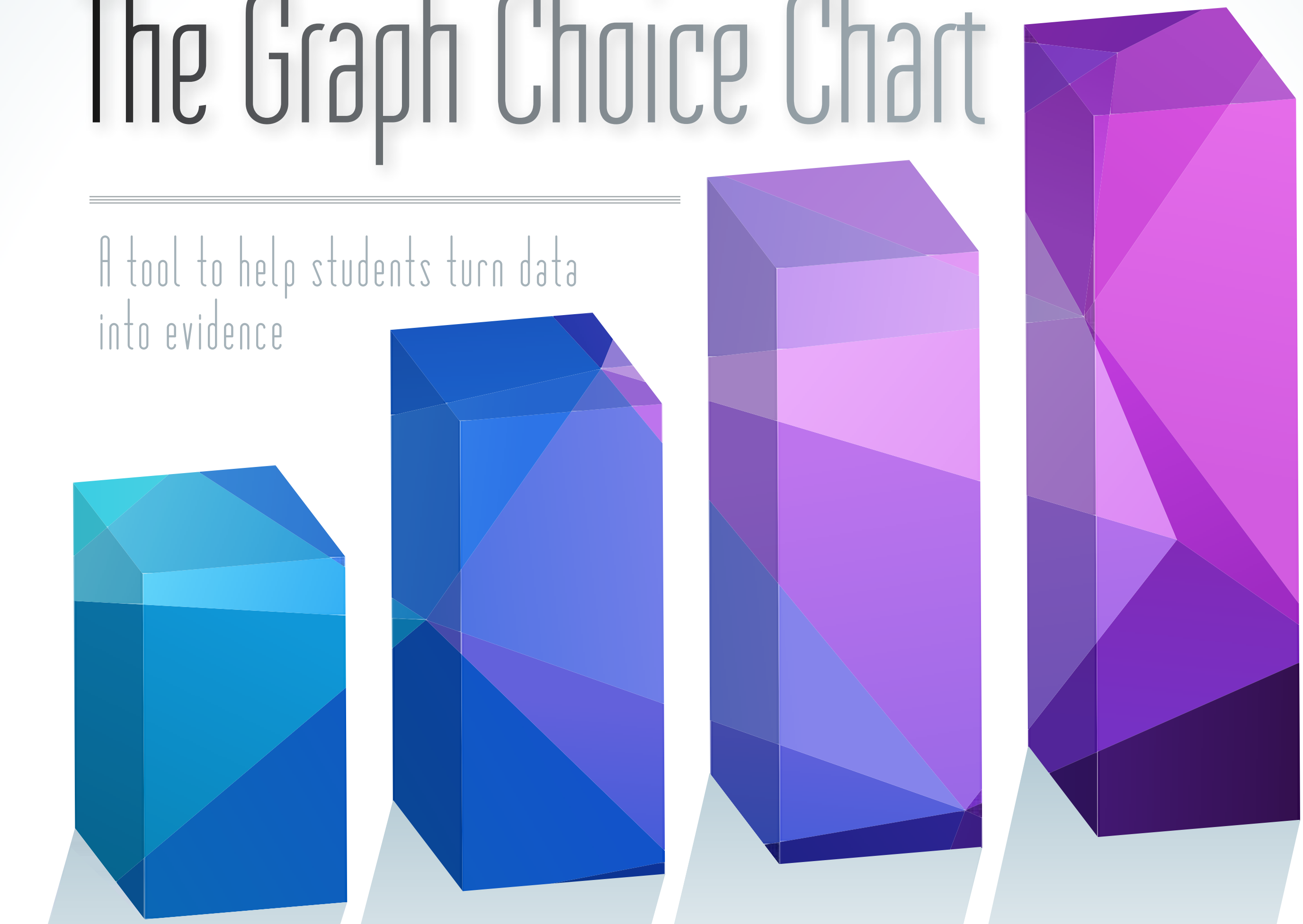
The Graph Choice Chart

A tool to help students turn data into evidence



Hannah Webber, Sarah J. Nelson, Ryan Weatherbee, Bill Zoellick, and Molly Schauffler

Data literacy is complex. When students investigate the natural world, they must be able to gather data, organize it in tables and spreadsheets, analyze it in context, and describe and interpret it—usually as evidence to support a scientific argument (Jiménez-Aleixandre, Bugallo Rodríguez, and Duschl 2000; Kilpatrick 1985; Schoenfeld 1992).

