation

Example 6

- Percentage seed germination was collected for different amounts of rainfall. The data was plotted, and a best fit line drawn as shown on the next slide.

Graphs and their presentation



Graphs and their presentation

Example 6:

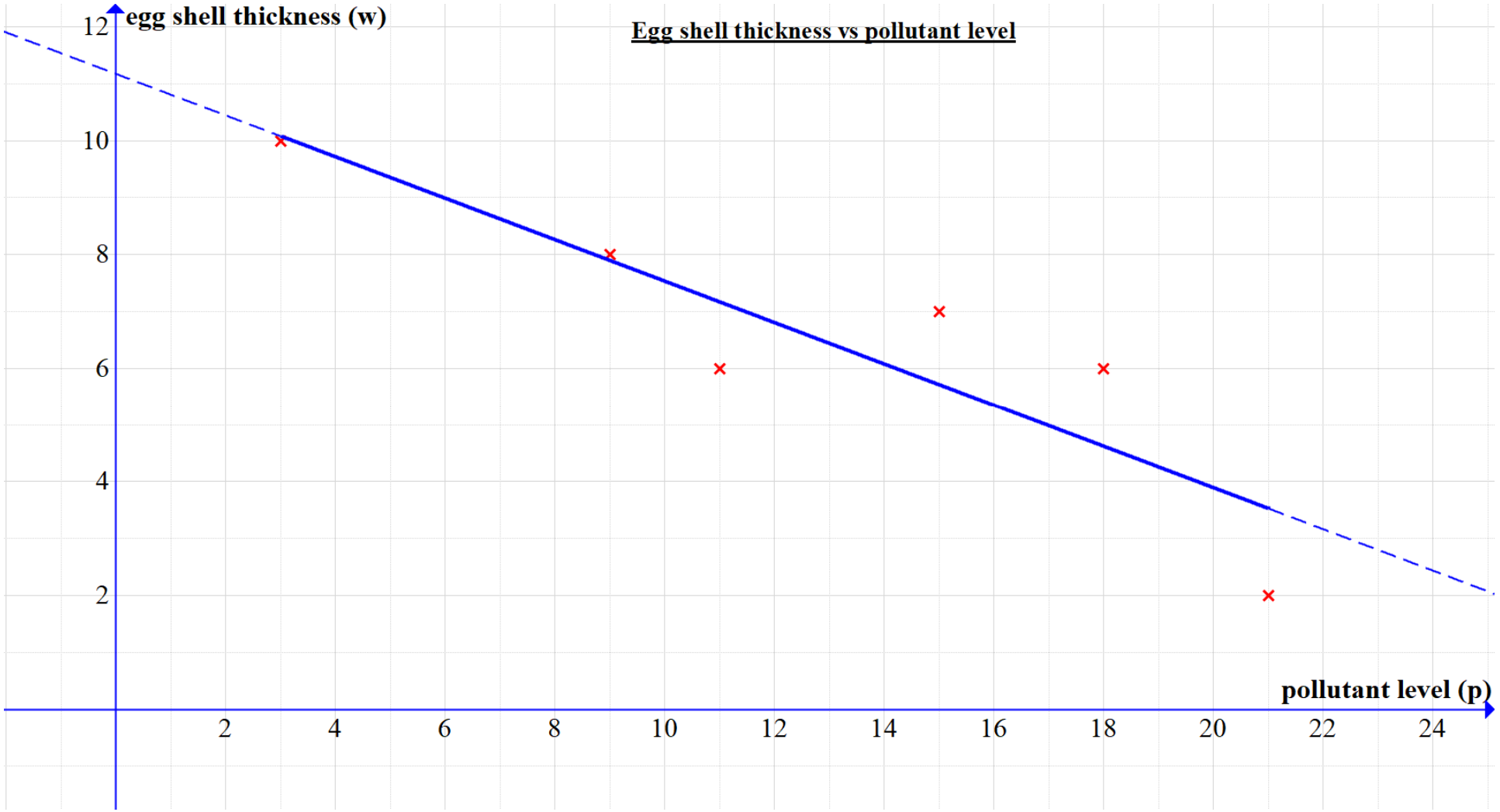
1) What is the predicted percentage germination for a rainfall of 30cm?
2) If the percentage germination is found to be 25% what is the estimated amount of rainfall?
3) What is the predicted percentage germination for a rainfall of 44cm? Can we trust this answer? If not, why not?

