Co-funded by the European Union

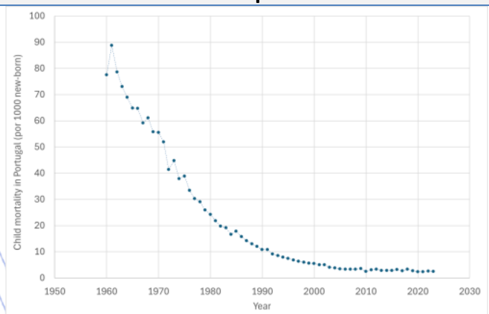
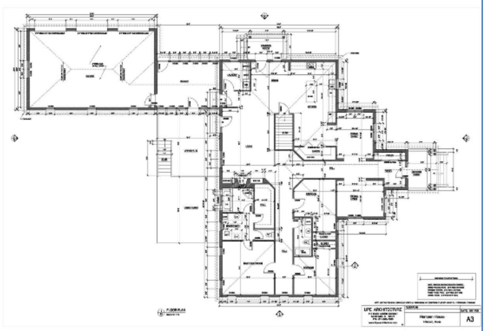
INTERPRET FIGURES AND MATHEMATICAL FORMULAE ORALLY AND IN WRITTEN FORM USING SPECIALISED VOCABULARY

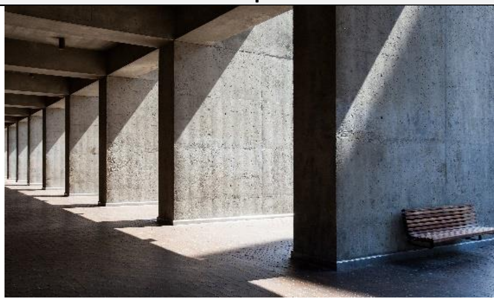
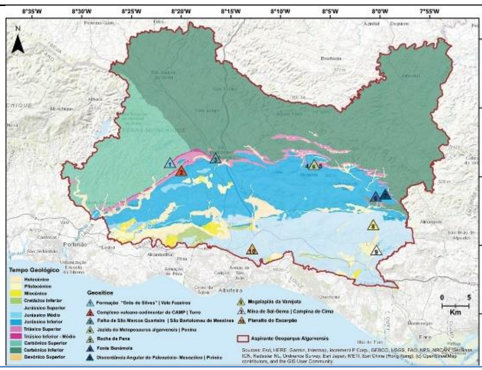
Handout Session 17

The first distinction we must do is about the different types of figures. Figures in a report are visual representations used to convey information. They encompass a wide range of elements, including graphs, diagrams, drawings, photographs/illustrations, and maps (Table 1). Essentially, any non-textual element besides tables is considered a figure. This means that in reports there will be commonly only "Figures" and "Tables" as non-textual elements.

Each figure element will have its set of mandatory features, which guarantee the correct and unique reading of the features in the element (Table 1). We will discuss below the mandatory features for each one the elements.

Table 1. Elements commonly included as figures in technical reports

Element	Example	Mandatory features
Graph		Two or more axis. Axis have numbered or coded scales. Axis are identified with legends (with units, when appropriate) May need colour or pattern legend.
Diagram		Should show the relationship between elements using connecting lines and/or arrows.
Technical Drawing		Intended to represent an object or organism. Made at scale representing the true relationship between dimensions. Should include legend with scale. Should include dimensions, when needed.

Element	Example	Mandatory features
Photograph/ illustration		Photographs are intended to show some physical entity (object, person, landscape, etc.). Should allow the correct interpretation of the entity, e.g., by using one- or two-point perspectives, with at least the vertical or horizontal parallel to image sides. Illustrations are freer, and can include, e.g., copies of computer code, free-hand drawings, etc.
Map		Representation of an earth's feature in two dimensions. Should contain legend, coordinates, scale, and orientation (direction of North).

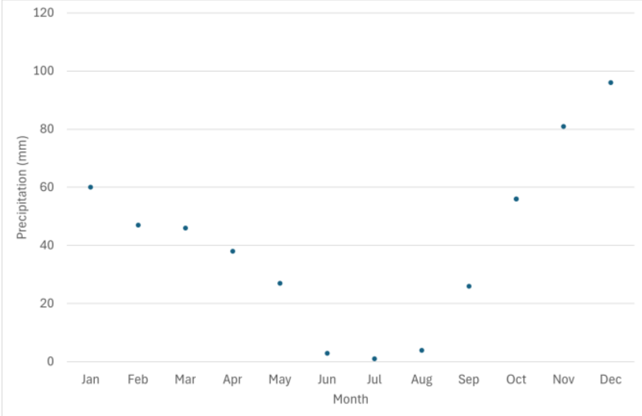
We will explore graphs with more detail. There are a very large variety of graphs, but the most common include: line graphs, scatter plots, bar graphs, pie charts, whisker plots (aka boxplot), high-low-close plots, ternary plots, and 3D-plots (Table 2).