These skills are echoed in the science and engineering practices of the *Next Generation Science Standards:* "Because raw data as such have little meaning, a major practice of scientists is to organize and interpret data through tabulating, graphing, or statistical analysis. Such analysis can bring out the meaning of data—and their relevance—so that they may be used as evidence" (NGSS Lead States 2013, Appendix F, p. 9).

But before students can identify patterns in data or use it as evidence, they must be able to graph it.

In 2007, we began working with scientists and teachers in Maine to explore students' data literacy skills. We found that when students began to organize, graph, and interpret their data, many were unsure about what kind of graph to make. Most made bar graphs, regardless of their research question. They also treated the graph like an end product in itself, instead of using it to see patterns and make arguments. Although students had the mechanical skills to generate graphs, they did not logically decide what kind of graph would best suit their particular research question.

Consequently, we developed the Graph Choice Chart (GCC), a tool to help students choose the appropriate graph.

This article describes the GCC and gives examples of how our partner teachers used it in their classrooms.

Background

Early in our project, we surveyed more than 200 high school students and asked them to draw graphs to illustrate simple comparisons between two groups and the relationships between two variables (Figure 1). In the first part, we asked students to draw a graph to help them determine whether the type of stream bottom—rocky or muddy—affected dragonfly abundance. The second part asked them to graphically show the correlation between fish size and the concentration of mercury. In the case of the dragonflies, only 23% of students made a graph—a frequency plot or a bar graph of group averages—that visually compared dragonfly abundance in the two habitats. In the fish example, only 58% of students correctly made a scatterplot to display the correlation between mercury concentration and fish weight. Based on our follow-up interviews with students, we concluded that, for many, the question "What kind of graph should I use?" did not occur to them.

Thus, the GCC we created takes the form of a decision tree, where a choice at each node, or decision point, leads to other choices and finally, to an outcome, or type of graph, for each branch (Figure 2, p. 40). This helps students make an informed decision about what kind of graph to use.

Focusing on the research question

The starting point for the GCC—and a requirement for it to work—is a precisely worded research question. Writing the question forces students to be clear and consistent—and to stick with one question—as they move through their analysis. Changing the wording of a question midstream can produce a different kind of graph and cause confusion. In the classroom, our partner teachers find that much of this confusion can be resolved by having students reconsider their research questions. The process of fitting a graph to a question encourages them to think more deeply about their data as they develop a claim or argument.

Classroom example

For example, one partner teacher works with her students to locate bird nests in a forest and measure the distance from each nest site to the nearby lakeshore. After looking at their data table and the bar graphs some draw, students conclude that birds build nests closer to the water because there may be more predators in the deep woods, and thus it is safer by the water—a conclusion that takes leave of the data and ventures into speculation. The following dialogue demonstrates how the teacher used the GCC to steer her students to a question that can be supported with the data collected:

Teacher: Okay, so what was your research question?

Student: Oh… (long pause). It's about the relationship between nests and distance to shore.

Teacher: How would you word that as a question?

Student: Umm…What is the relationship between nests and distance to shore?

Teacher: Okay, what kind of question is that? Use your GCC.

Student: (The student studies the chart.) It's like a correlation question.

Teacher: All right. A correlation question involves two numeric variables. What are the two variables you measured?

Student: (Long pause.) We measured distance to the water and… that's all.

