DATA *Nuggets* guide to graphing

The purpose of a graph is to visualize the data in a meaningful way, convey information to other interested parties, and provide as a way to support claims using evidence. Graphs simplify complex data into one image that can be used for interpretation and constructing explanations. An effective graph will allow someone who views it for the first time to easily identify each feature of the graph.

There are many different kinds of graphs, each appropriate for visualizing different types of data. Students should identify the graph type that is most appropriate for adequate display of the data they wish to graph. Below is a list of the graph types most commonly found in Data Nuggets:

- Bar graphs - suitable for when you have a categorical independent and continuous dependent variable. Used to make comparisons among groups and to represent data using summary statistics (mean, standard error).
- Line graphs (scatter plots) - suitable for when you have continuous independent and dependent variables. Points are plotted on the x-y coordinate plane. Lines of best fit are often added to examine trends in data, or the relationship between two continuous variables.
- Histograms - suitable for showing the distribution of continuous data. Breaks data into equal interval bins on the x-axis, with frequencies on y-axis.
- Pie charts - suitable for showing data that are parts of a whole. Used to visualize the proportional size of categories, out of 100%.

After choosing the appropriate graph type, the first step is to label the axes with the appropriate variables, and draw the scale. When done correctly the graph will have the following features:

- The independent (predictor) variable is placed on the x-axis (bar and line graphs only).
- The dependent (response) variable is placed on the y-axis (bar and line graphs only).
- Scale is present and intervals are uniform. Each tick along the axis represents the same number of units. These scales are most likely going to be different on the x and y-axes. A break can be added to the axis to represent a big jump in the scale, but it must be notated as a break.
- It is not necessary to label each interval on the scale. Labels can identify every five or 10 intervals, or whatever is appropriate.
- The graph makes clear whether the origin represents the point (0,0) or a different point along the scale.
- Units of measure included with axis labels, when applicable.
- Includes key/legend to differentiate variables or categories on graph, when applicable. In this case, each line must be clearly differentiated from the others - by a label, a different style (ex. dashes), or colors, indicated by the key/legend.

Now it's time to graph the data. The data will typically be presented in a table, and students will have to choose the appropriate data from the table for their graph.

- In some cases, it will be necessary to calculate summary statistics before graphing, such as a measure of central tendency (mean, median, mode). The students will then plot these statistics on their graph.
- For a scatter plot or line graph, a trend line (line of best fit) can be added after the points are graphed. The line should not be extended to the origin if the data do not start there. In addition, the line should not be extended beyond the last data point (extrapolation) unless a dashed line (or some other demarcation) clearly indicates that this is a prediction about what may happen.

Students can also identify and graph variation around the center point (mean, median, or mode in bar graph, or line of best fit in scatter plot). This variation can be quantified by several statistics. Students should include error bars on their graphical displays when appropriate and required by the teacher.

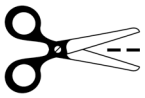
- The simplest measure of spread or variation in a data set is the range, which is the difference between the largest and smallest values in the data set. A discussion of range can help draw student attention to not only central tendency in the data, but also variation around the mean. Be sure to note with students that though range is easy to calculate, it can be misleading; one outlier can make it appear the dataset has a much wider spread than is a true representation
- Standard deviation (SD) is the most common measure of variation for normally distributed data. SD measures the average distance of all values from their mean. The smaller the bars, the less variation around the mean. It is calculated by taking the square root of the average of the squared deviations of the values from their average value.
- Standard error (SE) is the SD divided by the square root of the study's sample size ($SE=SD/\sqrt{n}$). Unlike SD, SE reflects uncertainty in our estimate of the mean. The larger our sample size and the less variation in the data, the more confident we can be in our estimate of the mean.
- Upper error bars are calculated by adding one SE or SD to the mean, and lower bars are calculated by subtracting one SE or SD from the mean.

Once students are done creating their graph, it's time to use it to answer the scientific question! A scientific question is answered by a claim supported by evidence. The evidence comes from the data plotted on the graph (or students can even refer back to the original table). The first step when interpreting the data is to identify any changes, trends, or differences seen on the graph. Students can draw arrows to point out what they see, and write one sentence describing what they see next to each arrow. Next, these observations can be used to construct an explanation. Explanations include a claim, the evidence used to support the claim, and the reasoning that links the evidence to the claim and ties the data back to the science context. These three steps are separated in each Data Nugget to emphasize the inclusion of each component in a student's explanation. As a writing extension, you can have students combine each the three parts to write a one paragraph summary or explanation that answers the scientific question.

Data Nugget Graphing Checklist:

Chose correct type of graph, AND includes all of the appropriate components below:

- ❑ One or both axes labeled correctly
- ❑ Units included with axes labels, *when applicable*
- ❑ Scale on both axes is correct
- ❑ Means or data points graphed correctly
- ❑ Includes key/legend to differentiate variables or categories on graph, *when applicable*
- ❑ Trend line or error bars are included and drawn correctly, *when applicable and required by teacher*
- ❑ Title included, *if required by teacher*



Data Nugget Graphing Checklist:

Chose correct type of graph, AND includes all of the appropriate components below:

- ❑ One or both axes labeled correctly
- ❑ Units included with axes labels, *when applicable*
- ❑ Scale on both axes is correct
- ❑ Means or data points graphed correctly
- ❑ Includes key/legend to differentiate variables or categories on graph, *when applicable*
- ❑ Trend line or error bars are included and drawn correctly, *when applicable and required by teacher*
- ❑ Title included, *if required by teacher*

These check boxes can be used as a grading rubric for graphs, or as a handout for students to help them remember all the graph components.