Graphs and their presentation

Exercise:

- Data was collected on the thickness of eggshells laid by birds of prey exposed to pollutants.
- A random sample was collected from 6 different nests, and tests for pollutant level p, and shell thickness w, was recorded, and plotted with a line of best fit as shown in the next slide.

Graphs and their presentation



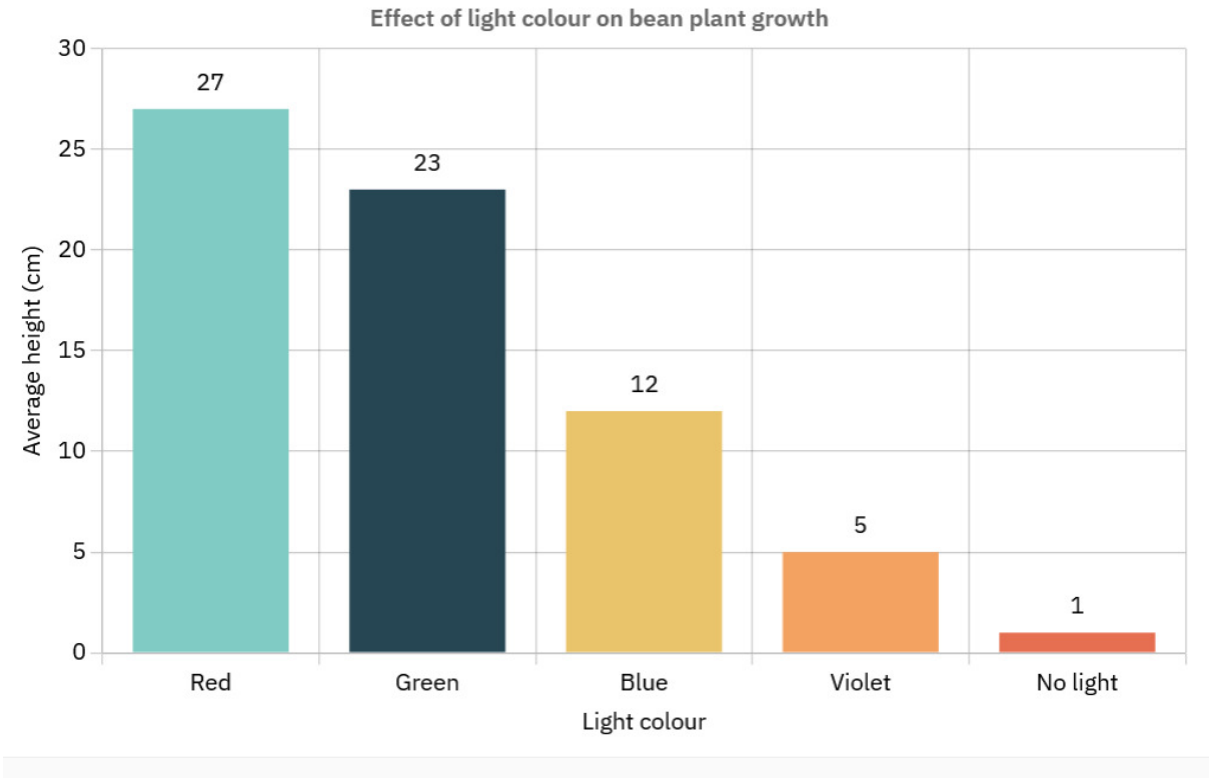
Graphs and their presentation

Questions

1) What can you deduce about the overall effect of pollution on eggshell thickness for these types of birds?
2) If you measured the thickness of an egg shell to be 4, what approximate pollutant level could you deduce?
3) What level of pollutants is required to decrease the thickness of an eggshell to 1? Can you trust this answer? If so, this would have to be biologically justified. If not, why not?