Teacher: So just one variable then. It sounds like maybe you are interested in what the distribution of nests is with regard to distance to the shore. Just one measured variable (i.e., distance) and one group (i.e., bird nests).

Student: So… (studying the GCC)… a frequency plot! That would show us how the bird nests are spread out in distance to the water.

In this example, the teacher realized that the student had lost track of her question and that prompting her to articulate it and use the GCC to reason through what kind of graph to make might result in a much richer (and clarifying) discussion about the data as evidence. The movement between the question and the graph choice is not unidirectional; thinking about the kinds of data needed for various graphs and the kind of data available enables the student and teacher to move among the framing of the research question, the nature of the data collected, and the kind of graph that might address the actual question the student has in mind.

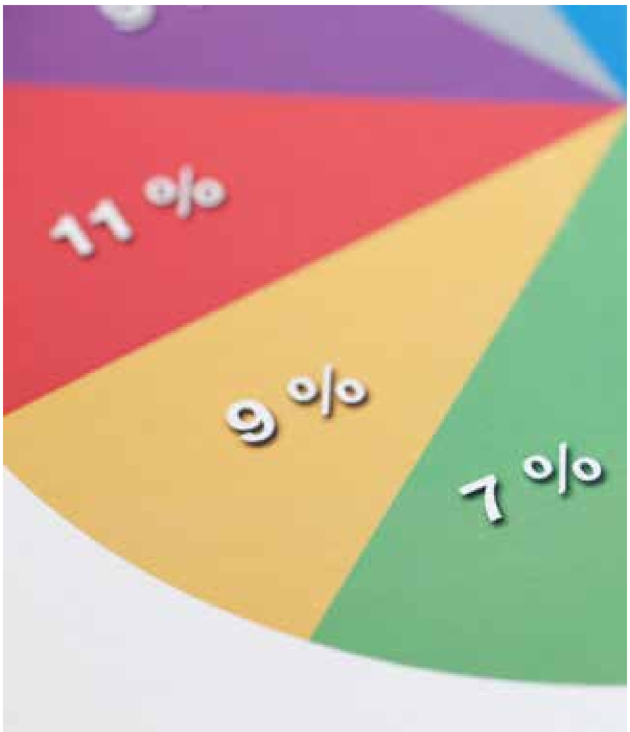
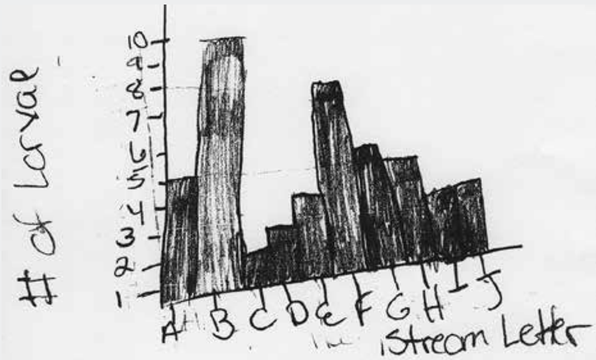


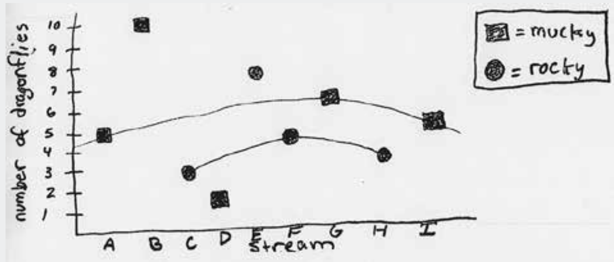
FIGURE 1

Examples of high school student comparison and correlation graphs

1 a. Dragonfly data graphs.

