Data literacy 101
How do we set up graphs in science?
BY KRISTIN HUNTER-THOMSON

We live in a data-rich and data-driven society. Not surprisingly, data are now integrated more purposefully into state science standards, as well as the Next Generation Science Standards (NGSS Lead States 2013). We know that data literacy is essential for our students to be successful in the 21st century (Fontichiaro and Oehrli 2016; Gummer and Mandinach 2015). But what is data literacy, exactly? And how do we add it into our already full curriculum?

My working definition of data literacy is “the ability to collect, process, interpret, analyze, and share data for ourselves and others to use and understand.” This definition encompasses the many aspects of working with data, from their initial collection to interpreting and communicating their meaning. This definition also stresses the collaborative nature of working with data; data are simply numbers and text until we make sense of them for ourselves and others.

Every few months, this column will explore ways to incorporate data literacy into our classrooms. We will explore strategies and tips addressing common struggles middle school students have as they build their data literacy skills. Let’s get started!

Data skill: Learning how to set up a graph or map in science
To study data, it is important to first understand what we do with them. There are three main kinds of questions that we ask with data (Berinato 2016):

1. Composition questions seek to make meaning of how the total is portioned into subgroups, or how the parts make up the whole.
2. Distribution questions seek to make meaning of how variable data points are in relation to one another.
3. Comparison questions have two components to them: They can investigate whether two or more groups are similar or different, or whether two or more groups or variables are correlated to one another.

Data can be collected by either your students or someone else, in real time or long ago. Regardless of data collection method, the type of question we are asking influences the kind of data and how much data we will need to answer our question. For example, if you are asking a:

- composition question, you need data on each subgroup and the whole group.
- distribution question, you need the full range of data on a group.

Kristin Hunter-Thomson (hunterthomson@marine.rutgers.edu) is the program director of Dataspire: Inspiring Data and Science Literacy and a teaching instructor of science communication at Rutgers University in New Brunswick, New Jersey.
comparison question, you need enough data on each group or variable to make a meaningful comparison.

Once we have the data, we need to do something with them so that we can better make sense of them. We use graphs and maps in science to organize data and make sense of it. In fact, we are able to make more meaning from data when we work with clear and relevant visualizations or representations (Few 2012; Berinato 2016). However, sometimes knowing where to start in making those visualizations is challenging for students (Webber at al. 2014). Therefore, we need to think of ways to help our students understand the foundation for how we create graphs. If they cannot create the graph effectively (or understand why it is set up the way that it is), then they will be less able to interpret or analyze the data successfully (Fontichiaro and Oehrli 2016).

How we set up a graph or map depends on the kind of question being asked and the graph type being used (to review common graph types, see the “Types of Graphs” sidebar at right). There are, however, general rules of thumb that are good places to start. Keep in mind that, like all rules, there are exceptions to the following.

Composition graphs
If you are looking at the composition of subgroups in relation to the whole group, the

Types of graphs
There are eight common graph types that are used in K-12 science classrooms [i.e., bar charts, line charts, pie charts, scatter charts], appear in the math standards and on science assessments [i.e., the previously listed types of charts, plus boxplots and histograms], and are often used in news media [i.e., the previously listed types of charts, plus stacked area charts and bubble charts]. Here is a quick refresher on what we use each graph type for:

- **Bar chart**: Plots the quantitative values for different categories of a group or variable. This lets you compare summary data across categories. Another variable can be plotted as an additional set of bars.
- **Line chart**: Plots a quantitative attribute in which there is a specific order to the data [e.g., ordered by time] along the $x$-axis and another quantitative value along the $y$-axis. These charts show local changes in data. The line indicates what a data value could have been if it had been collected.
- **Pie chart**: Plots how the quantitative values for categories make up the whole. This lets you explore the proportional relationship across the categories for how they make up the whole.
- **Scatter chart**: Plots quantitative values along each of the axes to investigate the relationship between the variables. A third attribute may be included within the formatting of the data points.
- **Box plots**: Plots the distribution and shape of quantitative values for one variable or different categories of a group. This lets you explore the range and the summary statistics of your variable.
- **Histograms**: Plots by groups the frequency of measurement and distribution of quantitative values for a variable. This lets you compare the frequency of different groups of values and the shape of that frequency.
- **Stacked area charts**: Plots a quantitative attribute in which there is a specific order to the data [e.g., ordered by time] along the $x$-axis and another quantitative value along the $y$-axis for different categories. This chart looks at both the local changes of data and the magnitude of changes across categories. This lets you examine how individual categories make up the whole over the ordered variable.
- **Bubble charts**: Plots quantitative values along the $x$- and $y$-axes, as well as the size of the data points, to investigate the relationship among the variables.
whole group cannot equal more than 100%, so you need to make sure you are plotting the proportion or percentage values for each subgroup rather than the measured values. For example, if we were comparing the types of birds that collided with windows in an experiment, the percentage of the two types of birds combined could only equal 100% of the birds in the experiment (Figure 1).