The selection of a graph type depends on the data (numeric, non-numeric), number of variables, and the purpose of the study. For instance, if the purpose is to "study the evolution of child mortality in Portugal between 1960 and 2023" [mortality = function (time)], then we need the data for the time (e.g, year) and for the child mortality rate. Time will be the independent variable, and mortality will be the dependent [dependent variable = function (independent variable)]. The data comes from the Portuguese Institute of Statistics in the following format:

Year	Child mortality rate (per 1000 new-born)
1960	77.5
1961	88.8
1962	78.6
1963	73.1
1964	69.0
1965	64.9
1966	64.7
(...)	(...)

The resulting graph is shown in Figure 1. Note that it includes legends in both axes and units, as mandatory (Table 2). The graph is a scatter plot as both variables are numeric. The symbols are connected by a line to convey the meaning of continuity, given that time is a continuous variable.

Table 2. Common graph types, what they represent, how the data should be prepared and how they look like

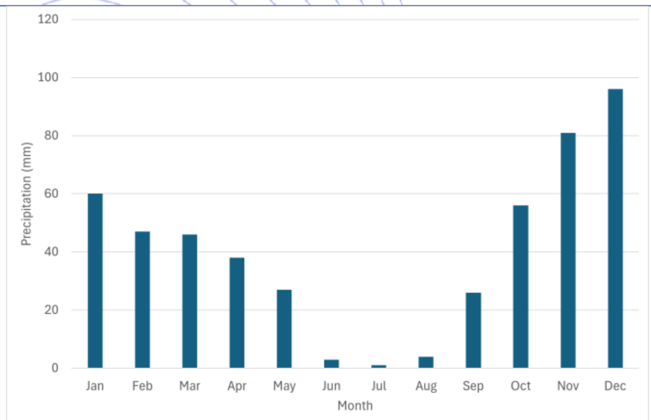
<table>
<tr><th>Graph type</th><th>Example</th><th>What it represents</th><th>Data format</th></tr>
<tr><td colspan="3">Graphs with one or more independent variable(s)</td><td></td></tr>
<tr><td>Line graph</td><td>
</td><td>It shows how the values of the dependent variable vary for different categories of the independent variable.

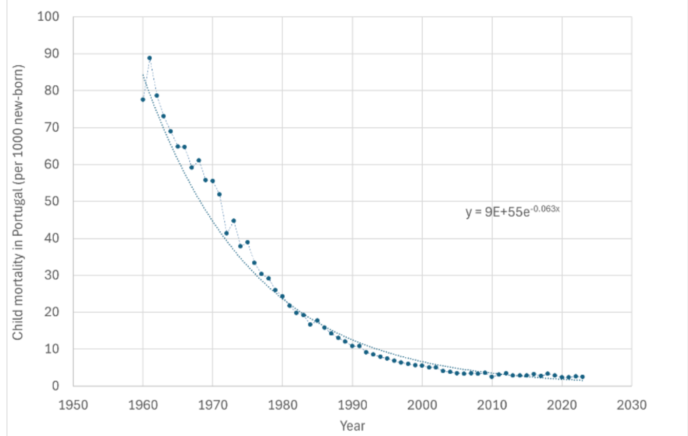
The independent variable is not a number (in the example it is the “month”).

Each value is represented by a symbol.

The distance in the x-axis for consecutive values of y is constant (for instance, if the distance between January and February is 1 cm in the horizontal, then the horizontal distance between June and July is also 1 cm).

Do not connect the symbols with lines to avoid confusion with scatter plots.