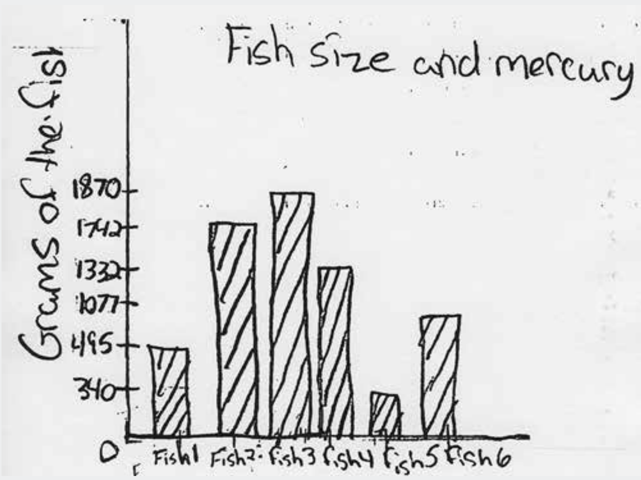
In a survey, students were given a table of data about dragonflies collected from rocky and mucky streams and the following prompt: “Draw one graph showing the data in a way that helps you figure out if the type of stream bottom has to do with dragonfly abundance (e.g., the number of dragonflies).” Seventy-seven percent of students made a graph that did not compare groups.





1 b. Fish data graphs.

Students were given a summary of two positively correlated variables and the following prompt: “Draw a graph to display the following data so that you can see if fish size and mercury are correlated. Don’t leave any fish out.” Forty-two percent of students made a graph that did not display a relationship.



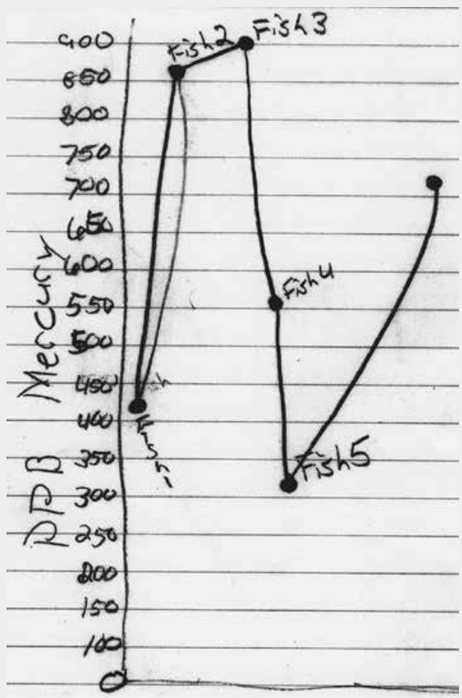
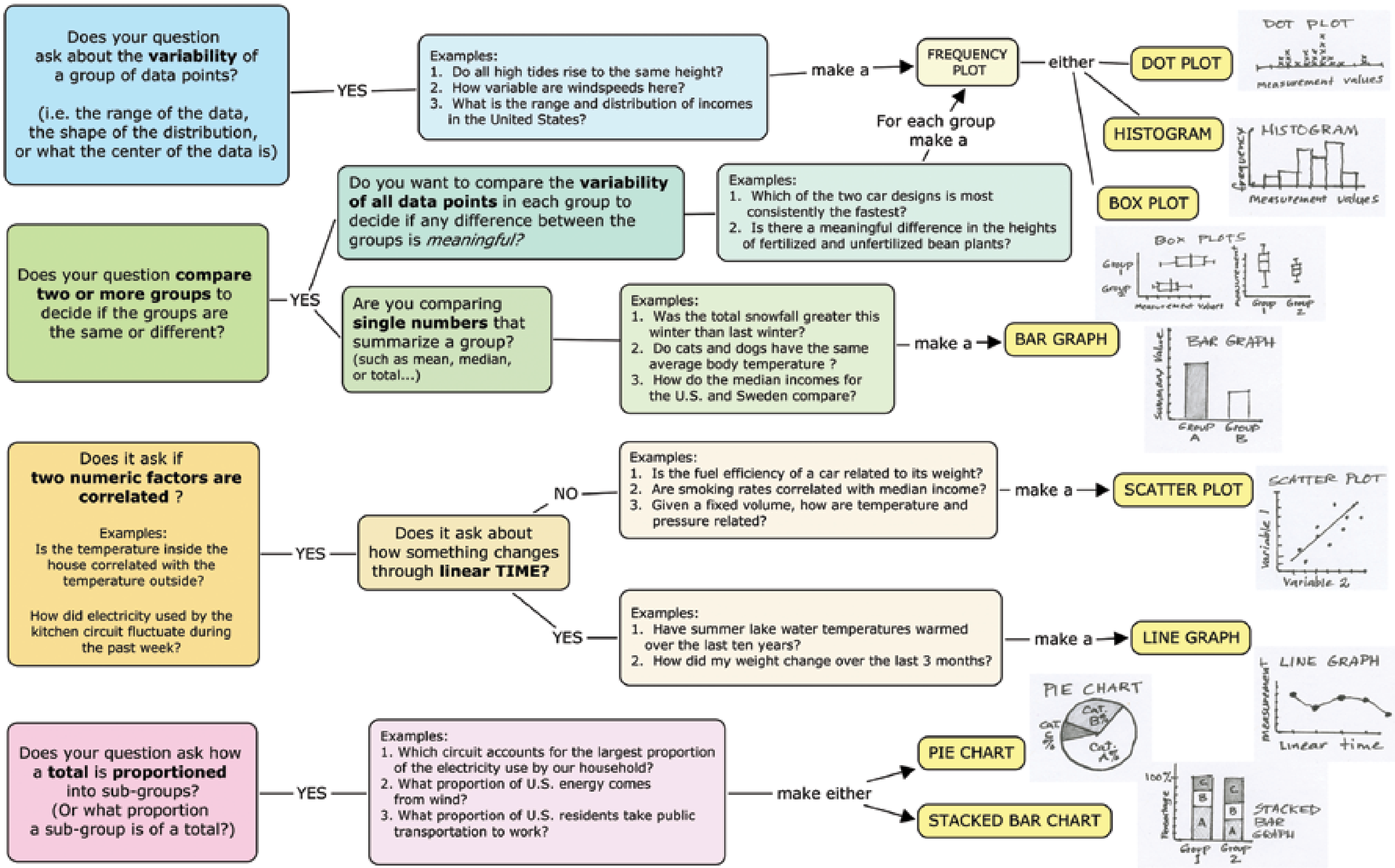