Post lesson exercises

1) Consider the bar chart below which illustrates the effect of light colour of decreasing frequency on bean plant growth

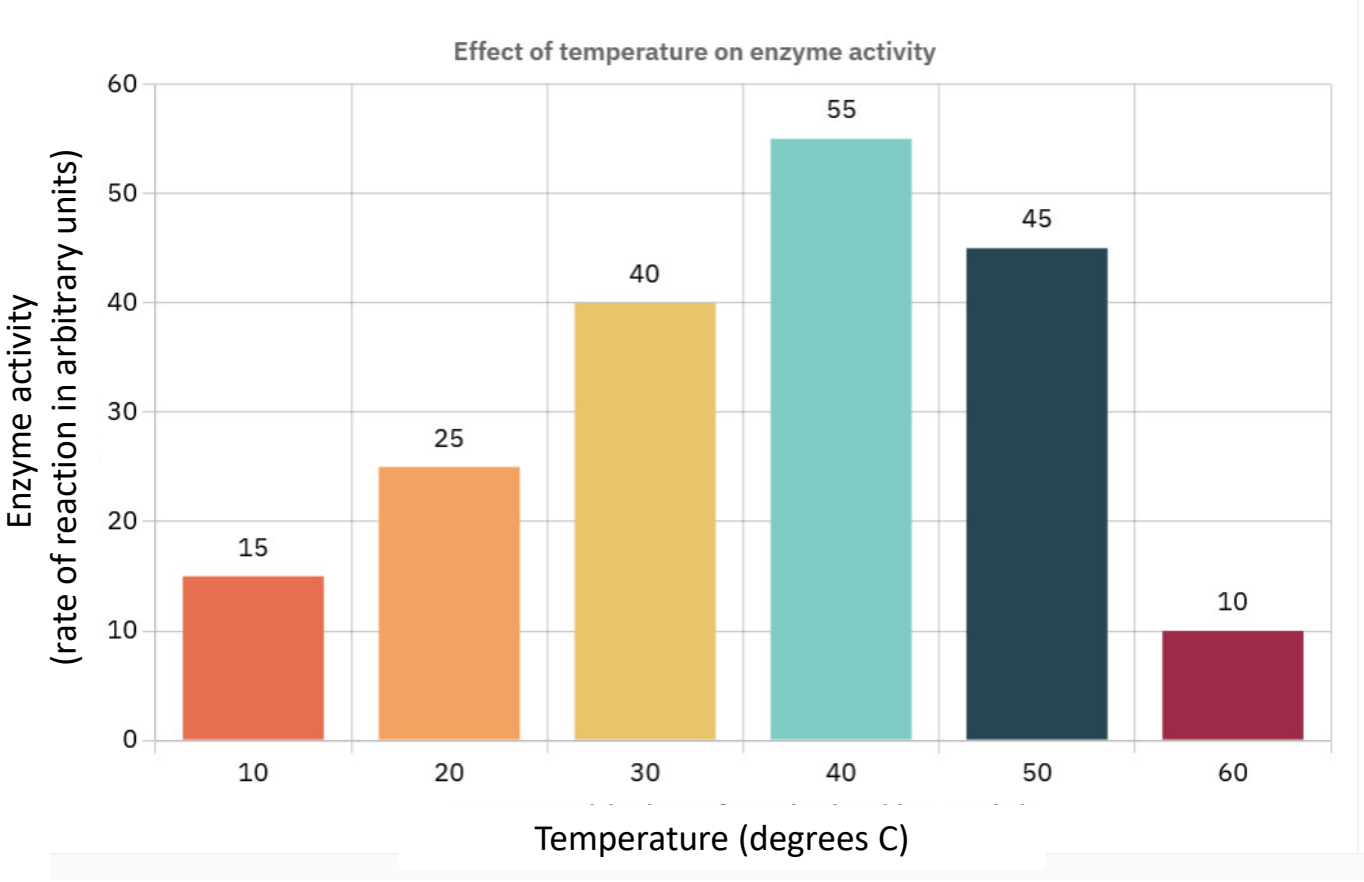


Post lesson exercises

a) Which light colour resulted in the tallest plants?
b) Which condition produced the least growth?
c) Calculate the average (mean) height of all plants.
d) Describe the effect of light frequency on the average height of bean plants.
e) Suggest a reason why plants grew poorly with no light.
f) What is one possible issue with the bar chart? This is a visualisation issue, not a biology issue.

Post lesson exercises

2) Consider the bar chart below which illustrates the effect of temperature on enzyme activity.



Post lesson exercises

a) At which temperature is enzyme activity the highest?

