Working with data in graphs

GLOBE		*Related SDGs:*	*Type of activity*
Sphere	**Protocols**		
All	All protocols	4 (Quality Education)	Evaluation Communication

Overview

Understand the importance of data analysis through charts and graphs in order to communicate the results of an investigation in an appropriate way.

Time
One double lesson or two 40-minute lessons.

Prerequisites

Notions of Excel or other applications that allow the elaboration of graphs.

School level
Fifth grade and above (Secondary school)

Purpose
To stimulate teachers to apply data analysis through charts and graphs with their students and to develop skills for their formulation and correct interpretation by the reader.

Student outcomes

- To understand how a chart is composed
- To create graphs from data tables
- To present the data in an appropriate way so that the graph represents exactly what we want to convey.
- To recognize the essential and non-essential data in a graph to select what is indispensable for the understanding of what wants to be conveyed.

Introduction

A **graph** is a type of representation of data, generally quantitative, by means of visual resources (lines, vectors, surfaces or symbols), to show the mathematical relationship or statistical correlation among them.

The graph is of high value to both the analyst and the end user, to whom we are communicating a data-driven story. (See Fig. 1).

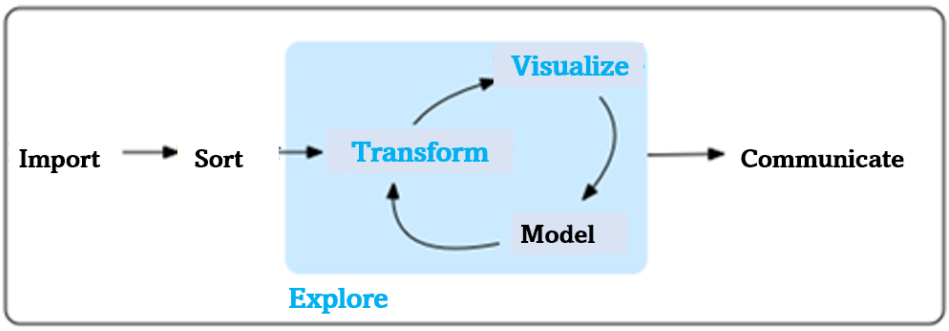

Fig. 1. Representation of a standard process in advanced data analytics. Source: datos.gob.es

Data represents variables (something that changes in space or time) at a particular time and place. They can be qualitative or quantitative. Qualitative data are used to describe the quality of something or qualify the appearance of something, these can be opinions or thoughts that are usually described in words. Quantitative data are those that are obtained from figures, numerical values that can be analyzed using mathematics.

When you define a research based on the hypothesis, the first thing it must be defined are the variables. A variable is a property of the research subject that can assume two or more values. If this does not occur, the observed characteristic is not a variable but a constant. You will have dependent variables then.

The independent variable is defined as the variable that changes or is controlled to see its effects on the dependent variable. For example, in a study investigating the effect of temperature over the course of a day, the hours of the day (or time) will be the independent variable and the temperature recorded will be the dependent variable. The independent variable is not affected by what the experimenter does or by any other variable within the same study; therefore, it is called "independent", controlled or predictive.

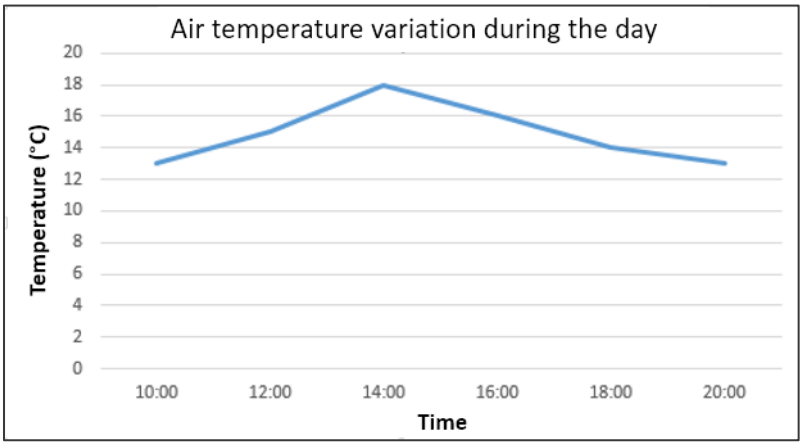

Fig. 2. Graph of air temperature variation on September 23rd, 2020, in Montevideo, Uruguay.