FIGURE 2

The Graph Choice Chart.

What question would you like to explore? Write your question as a complete sentence.



Choosing a graph type

Once students state their research question, they can use the GCC to identify the type of question they are asking and then link the question type to their choice of graph. The boxes on the left of the GCC are choice-points for identifying the type of question the student is asking (Figure 2). They include:

- Does your question ask about the *variability* within a group of data points? (one group, one variable);
- Does your question *compare two or more groups* to decide if the groups are the same or different? (two groups, one variable);
- Does your question ask if *two numeric factors are correlated?* (one group, two variables); or
- Does your question ask how a *total* is *proportioned* into subgroups? (Or, what proportion a subgroup is of a total?) (one group with subgroups, one variable).

If a student's question does not clearly fit with any of these choices, the example questions can be used to help the student rephrase (Figure 2). Once students have identified their question type, they follow the decision-making tree. Students answer the yes-or-no questions, review the examples for similarity to their own questions, and then determine a suitable graph type. (A sample GCC on the question "Has the bloom time of forsythia changed in the state of Maine over the last 30 years?" is available online; see "On the web.")

Classroom example

Some of our partner teachers also use the GCC in reverse. For example, using physical data collected by a balloon ascending through the atmosphere (see "On the web"), teachers ask their students to develop a question that can be answered with these data. Without the GCC, students ask such single-point questions as "What is the average air speed?" "How high does the balloon go?" or "Where did it land?" Students move beyond such questions using the GCC in reverse—starting with different graph types and working backward to see what kind of question might lead to that graph.