**Distribution graphs**

If you are looking at the distribution of a group, the categories of measured values should be on one axis and the frequency, or actual measured values, on the other axis (it does not matter which axis these are on, but they must be on different axes). For example, if we were looking at the range of heights of black cherry trees measured at school, we could use a histogram to see frequency of different heights measured (Figure 2A). Or, if we were looking at the different aspects of the range of measurements of carbon dioxide levels observed at Mauna Loa, Hawaii, in June, we could use a boxplot (Figure 2B).

**Comparison graphs**

If you are making a comparison among items (categories), the measured or summarized values should be on the y-axis and the items should be on the x-axis. For example, if we were comparing the average number of each kind of plant that herbivores ate in an experiment, we could use a column chart (Figure 3A).

If you are looking at an item over space, then in essence, latitude is on the y-axis and longitude is on the x-axis (which is how a map is typically oriented). To make comparisons across that space, the other variable should be represented in a way that indicates the measured values of the item or variable (e.g., dots, colored-in geographic areas). For example, if we were comparing the incidence rate of Lyme disease across states in 2016, we could color in the different states to represent the incidence rates (Figure 3B1). If, however, we are comparing magnitude of earthquakes across the United States for the past 30 days, the dot color could represent age and the dot size could represent magnitude (Figure 3B2).
If you are making a comparison between two variables, neither of which is an item or category, then each variable should be on one of the axes. When graphing data from an experiment that has a dependent and an independent variable, the independent variable should be on the $x$-axis and the dependent should be on the $y$-axis. For example, if we are looking for a relationship between ocean water temperature and salinity between January and April 2018, we could make a scatterplot of the measured data to compare how the values of each variable change in relation to one another (Figure 4A).

If you are making a comparison between two variables, and again neither are an item or category, but one variable has an order to its values, then we use a different kind of setup. For example, if your independent, ordered variable is time, then time should be on the $x$-axis, the measured values for the dependent variable should be on the $y$-axis, and the data points should be connected via a line. We connect the data points across time because time is a continuous variable (there is a never-ending number of possible values between each number) with a specific order (e.g., 12:05 p.m. always comes after 11:55 a.m., and there is time between those two values of recorded time). The key here is that although you may not have taken a data point from a time between 11:55 a.m. and 12:05 p.m., you could have, so we use a line to show that in the graph. If we
were investigating the percentage of male and female tuataras over time from a field experiment, we would plot our data on a line graph (Figure 4B).

**Conclusion**

A key first step in setting our students up for success in interpreting and analyzing data is to help them create graphs and maps in ways that will help them make meaning from the data. We need to teach our students not only how to make sense of graphs and maps in science, but also why we put them together the way that we do. This will help our students gain the necessary data literacy skills to make their own decisions regarding creating and understanding data visualizations.

In January, this column will explore how to help students learn which graph type is best, based on their data and question.

**REFERENCES**


**RESOURCES**


Centers for Disease Control and Prevention Lyme disease data tables—www.cdc.gov/lyme/stats/tabs.html

“Deadly windows” data—http://datanuggets.org/2017/05/deadly-windows

“Do insects prefer local or foreign foods?” data—http://bit.ly/2y2a8Cu


National Oceanic and Atmospheric Administration System Wide Monitoring Program—https://coast.noaa.gov/swmp/#/index