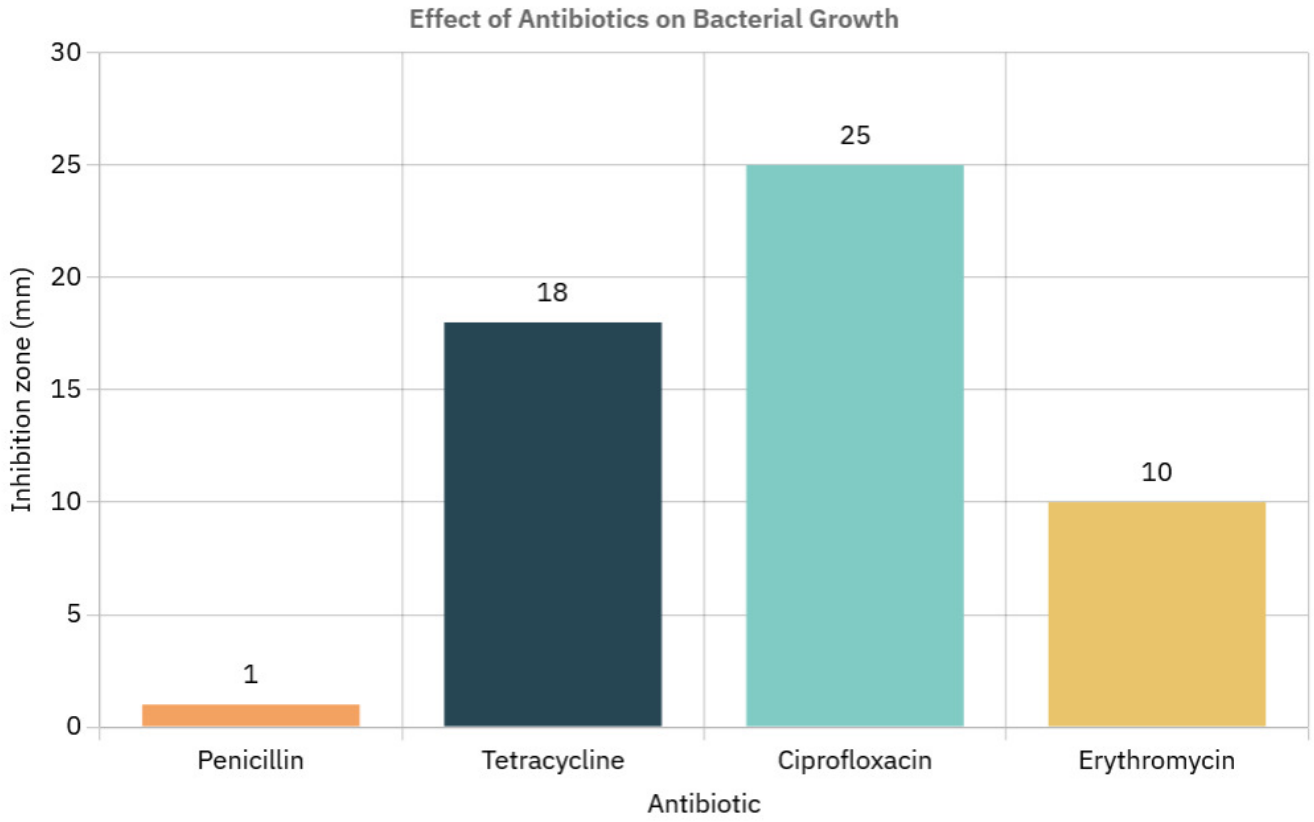
b) Calculate the rate of change in enzyme activity between 30°C and 40°C.

c) Describe the overall trend shown by the data.

d) Suggest a reason why enzyme activity increases as temperature rises up to 40°C.

Post lesson exercises

3) Consider the bar chart below which illustrates the effect of antibiotics on bacterial growth.



Post lesson exercises

a) Which antibiotic was most effective at inhibiting bacterial growth? Which was least effective?

b) Compare the effectiveness of Tetracycline and Erythromycin.

c) Calculate the percentage increase in effectiveness between Tetracycline and Erythromycin.

d) Penicillin showed no inhibition of *E. coli*. Suggest a biological reason for this observation.

Post lesson exercises

4) You are investigating the effect of light intensity on the rate of photosynthesis in a plant species.