For example, students can discuss each question type and then, using the balloon data, write one question of each type on a 3 × 5 card (except proportional questions), with the writer's name on the back. Students then put their questions in a pile. Each student chooses a question and, again, using the GCC, determines the appropriate graph type to answer that question. If the question is worded so that it is hard to determine the appropriate graph, students ask the writer to explain and rewrite the question.

Refining students' process

Single-number comparisons and variability

Without prior instruction, few of our project students make

FIGURE 3

Average kWh data.

What was the mean kWh used per month by U.S. households in 2009? (Data source: *EIA.gov*)
The question can be addressed by either a bar graph or dot plot. The dot plot inspires discussion and new questions more than the bar graph does because it shows the variability among the states. (The vertical line in the dot plot is the mean.)

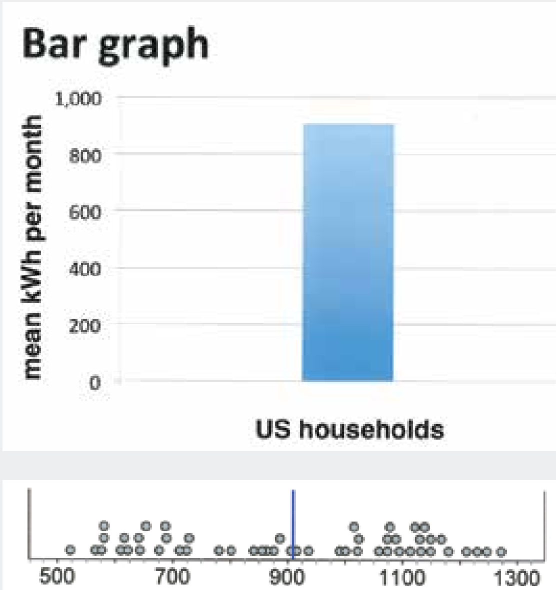
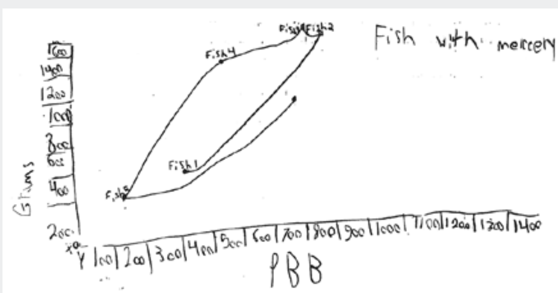


FIGURE 4

Scatterplot.

Many high school students inappropriately connect points in a scatterplot. In this case, a ninth-grade student connected the dots in the order that they appeared in the data set.



a frequency plot. Yet, frequency plots are often best when determining whether there is a meaningful difference between groups, as in Figure 3 (p. 41), where students explore what the mean kilowatt-hour per month was for U.S. households in 2009. The GCC moves students toward graphical displays of variability when such data are available and the question warrants (e.g., "Which solar car has more consistent race times over 10 trials?") and confines the use of bar graphs to comparisons between single numbers such as mean, median, and sum (e.g., "Was total rainfall greater in July or August?").

Classroom example

As previously noted, our early survey work showed that bar graphs were the first choice for many students, regardless of their research question. Teachers using the GCC help students understand why other kinds of graphs are useful and when to use them.

To explore variability in chromosomes, for example, one teacher gives students a table showing the number of chromosomes for a variety of animal and plant species and asks: "Do plant species tend to have more chromosomes than animal species do?"

This teacher finds that students tend to graph these data in one of two ways: They either calculate the average number of chromosomes for each kingdom and plot the averages as two bars on a graph, or they graph every organism as an individual bar, resulting in a graph with too many bars to enable easy comparison. (See "On the web" for more on this lesson.)