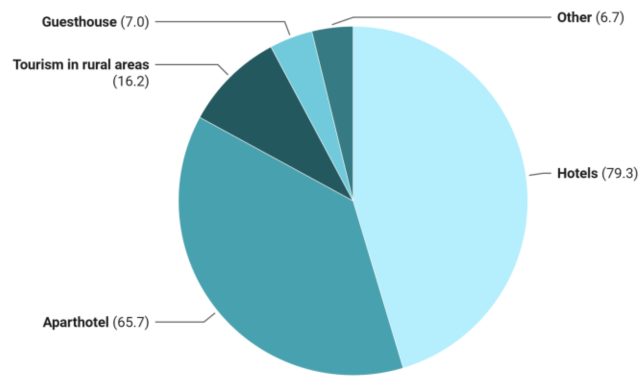
The example shows the value of rainfall in millimetres per month at a given location.</td><td><table><tr><th>Month</th><th>Precipitation (mm)</th></tr><tr><td>Jan</td><td>60</td></tr><tr><td>Feb</td><td>47</td></tr><tr><td>Mar</td><td>46</td></tr><tr><td>Apr</td><td>38</td></tr><tr><td>May</td><td>27</td></tr><tr><td>Jun</td><td>3</td></tr><tr><td>Jul</td><td>1</td></tr><tr><td>Aug</td><td>4</td></tr><tr><td>Sep</td><td>26</td></tr><tr><td>Oct</td><td>56</td></tr><tr><td>Nov</td><td>81</td></tr><tr><td>Dec</td><td>96</td></tr></table></td></tr>
<tr><td>Bar chart</td><td>
</td><td>It represents the same information as the line graph, but the symbol is replaced by a column.</td><td><table><tr><th>Month</th><th>Precipitation (mm)</th></tr><tr><td>Jan</td><td>60</td></tr><tr><td>Feb</td><td>47</td></tr><tr><td>Mar</td><td>46</td></tr><tr><td>Apr</td><td>38</td></tr><tr><td>May</td><td>27</td></tr><tr><td>Jun</td><td>3</td></tr><tr><td>Jul</td><td>1</td></tr><tr><td>Aug</td><td>4</td></tr><tr><td>Sep</td><td>26</td></tr><tr><td>Oct</td><td>56</td></tr><tr><td>Nov</td><td>81</td></tr><tr><td>Dec</td><td>96</td></tr></table></td></tr>
</table>

<table>
<tr><th>Graph type</th><th>Example</th><th>What it represents</th><th>Data format</th></tr>
<tr><td>Scatter-plot</td><td>
</td><td>It represents the relationship between two or more numeric variables.

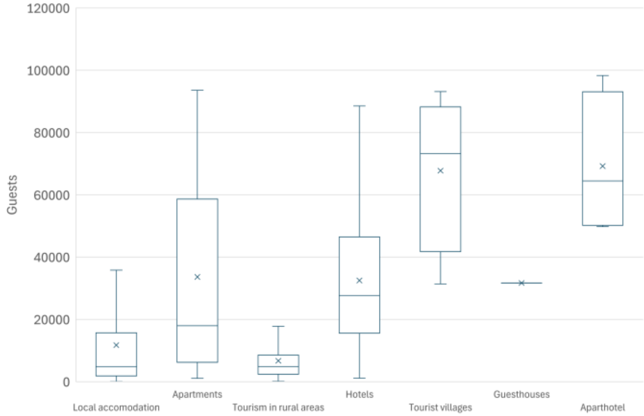
In the example, it shows the fast decrease in child mortality rate in Portugal over the last six decades.

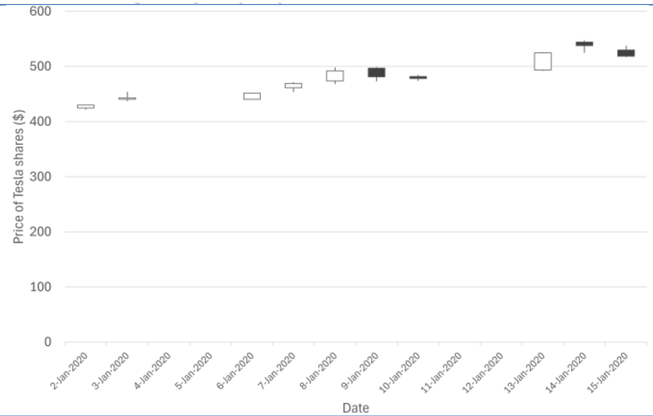
Each year is a numeric value (x-axis) and the mortality rate is the y-axis).

The axes include the legend and unit of measurement.