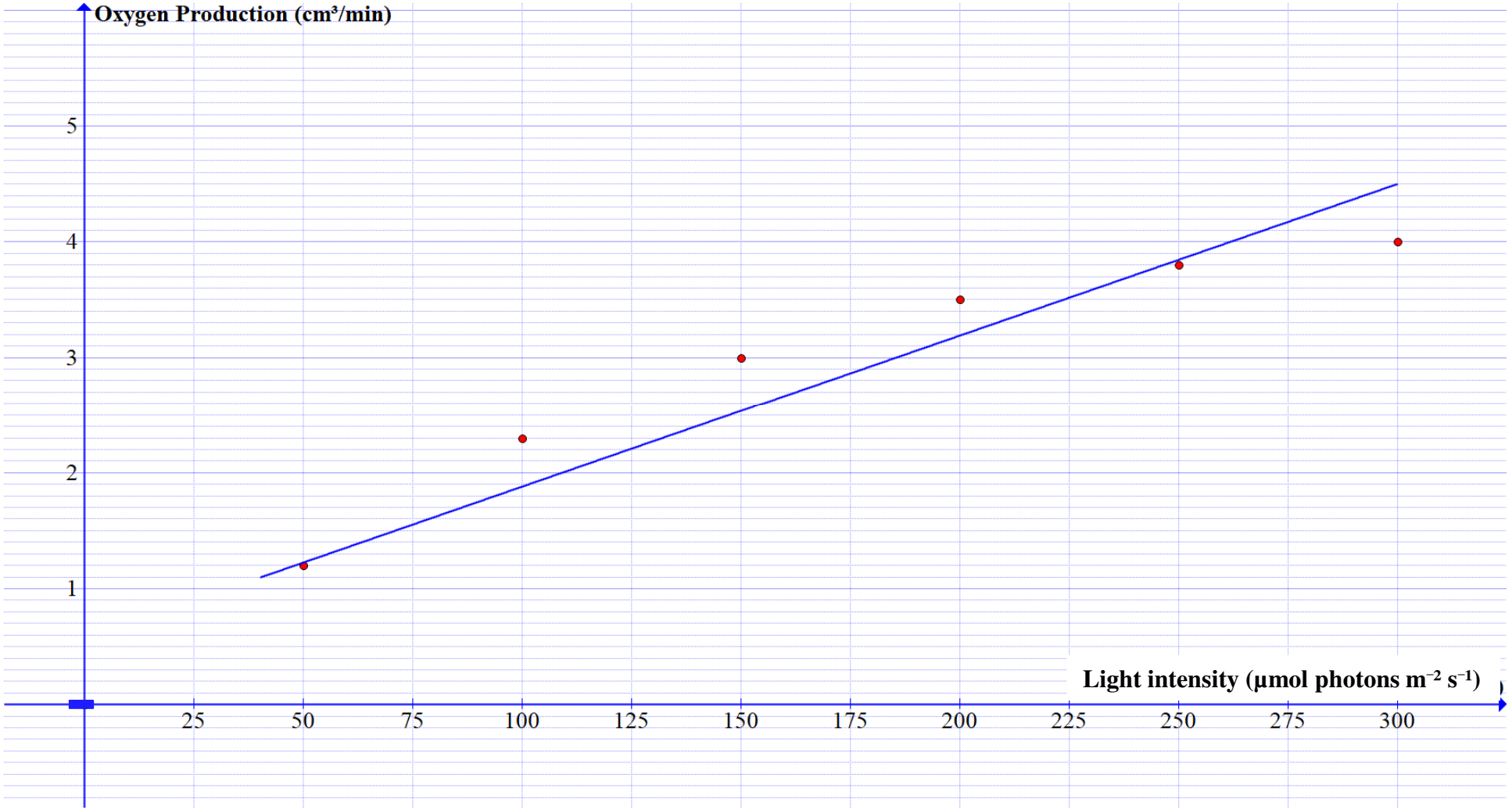
   The rate of photosynthesis was measured as the amount of oxygen produced (in cm^3 per minute) at different light intensities (in µmol photons m^{-2} s^{-1}).

   Data collected from an experiment is shown in the table on the next slide, with the scatter plot and line of best fit shown on the slide after that.

Post lesson exercises

Light Intensity (relevant units)	Oxygen Production (cm^3/min)
50	1.2
100	2.3
150	3.0
200	3.5
250	3.8
300	4.0

Post lesson exercises



Post lesson exercises

a) Predict the oxygen production at a light intensity of 75 units using your line of best fit.

b) Calculate the gradient of the line of best fit. What does this gradient mean biologically?

c) Find the light intensity at which photosynthesis reaches 90% of its maximum value.

Post lesson exercises

d) Suggest a biological reason why the rate of photosynthesis does not keep increasing indefinitely with light intensity.

e) Oxygen production seems to plateau/level off as light intensity increases. If this is the case, how valid/suitable is it to use a line of best fit? Explain your answer.