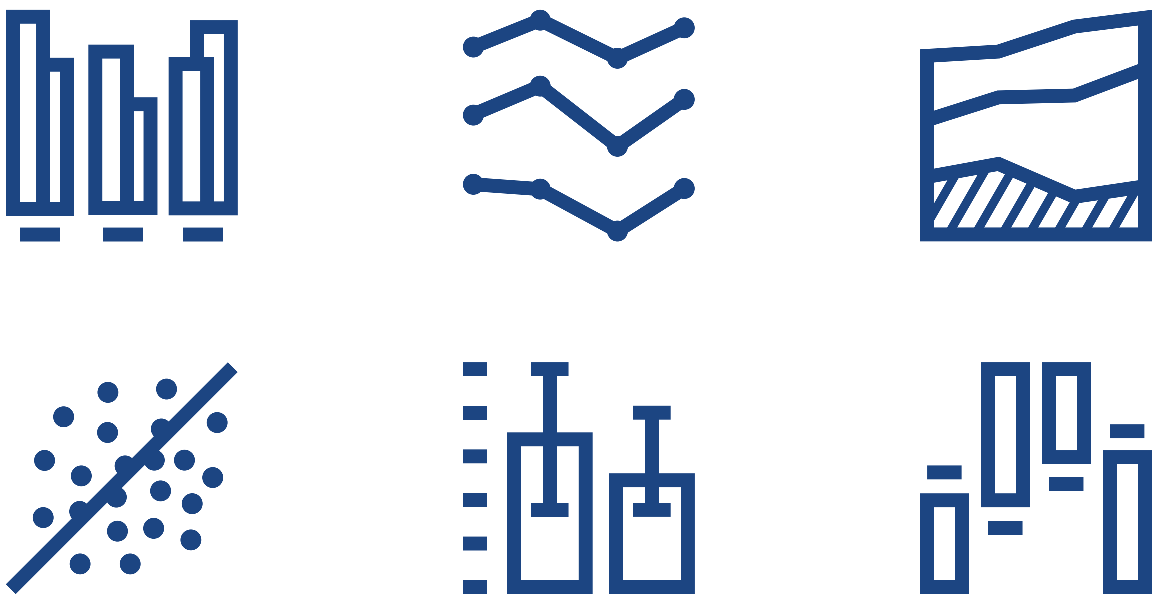
Students typically state that plants have, on average, more chromosomes than animals, or they find that they cannot make a statement about the groups because plotting all of the data points creates a confusing graph.

The teacher then asks students to use the GCC and consider the following: "Does the question ask for a comparison of single numbers that summarizes the two groups, or does it ask for a comparison of groups of data points?" Although students are reluctant to give up the single-number average or any of the individual data points, after a discussion, they often decide that the word "tend" in the research question suggests that they should not reduce these data to an average. Using the GCC, students decide to use boxplots instead.

Once the data are graphed as boxplots, students discuss the graphed evidence with a richness and nuance that cannot emerge from a bar graph. For example, one group said: "There is a lot of overlap, and the median for animals falls within the interquartile range for plants. Perhaps there is no real difference, but the boxplot shows that animals tend to have slightly more chromosomes. The field horsetail and rattlesnake fern are extreme values that really raise the average for plants."

Correlation and time series

We also find that students generally do not consider whether it makes sense to connect data points in an *x-y* plot. When choosing a graph type to show correlation (e.g., a scatterplot), many students incorrectly connect the dots instead and produce a line graph (Figure 4, p. 41). However, it does make sense to connect data points when students are graphing change through continuous time. To help them make a deci-



sion about when to connect data points, the GCC separates time series from other correlations.

Proportion

Students also tend to incorrectly use pie charts to answer questions about group comparisons or correlations. To help students recognize when pie charts are appropriate, the GCC has a fourth kind of question: questions about proportions of a whole.

Teacher findings

Many partner teachers report that before using the GCC, their students did not know that so many different types of graphs existed. They say that students routinely pull out their GCCs, originally distributed in science class, in their math classes as well.