It is represented on the abscissa axis (x) in a graph. Generally, only one independent variable is chosen in a chart to clearly establish that only that variable is the one that will influence the dependent variable; if we had more than one it would complicate the interpretation since the dependent variable would be influenced by more than one cause and we would not be able to determine exactly to what extent each one is influencing it and how it does so.

The dependent variable represents a quantity whose value depends on how the independent variable is modified. It is about the effect, what is measured. It is represented on the y-axis of a graph. It is also known as the experimental, measurement or response variable. For example, in a graph where you want to represent the rainfall that was recorded throughout the month, the dependent variable will be the amount of rainfall recorded, and the independent variable will be the days of that month.

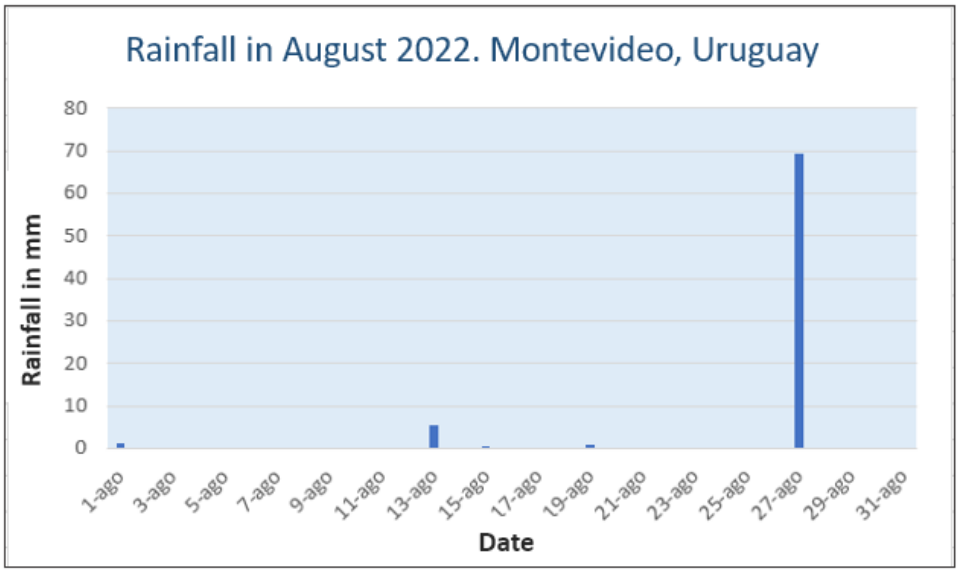

Fig. 3: Graph that shows the rainfall during August 2022.

In the elaboration of the graphs, we must consider including some basic data that contribute to the understanding of the graph:

Title of the graph: it is the presentation of the graph and the indicator of what we expect to find in it. It should be brief, clear and concise. It should be placed at the top in a larger font than the titles of the axes so that it stands out and is centered in the image. It is essential for the understanding of the graph in the context of the report or research.

Axis titles: these are the names of the dependent and independent variables, which will go on the respective axes. In the case of the dependent variable, it will go vertically to the left. In the case of the independent variable, it will be horizontal, below the values of the variable.

Units used: Special care should be taken to clarify the units of measurement in which the variable has been recorded correctly, so as not to leave it to free interpretation and to avoid drawing an erroneous conclusion. For example, in Fig. 3 it is made clear that rainfall was taken in mm (and not in liters). For the previous case in Figure 2 (temperature vs. time), it is clarified that the temperature was taken in Celsius degrees (and not in Fahrenheit).