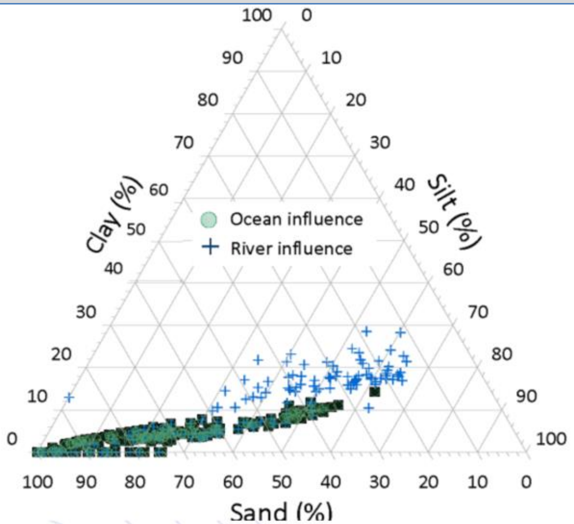
In the scatterplot the difference between two consecutive values of x (lag) does not have to be constant (in opposition to the line chart above).</td><td><table><tr><th>Year</th><th>Child mortality</th></tr><tr><td>1960</td><td>77.5</td></tr><tr><td>1961</td><td>88.8</td></tr><tr><td>1962</td><td>78.6</td></tr><tr><td>1963</td><td>73.1</td></tr><tr><td>1964</td><td>69</td></tr><tr><td>1965</td><td>64.9</td></tr><tr><td>1966</td><td>64.7</td></tr><tr><td>1967</td><td>59.2</td></tr><tr><td>1968</td><td>61.1</td></tr><tr><td>1969</td><td>55.8</td></tr><tr><td>1970</td><td>55.5</td></tr></table>Etc..</td></tr>
<tr><td>Pie Chart</td><td>
</td><td>Represents percentages.

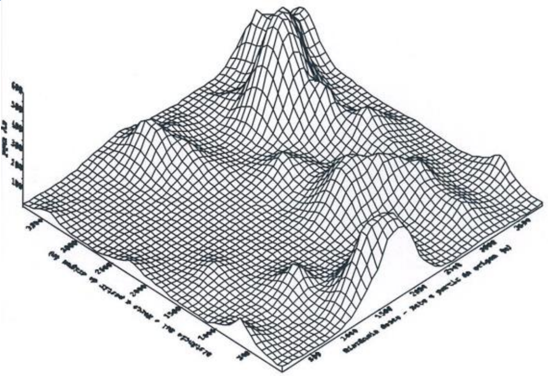
In the example, the pie chart shows the percent division of guests in Portugal per type of accommodation.</td><td><table><tr><th>Type of accomodation</th><th>%</th></tr><tr><td>Hotels</td><td>65.7</td></tr><tr><td>Aparthotel</td><td>7.0</td></tr><tr><td>Guesthouse</td><td>1.2</td></tr><tr><td>Apartments</td><td>3.9</td></tr><tr><td>Tourist villages</td><td>1.6</td></tr><tr><td>Local accommodation</td><td>16.2</td></tr><tr><td>Tourism in rural areas</td><td>4.5</td></tr></table></td></tr>
</table>

<table>
<tr><th>Graph type</th><th>Example</th><th>What it represents</th><th>Data format</th></tr>
<tr>
<td>Whisker plot/box-plot</td>
<td>
</td>
<td>Represents statistical information for a large dataset, divided by categories. Mean: x; Median: central line in the box; Interquartile range: top and bottom of box; Maximum and Minimum of data: top and bottom whiskers.

The example shows the statistics of the number of guests per type of accommodation, considering data for all Portuguese municipalities.</td>
<td>
<table>
<tr><th>Municipality</th><th>Type of accomodation</th><th>Guests</th></tr>
<tr><td>Moita</td><td>Local accomodation</td><td>67</td></tr>
<tr><td>Lourinhã</td><td>Apartments</td><td>3278</td></tr>
<tr><td>Sátão</td><td>Tourism in rural areas</td><td>125</td></tr>
<tr><td>Penedono</td><td>Tourism in rural areas</td><td>136</td></tr>
<tr><td>Pedrógão Grande</td><td>Tourism in rural areas</td><td>140</td></tr>
<tr><td>Penedono</td><td>Hotels</td><td>1135</td></tr>
<tr><td>Ourique</td><td>Local accomodation</td><td>225</td></tr>
<tr><td>Monção</td><td>Local accomodation</td><td>335</td></tr>
<tr><td>Rio Maior</td><td>Local accomodation</td><td>336</td></tr>
<tr><td>Vila Pouca de Aguiar</td><td>Local accomodation</td><td>349</td></tr>
</table>
Etc..</td>
</tr>
<tr>
<td>High-low-close plot</td>
<td>
</td>
<td>Represents the values of trade prices of commodities or shares. Shows the market open price, the maximum, minimum, and market closing price for a given period.