Teachers also find that students tend to start projects with vague, unformed questions. For example, "My question is about loons" could be: "How do loon populations on Eagle Lake and Moosehead Lake compare?" "Is loon population correlated with lake size?" or "Has the summer loon population changed through time?" Each question is about loons and yet each warrants a different graph.

Teachers also find that the GCC forces students to work on one question at a time. One teacher uses the motto "one question, one graph" to help students refine a compound question into two single questions. For example, "Does air temperature change more under the trees or in the open and does the difference change through the season?" becomes "How does air temperature change under trees compare with air temperature change in open fields?" (group comparison) and "How does air temperature difference in two locations change through the season?" (time series).

Another finding is that some students resist the thought-process altogether and simply jump to their favorite graph on the right side of the chart—usually the bar graph—regardless of their question. To emphasize the thought sequence, one teacher slides a piece of paper over when moving from left to right on the GCC and discusses each "column" or decision point with students, referring back to the written question before moving along.

Conclusion

The GCC sets up a framework for thinking about data analysis that is based on real questions and reasoning, rather than an absolute set of steps. Feedback from partner teachers who have used the GCC is positive and often enthusiastic. They indicate that students feel empowered when they realize they have a choice about what kind of graph to use.

Over the last several years, we have worked with partner teachers and their students to develop a set of probes and rubrics for use with the GCC to make formative assessments about students' proficiency in selecting appropriate graphs. Over the coming year, we will continue to work toward validating and assessing the reliability of these instruments.

The GCC, however, is not perfect. It does not represent hard and fast rules for graphing. But at its core, it helps students initiate a process of reasoning about graphing based on purpose, rather than on didactic instruction. Once students master linking data analysis to their research question, they can move beyond the GCC and combine options based on reasoned decisions. ■

Hannah Webber (hwebber@schoodicinstitute.org) *is education projects manager at the Schoodic Institute in Winter Harbor, Maine; Sarah J. Nelson* (sarah.j.nelson@maine.edu) *is associate research professor in Watershed Biogeochemistry at the University of Maine in Orono; Ryan Weatherbee* (ryan.weatherbee@umit.maine.edu) *is a research associate in the Satellite Oceanography Data Lab at the University of Maine in Orono; Bill Zoellick* (bzoellick@schoodicinstitute.org) *is education research director at the Schoodic Institute; and Molly Schauffler* (mschauff@maine.edu) *is an assistant research professor in Earth and climate science at the University of Maine in Orono.*

Acknowledgments

The Graph Choice Chart is part of a larger data literacy framework developed in our Maine Data Literacy Project, which is funded by two Title II Math Science Partnership grants to the Schoodic Institute through the State of Maine Department of Education. We wish to acknowledge the invaluable contributions of our partner teachers as well.

On the web

Animal and plant species chromosomes: *http://participatoryscience.org/data-activity/practice-comparing-groups-chromosome-number-data*

Balloon ascent data: *http://participatoryscience.org/data-activity/practice-asking-questions-balloon-ascent-data*

Graph Choice Chart: *http://participatoryscience.org/file/graph-choice-chart*

Sample Graph Choice Chart: *www.nsta.org/highschool/connections.aspx*

References

Jiménez-Aleixandre, M. P., A. Bugallo Rodríguez, and R. A. Duschl. 2000. "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education* 84 (6): 757–792.

Kilpatrick, J. 1985. A retrospective account of the past 25 years of research on teaching mathematical problem solving. In *Teaching and learning mathematical problem solving: Multiple research perspectives,* ed. E. A. Silver, 1–16. Mahwah, NJ: Lawrence Erlbaum Associates.

NGSS Lead States. 2013. *Next Generation Science Standards: For states, by states.* Washington, DC: National Academies Press.

Schoenfeld, A. H. 1992. Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In *Handbook for research on mathematics teaching and learning,* ed. D. Grouws, 334–370. New York: MacMillan.