Scales: The scales of the axes should start at 0 and cover the whole range of data present in the table: for example, if the minimum precipitation value is 3 mm, the graph will show the scale from 0 mm and not from 3 mm onwards. You will also need to scale the axes proportionately to the measurements taken, for example, instead of unit by unit, you may require it to be 50 by 50 or 100 by 100, or 0.5 by 0.5.

Data labels: optionally you can add the label to the data. This means that once we select it, the value of the data will appear above the corresponding bar. All these options are offered by Excel by selecting the area of the chart and clicking on the + symbol that will appear in the pop-up window.

Trendline: it is additional information that we can add to the chart. It is an analysis tool that allows us to visualize points on a chart that go beyond the actual data and represent possible future values according to their trend**.** It offers several layout options: linear, linear extrapolate, two-period moving average, logarithmic, polynomial, exponential.

Grid lines: it is an accessory. We can add horizontal or vertical lines, or both, as the background of the graphic. In papers it is recommended not to add them to make it more easily read.

By selecting the chart area and clicking on the brush symbol, different styles and colors to apply to the chart will appear in the pop-up window. Finally, by selecting the funnel symbol in the pop-up window, you can select which data points and names are visible on the chart and modify the data series and categories.

According to what we want to emphasize in the graph or the type of data handled, different types of graphs will be used. This will allow us to adequately visualize what we want to communicate to the reader. We can choose between the following types of graphics:

- **Bar charts:** or column charts. They only consist of two axes, X and Y. On one of the axes the items to be compared are placed, and on the other one the values of each one are presented. They are easy to understand. An example of a bar chart: figure 3. A variant of the bar charts are the **histograms:** the difference is that in this type of chart you can only see the variation of a variable with respect to the data being analyzed. They have the particularity that there is no separation between the bars.

- **Line graphs:** in these graphs the points will vary on each X-axis, and then are joined with straight lines to show the increase, decrease or equality from one axis to the next. It is one of the most useful because it allows you to analyze and understand how a variable change over time. Example: Fig. 2 shows the rise and fall of temperature over the course of the day.

- **Circular, pie, cake or sectorial:** it is used when you already know the absolute value or quantity of what you are measuring, but you need to know how many parts the total is divided into. It has a geometric style that stands out for its visual simplicity, it is very easy to remember. Example: Fig. 4 highlights the predominance of Eucalyptus over other species in the study site.

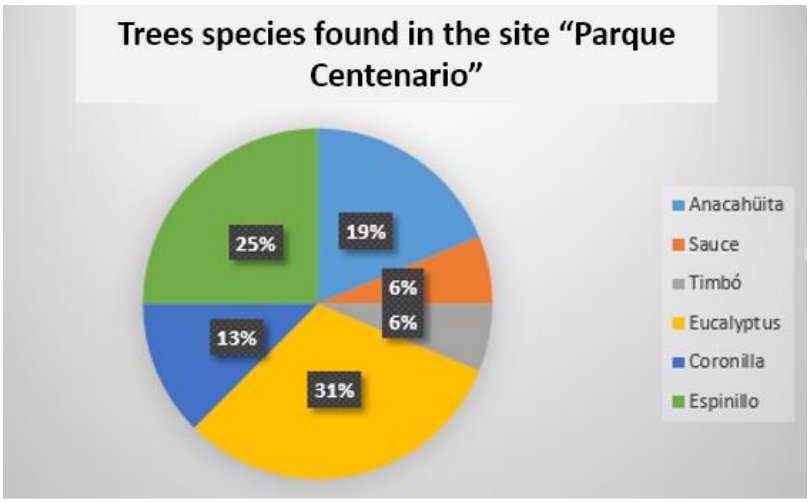

Fig. 4. Example of a pie chart: Distribution of Tree species in the site "Parque Centenario".