Post lesson exercises

5) You are investigating how substrate concentration affects the rate of reaction catalysed by an enzyme. The rate of reaction is measured in µmol/min.

   Data collected from an experiment is shown in the table on the next slide.

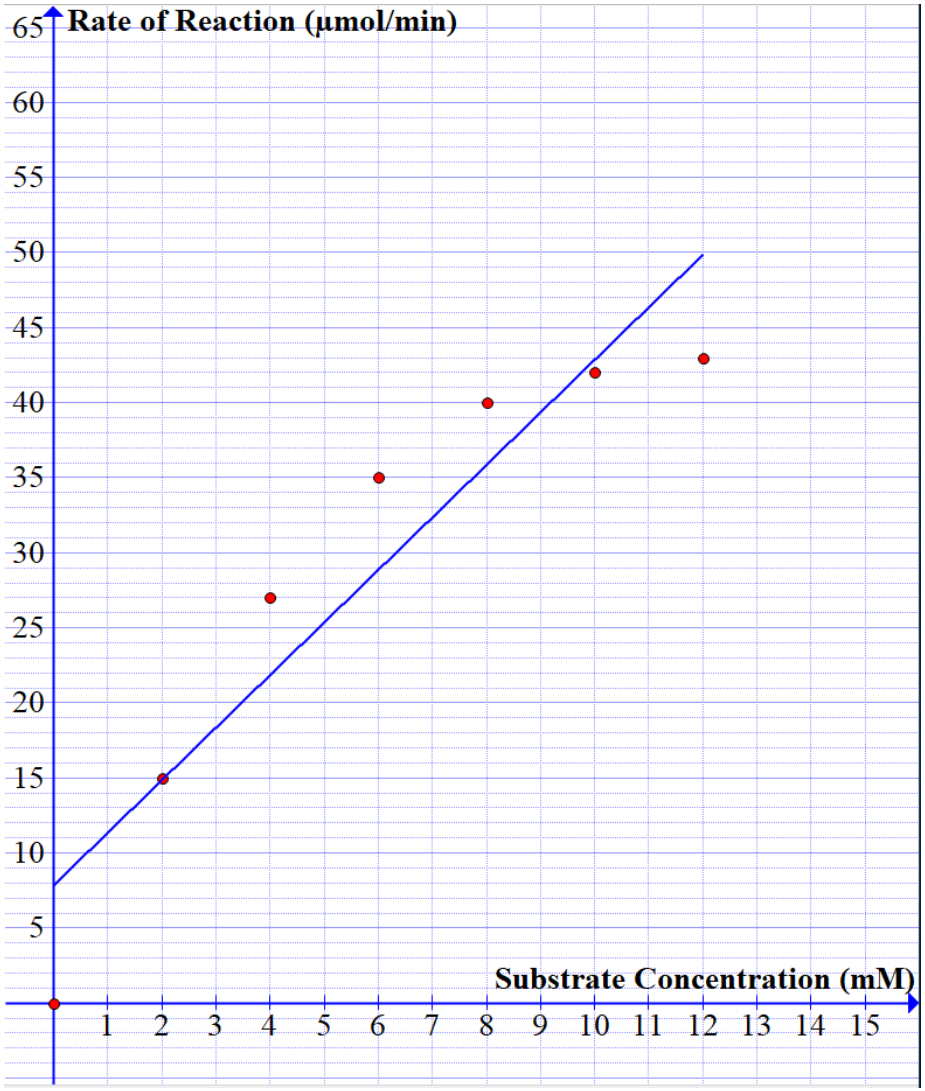
Post lesson exercises

Substrate Concentration (mM)	Rate of Reaction (µmol/min)
0	0
2	15
4	27
6	35
8	40
10	42
12	43