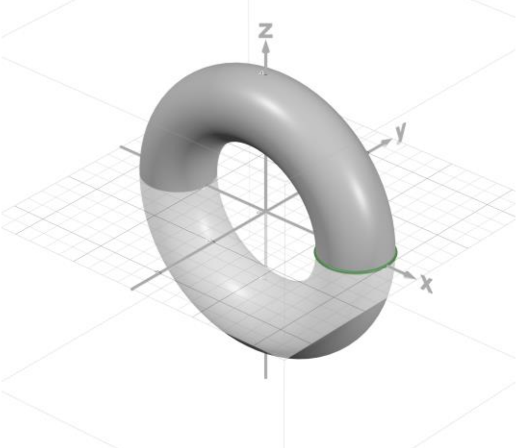
The example shows these statistics for Tesla shares during the first 15 days in 2020.</td>
<td>
<table>
<tr><th>Date</th><th>open</th><th>high</th><th>low</th><th>close</th></tr>
<tr><td>2020-01-02</td><td>424.5</td><td>430.7</td><td>421.7</td><td>430.3</td></tr>
<tr><td>2020-01-03</td><td>440.5</td><td>454</td><td>436.9</td><td>443</td></tr>
<tr><td>2020-01-06</td><td>440.5</td><td>451.6</td><td>440</td><td>451.5</td></tr>
<tr><td>2020-01-07</td><td>461.4</td><td>471.6</td><td>453.4</td><td>469.1</td></tr>
<tr><td>2020-01-08</td><td>473.7</td><td>498.5</td><td>468.2</td><td>492.1</td></tr>
<tr><td>2020-01-09</td><td>497.1</td><td>498.8</td><td>472.9</td><td>481.3</td></tr>
<tr><td>2020-01-10</td><td>481.8</td><td>484.9</td><td>473.7</td><td>478.2</td></tr>
<tr><td>2020-01-13</td><td>493.5</td><td>525.6</td><td>492</td><td>524.9</td></tr>
<tr><td>2020-01-14</td><td>544.3</td><td>547.4</td><td>524.9</td><td>537.9</td></tr>
<tr><td>2020-01-15</td><td>529.8</td><td>537.8</td><td>516.8</td><td>518.5</td></tr>
</table></td>
</tr>
</table>

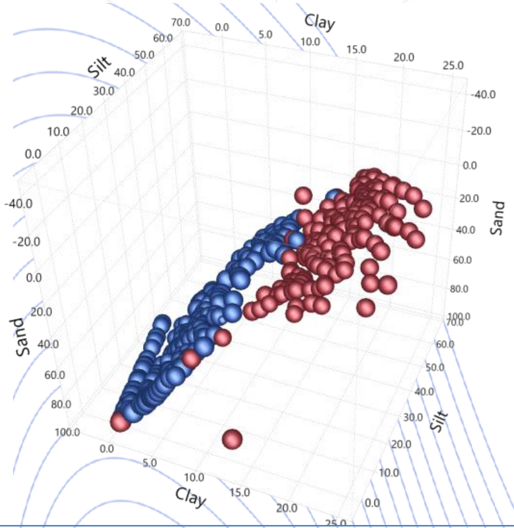
<table>
<tr><th>Graph type</th><th>Example</th><th>What it represents</th><th>Data format</th></tr>
<tr>
<td>Ternary plots</td>
<td>
</td>
<td>Represent simultaneously the values of three variables (which sum to 1 or to 100) in a triangular graphical system.

The example shows the representation of granulometric fractions of sediments in a Brazilian mangal under the influence of riverine and maritime influence.</td>
<td>
<table>
<tr><th>Influence</th><th>% sand</th><th>% silt</th><th>% clay</th></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>99.8</td><td>0.2</td><td>0.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>77.4</td><td>17.5</td><td>5.1</td></tr>
<tr><td>River</td><td>87.0</td><td>0.0</td><td>13.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>99.9</td><td>0.1</td><td>0.0</td></tr>
<tr><td>River</td><td>100.0</td><td>0.0</td><td>0.0</td></tr>
<tr><td>River</td><td>99.7</td><td>0.3</td><td>0.0</td></tr>
</table>
Etc..</td>
</tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr>
<td>Pseudo-3D and 3D-plots</td>
<td></td>
<td>Pseudo-3D represents one variable in relation to two other which are assumed to be a coordinate system.

The example shows the concentration of the metal arsenic (As) in soil at a given location in Portugal. The x and y-axes are longitude and latitude, respectively.

3D graphs show the features of an object or data values in a three-dimensional coordinate system.</td>
<td>
<table>
<tr><th>Longitude (x)</th><th>Latitude (y)</th><th>Arsenic concentration (ug/kg)</th></tr>
<tr><td>35621</td><td>54585</td><td>56</td></tr>
<tr><td>75512</td><td>78235</td><td>154</td></tr>
<tr><td>88225</td><td>55225</td><td>89</td></tr>
<tr><td>85326</td><td>85422</td><td>235</td></tr>
</table>
Etc..</td>
</tr>
</table>

<table>
<tr><th>Graph type</th><th>Example</th><th>What it represents</th><th>Data format</th></tr>
<tr><td></td><td>


</td><td>The second example shows the 3D representation of a parametric surface of revolution:

$$f(t) = (3 + \cos t,\ \sin t)$$

The third example shows the 3D plot of texture fractions in Brazilian mangrove sediments where the coordinates are % of sand, silt and clay.</td><td>

Influence	Sand	Silt	Clay
River	100.0	0.0	0.0
River	100.0	0.0	0.0
River	100.0	0.0	0.0
River	99.8	0.2	0.0
River	100.0	0.0	0.0
River	77.4	17.5	5.1
River	87.0	0.0	13.0
River	100.0	0.0	0.0

Etc..