- **Scatter plot:** it represents the data obtained through observation, in the form of dots, to then compare them with certain trends**.** It is also known as XY graph and it is used in cases of ordered pairs to know the type of relationship established between variables or when a large number of observations are available.

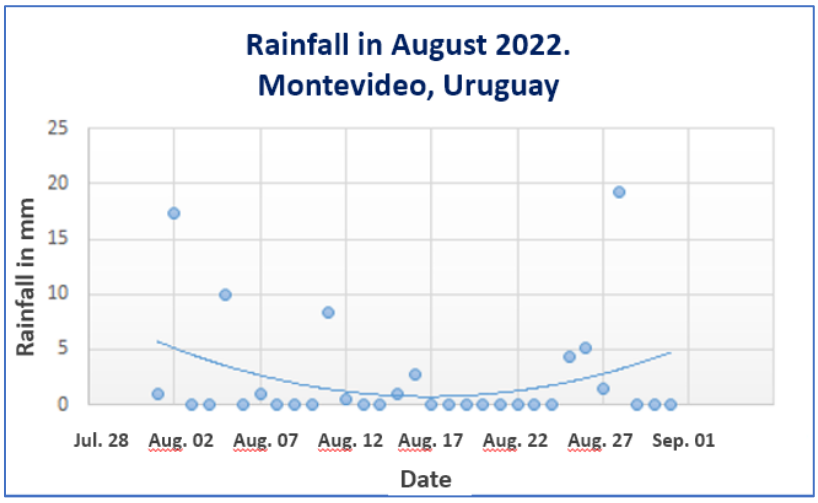

Fig. 5: Rainfall scatter plot.

- **Area graph:** it is used to see the behavior of two or more variables in a period of time. Two axes are used to make it, one vertical and the other horizontal. On the horizontal axis the time measures are placed, and on the vertical axis the values of each variable are placed. First, the variable with the highest value is placed and the points are joined by means of lines or curves, then the other variables are added, from highest to lowest. For a better visualization, the resulting areas among the curves are painted with different colors.

- **Stacked Area Chart:** Used to compare total magnitudes and to show the distribution of a known total. It allows both actions to be achieved at the same time,

combining the bar chart and the pie chart. It can be used to visualize, for example, monthly or yearly accumulated rainfall (see example in Figure 6).

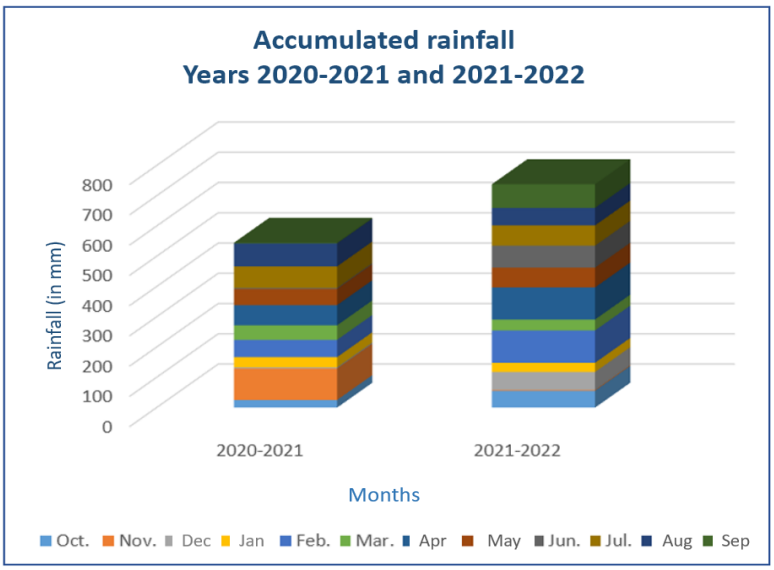

Fig. 6. Accumulated rainfall graph in the site Atmósfera en casa, in Montevideo, Uruguay.