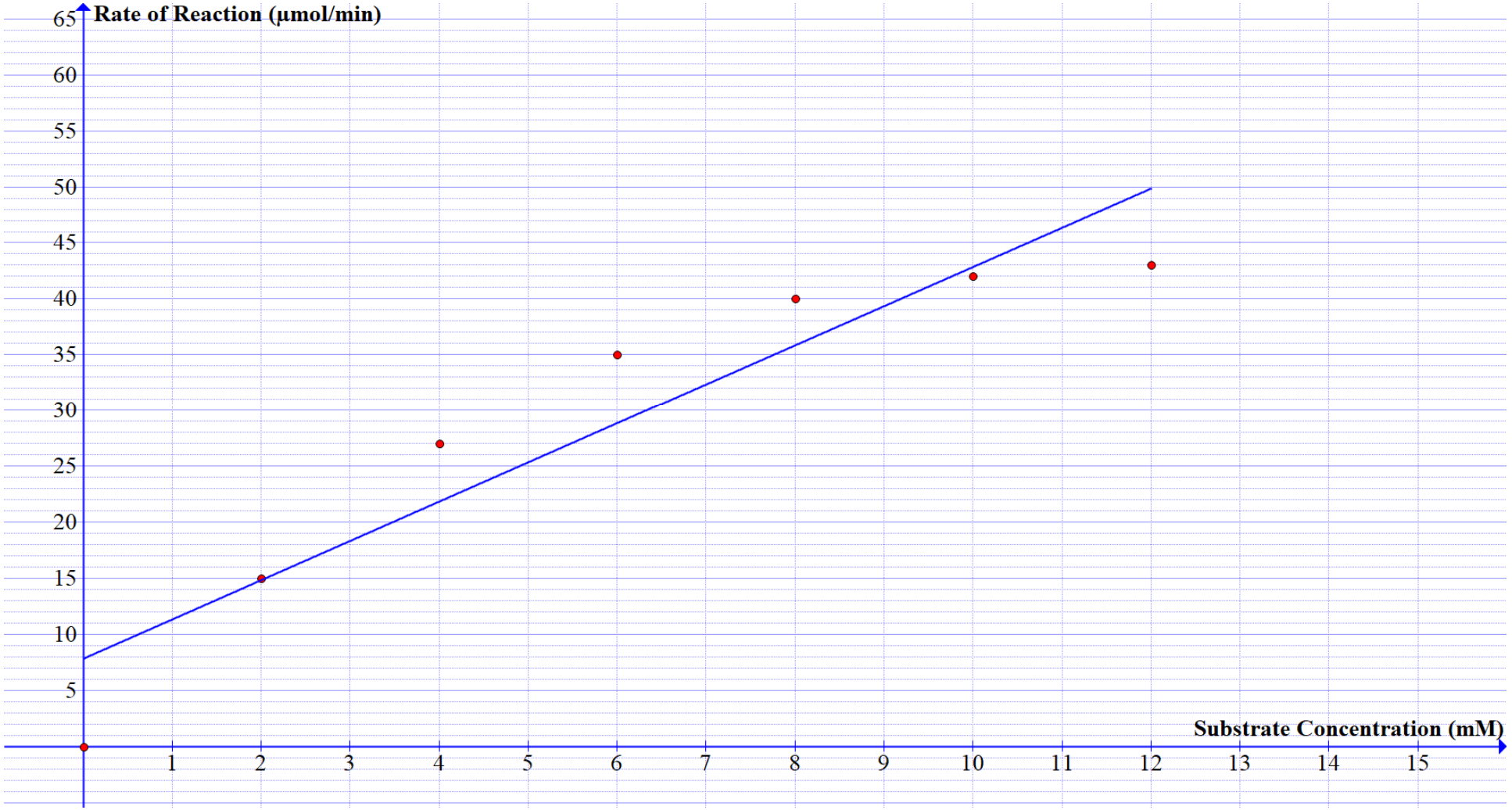
Post lesson exercises

Two scatter plots, along with their lines of best fit, are shown, one here and one on the next slide.

- Which line of best fit is correct?
- Are the lines of best fit different? Explain why these graphs look different.



Post lesson exercises



Post lesson exercises

a) Predict the rate of reaction at a substrate concentration of 3 mM.

b) Predict the rate of reaction at a substrate concentration of 14 mM. Can your prediction be trusted? Why, or why not?

c) Identify the point where the reaction rate begins to plateau.

d) Explain why the reaction rate begins to plateau.