Same data as above</td></tr>
</table>

The graph shows a fast decrease in mortality (Figure 1), but how can we describe it in technical words? For that we need to look at the shapes of known functions. Common mathematical functions are linear, power, quadratic, polynomial, exponential/geometric, logarithmic, and sinusoidal (Table 3).

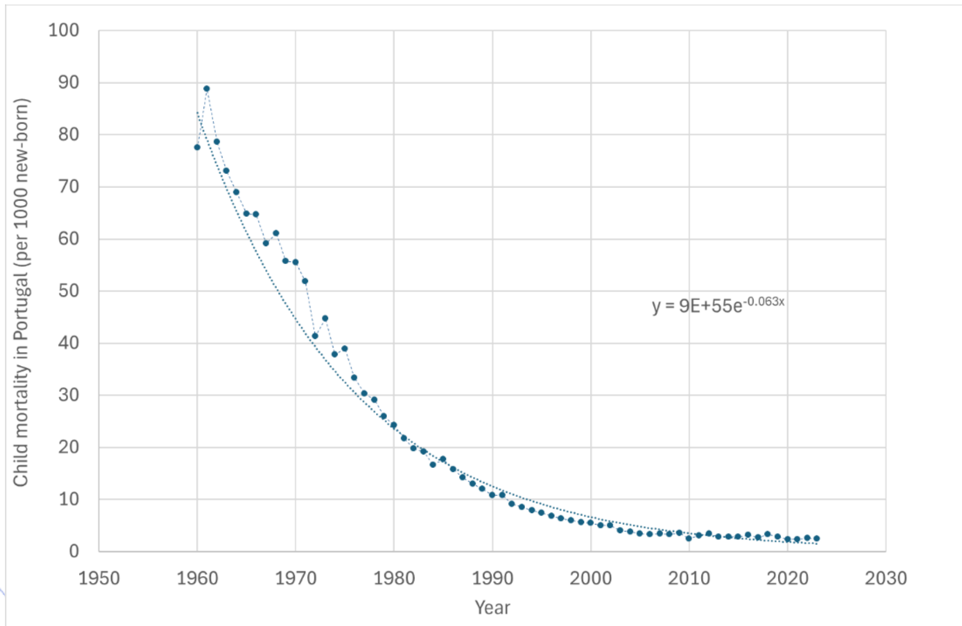
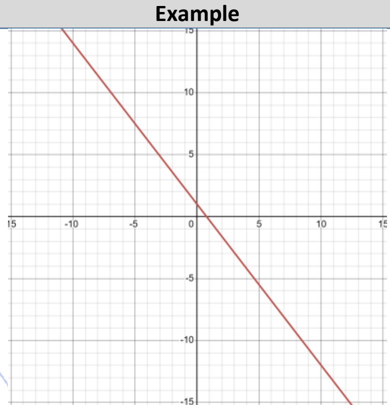
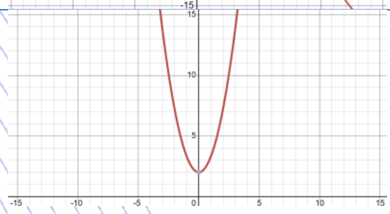
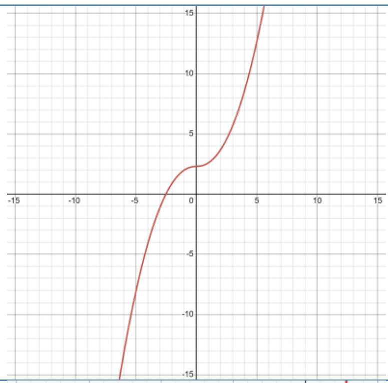
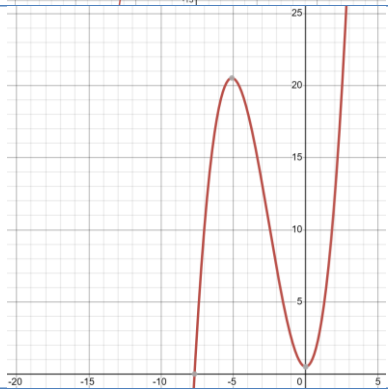
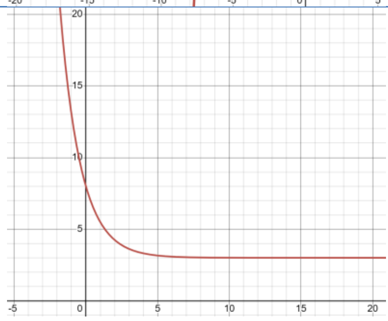
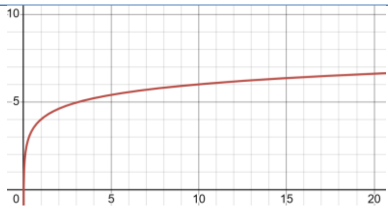
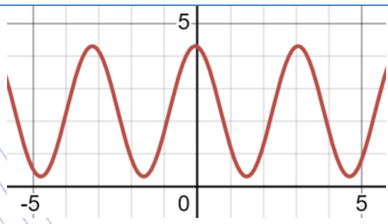


