Teacher Tune-up

Quick Content Refresher for Busy Professionals

Misleading Representations of Data

An important part of being a scientist is being able to communicate one's ideas, research, and findings. While we put a lot of emphasis on the verbal side of communication—writing, argumentation, and how to think like a scientist—the secret weapon for budding scientists is the ability to shape their data into the right visual form to tell the story as clearly as possible.

Distortions of data obscure a scientific narrative. While some people may intentionally distort data to persuade someone to agree with a particular point of view, any scientist's goal should be clarity of data. Misleading representation of the data is a cousin to bad data, and neither should have a home in your classroom.

Unfortunately, the way that data visualizations are often taught in elementary schools can breed some bad habits you may need to undo with your students. We have put together a list of Dos and Don'ts to help navigate good and bad practices in visual representations.

Show everything, including outliers.

It is usually a good idea to include all the data in your graph. If students make the graph *by hand*, this documentation can be quite a chore. Deciding on a scale by hand is a serious application of number sense and use of proportional reasoning. As a practical matter, though, unless you're using graph-making as a lesson in itself, consider telling the students what limits and scale to use, or leading the whole class through the decision-making process. If, on the other hand, students use *technology* to graph, life is easier; but students still need to recognize the importance of scale. Most systems have some sort of zoom feature that makes it easy or even automatic to include all the data.

Show your scale.

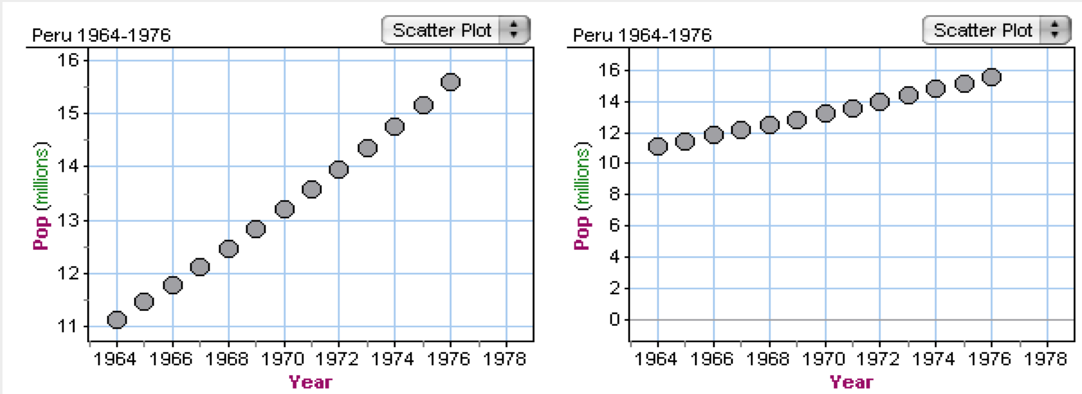
Choose the right proportions to show a trend accurately. Don't skip over numbers. Don't cut off or raise the bottom of the graph. Keep your scales linear; if you must make a scale logarithmic, make it very clear that you are doing so.

Students often have trouble deciding on the scale for their graphs. When they make graphs by hand, this amounts to figuring out how long to make the axes on their paper and how to label those axes, which also means deciding how much each square or graph "tick" represents. It's actually pretty complicated and can depend on the context—for example, whether they are using technology to make the graph, and what the data are about (see *zero suppression* below). As a teacher, you will have to figure out on the fly what students understand and how far you can nudge them.

Choose what the right minimum and maximum values should be.

Students won't call it this, but teachers might benefit from knowing the term *zero suppression*. Should your axis range always include zero? Not always, but in certain circumstances, you must include zero. Consider these two graphs.

The Population of Peru, 1964–76



*The graph on the left, using the software's default scaling, does not include zero on the vertical axis. If you don't notice this, it looks as if the population doubled between 1964 and 1965. In the graph on the right, zero is included. We see that the population is increasing less dramatically that we might have thought. So it's a fairer visual representation of the overall population. On the other hand, if you wanted to see greater detail about the change in population—and recognized that the zero was suppressed—the one on the left is better. Notice that we suppress the zero on the horizontal axis without even thinking: no one considers that the **Year** axis should start at zero.*

Use the right scale on each axis.