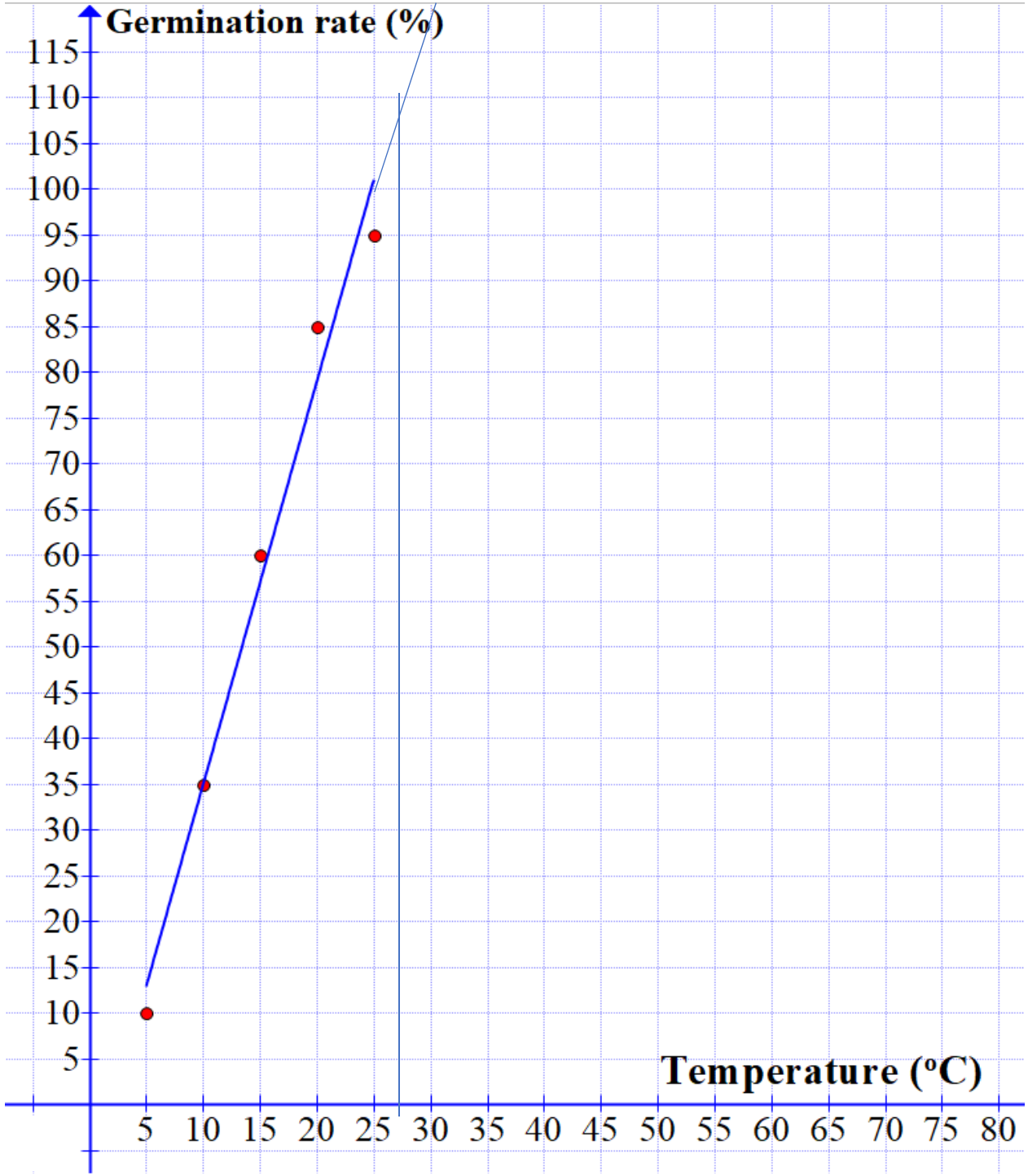
Post lesson exercises

6) You are investigating the effect of temperature on the germination rate of seeds stored inside a greenhouse. Germination rate is measured as the percentage of seeds germinated after 48 hours.

   Data collected from an experiment is shown in the table on the next slide, with the scatter plot and line of best fit shown on the slide after that.

Post lesson exercises

Temperature (°C)	Germination Rate (%)
5	10
10	35
15	60
20	85
25	95


Germination rate (%)
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
Temperature (ºC)
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80

Post lesson exercises

a) Estimate the germination rate at 12°C.

b) Find the percentage rate of germination.

c) Predict the germination rate at 27.5°C and explain whether or not your answer is valid in this case.

d) Two more data points are collected from the experiment,
as shown:

Temperature (°C)	Germination Rate (%)
30	90
35	70

Post lesson exercises

d) **Continued**: Plot these two data points on the previous graph.

i. Is it still valid to consider the whole data set as having a linear trend? Why?

ii. Is your answer to c) valid?