Figure 1. Evolution of child mortality (per 1000 new-born) in Portugal between 1960 and 2023

Given the shape of the evolution of child mortality over time, and the fact that it follows closely a geometric decrease, we could hypothesize that "the mortality rate seems to have followed a geometric decrease in the analysed period". We would then fit a mathematical model to the data and do the necessary statistical analysis to validate the model.

Table 3. Common mathematical functions. Their designation, function and application

Designation	Function	Example	Application
Linear	$y = ax + b$		The dependent variable is linearly dependent on the independent. As the slope becomes more vertical, the rate of change becomes faster. When the line passes through the origin (0,0), the function represents direct proportionality between one variable and the other.
Quadratic	$y = ax^2 + bx + c$		The dependent variable varies fast with the changes in the independent and has one inflection point (change of signal). It models, e.g., the trajectories of projectiles.

Designation	Function	Example	Application
Power	$Y = ax^k$ (both a and k are reals)		The dependent variable varies as a power of the independent. Represents fast changes. Examples include population growth, radioactive decay, and fluid resistance.
Polynomial	e.g., $y = ax^3 – bx^2 + cx + d$ (involves only non-negative integer powers of x, combined with addition, subtraction, and multiplication		The dependent variable has a complex non-linear relationship with the independent. Simple examples include the trajectory of projectiles that do not follow quadratic function.
Exponential/ geometric	$Y = a\ b^x$		Represent very fast changes in the dependent variable as a function of the dependent. Can be used to model, e.g., population growth (at early stages). Geometric sequences have discrete terms (specific points), whereas exponential functions can continuously take any real number input.
Logarithmic	$Y = \log_b (x)$ (it is the inverse of the exponential)		It works as the inverse of the exponential function. Examples of use include the Richter scale for earthquake magnitude, the decibel scale for sound intensity, pH for acidity.
Sinusoidal	$y = a \sin(b(x - c)) + d$ or $y = a \cos(b(x - c)) + d$, where: a is the amplitude, b affects the period, c is the phase shift, and d is the vertical shift.		It is used to model periodic relationships between the dependent and the independent. Examples include tides, air temperature, and sound waves.

Tables

Tables are used to present comparative data, statistics, schedules, pricing information, research results, contact lists, or any information that benefits from organized categorization. They are particularly useful to show relationships between different types of data or when readers need to quickly locate specific information.

They are structured arrangements of data organized into rows and columns, creating a grid format that makes information easy to read, compare, and analyse. Tables consist of several key components, namely: Table caption (usually numbered), rows (horizontal lines of data), columns (vertical divisions of information), cells (individual boxes where rows and columns intersect), column headers, and row identifiers when needed (Table 4).

Table 4. Example of table with column and row identifiers [caption]

	Column header	Column header	Column header	Column header
		Group header		
Row identifier	cell			
Row identifier				
Row identifier				
		Group header		
Rows identifier				

When filling the table with the data, always represent the numerals with the same number of significant figures as those of your initial data, or at least use the same number of decimal places as in the original data. See more details below.

All tables must be cross-referenced in the text. They should be inserted in between text paragraphs, not at the start of sections.

Equations

Equations are a very compact and unique way of conveying relationships between variables, no matter how complex. They can be placed In line with text when the equation is simple and short, is part of the sentence flow, and does not disrupt reading. For instance, "The decrease of concentration of arsenic in water follows a first-order kinetics, $C = C_0 e^{-kt}$, ..."

If the equations are complex or long, contain fractions, integrals, or multi-level expressions, or represent key result or important relationships, then they should be placed separately. In this case they should be numbered sequentially, inserting the number to the right of the equation, inside brackets. All variables and parameters in the equations should be identified, before or after the equation, the first time they appear (including their units). For instance:

$$C_t = C_0\, e^{-kt} \qquad [1]$$

Where C_t and C_o are the concentrations (mg/L) at time *t*=*t* and *t*=0, respectively; and *k* is the degradation constant (h^{-1}).