- **Other graphics**: **pyramid (**to represent characteristics of a population of a city or country), **cartogram** (it is made on maps and different marks or references are used to show the results of a certain event or situation), **fluctuation graph**, **bullet graph**, **highlight table** or color map, **tree map**, **box and whiskers diagram**, **candlestick graph, pictogram, arc diagrams, spider graph, 2D and 3D graphics,** etc.

Guiding Research Questions

- What data can we represent in a graph? What are dependent and independent variables? How are they applied in a graph?
- Should the chart have a title? Why?
- What do the X and Y axes represent on a graph?
- Are all charts suitable for all data or are there chart formats that are more suitable for each thing?

Scientific concepts

- Dependent vs. independent variables
- Bar chart, pie chart, line graph, etc.

Materials and tools

- Excel Spreadsheet App
- Legos or other stackable, colorful materials to represent data

What to Do and How to Do it

Beginning -

Depending on the age of the students, you can introduce them to graphing through representation with Legos or any other nesting cubes. It is also popular to graph with food, but other teachers, school administrators, and parents may not like the idea of using food for teaching purposes. Another way to start working on graphing with young children can be to use an open abacus.

Fig. 7. Open abacus for working with graphics with children.

In the case of older students, graph paper will be used for making graphs or, in secondary school, the Excel calculation program can be used in its graphing options.

Development -

- Organize the students into groups, and armed with tape measures and science notebooks, go out to the school yard or a park to measure the height and circumference of the trees in a 10 x 15m pixel. If you already use a site for different biosphere measurements, use the same one as it will be useful to have these measurements in the future.
- Each group chooses a tree, tries to identify it with the help of the teacher or a local flora identification guide, and will measure the height and circumference of the chosen tree. Try to organize as many groups as there are trees, or have some groups measure 2 trees until all the specimens in the pixel are complete. It is advisable to mark each marked tree with a tape or a flag so as not to repeat the specimens already measured.
- To estimate the height, they can use the GLOBE Tree Observer app, and record the circumference there as well. In addition, have students record them in the science notebook.
- Return to class with the students and ask each group to record on the board the height of the tree they measured in one column and the circumference of the same tree in another column. They will make a table with the data of the specimens measured, the heights, and the circumferences.
- Then, in pairs, some students will make graphs comparing species to tree height, and others will make other graphs using the circumference measurements. They can use graph paper or the Excel spreadsheet.

Ending -

All students will display their printed graphs on the board and look at the different ways they each chose to lay out their graph (types of graphs they chose, colors, how they used titles, how they moved data from the charts to the axes of the graph, etc.).

Guided by the teacher, they will decide which graphs were the most appropriate to communicate the information they want to convey: the heights and circumferences of the trees according to the respective species. They will detect errors and lack of relevant information for the correct communication of the data. They will draw conclusions about what they have worked on.

You can create a Padlet and upload the photos of the measured trees, along with their heights and circumferences, and the graphs you created with the measurement data. It will be a record for future research and for the field portfolio.

Frequently Asked Questions

Can I work with graphs with younger children?

Yes, but make sure you use Legos or something similar, that it is easy for them to represent the measured objects and that the measurements are in whole numbers to transfer the data to the Lego cubes.

Suggested resources for further information:

https://www.experimenta.es/uncategorized/build-chrome-juego-construccion-google-lego/

Bibliography

Mejia Jervis, Tatiana (July 22, 2020). Graphics. Lifeder. *https://www.lifeder.com/graficas/*

Risso, Ignacio. August 12, 2022. +20 types of graphs to analyze and present your data properly. *https://www.crehana.com/blog/negocios/10-tipos-de-graficas/*

The importance of graphical data visualization (Blog). 22 August 2018. At: *https://datos.gob.es/es/blog/la-importancia-de-la-visualizacion-grafica-de-los-datos*.