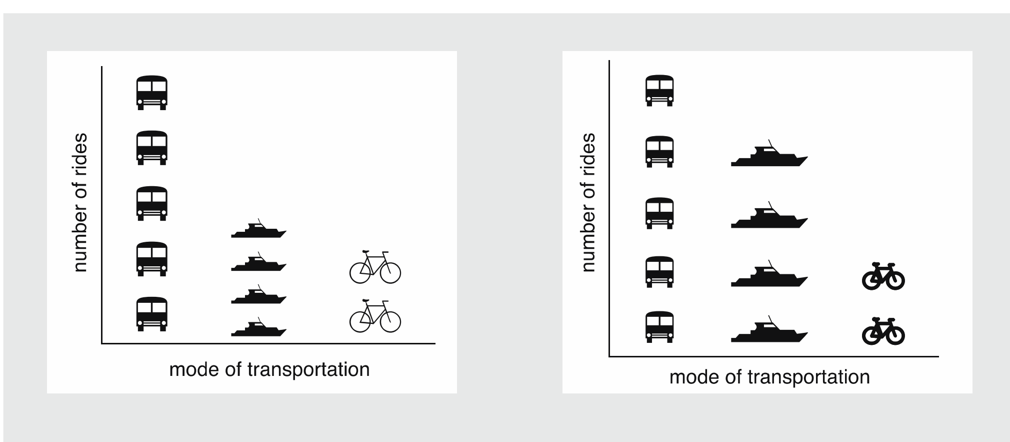
If you're making a plot (such as a scatter plot) that lets you see the relationship between two variables (such as a trend over time), your two axes will probably not be scaled the same. If you're using technology, you have to learn how to scale axes independently. This process might involve dragging things associated with axes (e.g., using programs such as Tuva, CODAP, TinkerPlots, Fathom) or assigning minimum and maximum values (e.g., with Desmos, graphing calculators). Teachers need to be familiar with these controls in advance.

Say no to pies.

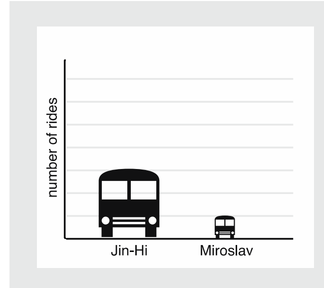
Tables or bar charts are easier to read and interpret than pie charts, in part because we can more easily "read" quantity linearly rather than angularly. Save your pies for dessert, not data.

Resist using pictograms.

If the student has their heart set on using pictograms, line up the icons, as in the graph on the right, and/or use icons of similar sizes.



Don't use pictographs.



These are the types of graphs that use illustrations rather than simple geometric shapes to represent value. If for some reason your students insist on using illustrations, at minimum coach them to scale the illustrations on both axes, using the total **area** as the value, not just the height or width of the illustration. Doing it the other way makes a value that is two times as large appear to be four times as large. We assign value, psychologically, based on the overall apparent "visual mass" of objects in graphs. So if we see several different shapes in a graph, we unconsciously compare their *areas* rather than their heights or widths.

Stay two-dimensional.

Your program may give you an option to make your graphs three dimensional. Stomp out this urge. Don't make your graphs 3D. Just as we intuitively assign value based on the overall apparent visual mass of objects in pictographs, when we see something that looks 3D, we compare volumes. Therefore, avoid 3D, and avoid using scaled pictures to represent quantities.

Use areas only when appropriate.

While we don't usually recommend using anything beyond very simple geometric dots, lines, and bars to communicate simple data sets, in some situations data can take advantage of the visual mass phenomenon to good effect. Consider this graph of the world, which shows area of countries scaled by population. We do not expect students to make such a map, but we can ask them to understand it and answer questions based on it.

After introducing the concept of misleading representations of data, you could follow up with a discussion of the more general topic of propaganda and how people "lie with statistics." Distinguish between misuses of statistics that are apparently intended to convince the reader of something and the related, less nefarious category of mistakes and misconceptions. They are related, but different. In the case of persuasively misrepresented data, your aim is to inoculate students against being deliberately deceived, often by published visualizations. For the second, help students learn to communicate results fairly and correctly, often using their own data.

Finally, if you enjoy thinking about these topics, take a look at the books of Edward Tufte, an economist cum artist who has made a career of critiquing data design.