(Note: to put the number in front of the equation, insert a table with one line and two columns, set the width of the columns to fit the equation and the number in the correct position, and then set the margins of the table to transparent)

Numbers and units

When reporting measurements, the values should include all the digits one is certain of, plus one final digit which is an estimate. For example, when reading the voltmeter where the electric potential is between two markings (5 and 6 V), the last digit in the volt measurement is estimated based on where the needle is compared to the nearest markings (in the example, approximately 5.2 V) (Figure 2).

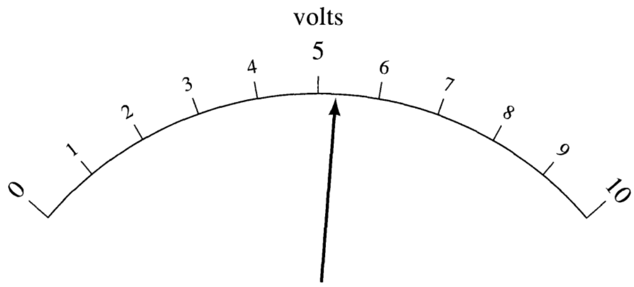


Figure 2. Reporting measurements – example. The uncertainty of the measurement is between 5 and 6, so on the first decimal place

If the measurement is repeated, most probably the values will not be the same (for instance, when we take our body temperature to check for fever, two or three consecutive measurements are usually not equal…). So, we should be able to report that uncertainty (a kind of quantitative "more-or-less 38,5 °C").

The best way to determine the reliability of results is to repeat the measurement and then calculate the average and standard deviation of the results. The standard deviation quantifies how much scatter (or uncertainty) there is in the measured data.

Rounding the standard deviation to one significant figure will tell in which decimal place the uncertain digit of the result lies. For instance, in our example above, the significant digit is on the first decimal place, 0.1 V.

The results are then reported as: mean ± standard deviation. In our example, 5.2 ± 0.1 V. Note that the uncertainty term (after the ±) has at most one or two digits.

Each measurement and the results of computations should be reported using the same number of significant figures.

To know how many significant figures to set, use the following criteria:

- Check the limit of quantification (LOQ) of the method of measurement. It will report the lowest value measured with certainty. Search the lowest value in the dataset and add one digit to the right to establish where the uncertainty lies. Count the number of significant figures. All the other values in the dataset should be reported using this number of significant figures.

- If doing a measurement, add one digit to the right of the reading from the measurement device. Count the number of significant figures. All the other values in the dataset should be reported using this number of significant figures.

To compute the number of significant figures (s.f.), use these rules:

1. All nonzero digits are significant	12.546 has 5 s.f.
2. Zeroes between nonzero digits are significant.	4.056 has 4 s.f.
3. Zeroes to the left of the first nonzero digits are not significant.	0.0012 has 2 s.f.
4. Zeroes to the right of a decimal point in a number are significant.	0.100 has 3 s.f.
5. When a number ends in zeroes that are not to the right of a decimal point, the zeroes are not necessarily significant. To avoid ambiguity, use scientific notation.	5.5×10^2 has 2 s.f. 5.50×10^2 has 3 s.f.

When doing mathematical operations, use these rules:

1. For addition or subtraction, round off the result to the leftmost decimal place.	40.123+20.34 = 60.46
2. For multiplication or division, round off the result to the smallest number of significant figures.	1.23 x 2.0 = 2.5
3. For logarithms, retain in the mantissa (the number to the right of the decimal in the result) the same number of significant figures as there are in the number whose logarithm is being taken.	log(12.8) = 1.107 The mantissa is .107 and has 3 s.f. because 12.8 has 3 s.f.
4. For exponents, the number of significant figures is the same as in the mantissa.	$10^{1.23} = 17$ Because the mantissa (.23) has 2 s.f.
5. For multiple calculations, compute the number of significant digits to retain in the same order as the operations:	logarithms -> exponents -> multiplication and division -> addition and subtraction

When transcribing values from spreadsheets that use scientific notation, e.g., 3.26E-3, convert to the correct mathematical representation by replacing the "E-"by "x10", and raising the exponent. The value above becomes 3.26×10^{-3}.

All measurements should be stated in metric units based on the International System of Units (SI). Some exceptions are allowed for units of common use, such as litre (equal to 1000 ml, 100 cl, 10 dl), represented by L, and hectare (equal to an area of 100 m x 100 m = 10 000 m^2), represented by ha.