Reading Data Visualizations – Tutorial Script

An important part of critical and skeptical thinking is knowing how to read and interpret data shown in visualizations such as graphs and maps. What follows is a tutorial for how to read these data sets and evaluate them critically.

1. Read the title and caption of the visualization – what do they claim we are seeing?

For example, this graph is showing general variations of the concentrations of carbon dioxide and oxygen gas with depth in the world's oceans. This map is showing global chlorophyll averaged over 1 year's time on the planet. Plenty of questions still, but that's how we start.

2. Look for numbers, scales, and axes. Read the labels. Read the units. Determine the ranges shown.

For example, this graph has a y-axis label showing depth in meters, with 0 depth (sea surface) at the top, and increasing depth as we move down – reaching 1200 meters depth at the bottom of the graph. The x-axis shows two scales – one on top, which refers to the oxygen graph (solid line) and one on the bottom, which refers to the carbon dioxide graph (dashed line). Both gas amounts are given in the same unit: mL/L (a ratio of how much of the volume of a given amount of seawater comes from each gas), with values increasing to the right. For oxygen gas, values frange from 0 on the left to 6 mL/L on the right. For carbon dioxide, values are much much higher, ranging from 46 mL/L at the lowest to more than 52 mL/L at the highest. This map has the latitude and longitude degree marks on its edges and is showing us the entire planet. Colors are used to show us varying amounts of chlorophyll (a pigment found in photosynthesizing autotrophs). Chlorophyll values vary from highs of 35 mg/m^3 to none at all (0). What is a mg/m^3? A density – so the density of chlorophyll found in surface ocean water.

3. What patterns do we see in the data? Do the data vary in a consistent or descriptive way? Are all the values bunched up in one location? Or distributed in an regular pattern?

For example, in the global chorophyll map, we see large dark areas in the centers of the northern and southern Atlantic, northern and southern Pacific, and Indian Oceans where there appears to be no chlorophyll. We see high chlorophyll values in the Arctic Ocean, Antarctic Ocean and along all coastlines. We see moderate values in strips along the equator, thicker in the east than in the west. In this graph, For the oxygen curve, we see the highest oxygen at the surface. We see it drop quickly around 100 m depth, staying consistently low at almost nonexistent values from 200 to 800 meters depth, then increasing at depths below 800 meters, but not as high as at the surface. For the carbon dioxide line, we see the lowest values (45 mL/L – so still pretty high!) at the surface, and then a steep increase from the surface to 300 m and an even steeper increase as we descend below 300 m towards the seafloor.

4. If there are multiple data sets shown in one visualization, do any of the data trends correlate (increase or decrease at the same rates under the same circumstances) or have an inverse correlation (under the same circumstances, one value increases while the other decreases)?

For example, at the surface, down to 200 meters depth, the two are inversely correlated – oxygen decreases as carbon dioxide increases. From 200 to 800 meters there is no correlation at all. Below 800 meters, there is a correlation.

5. Were the data validly gathered?

It's always important to be sure we know where and how these data were gathered. For example, in the graph, we have no source cited, and the graphs are considered generalized variations, not actual data. So someone has reviewed depth profiles of these gases across the world's oceans and then generalized them to these curves. That means that these curves might not represent the situation in all areas of the ocean. And who generalized them? Is there an organization that stands behind them and can show the process through which the generalization happened? Not that we know of. How were the samples taken that were originally part of the larger data set? We don't know (though we can research and apply known techniques for measuring dissolved gases in seawater). But we don't have specifics. So we have to be careful about applying these data to solve specific problems. We should hold onto some skepticism and not expect all locations to look exactly like these lines. For the map, the data come from NOAA. If you follow the data links at the NOAA website, you can download the actual data points that were gathered through satellites based on the natural fluorescence of chlorophyll, which is detectable on the surface of the oceans by satellites. We can read far more about the gathering technique. But we also know NOAA to be an organization committed to long-term data gathering with multiple organizations worldwide using their data and reviewing their techniques. So we can choose to trust these data and the techniques used.

6. What are the limitations of these data? Assumptions?

In the Chlorophyll map, we are limited primarily in scale. By looking at the entire globe, we can't see fine distinctions at a more granular level. In addition, we are looking at averages over a year, so we can't see how each of these areas varies in chlorophyll content with the seasons. In the graph, we are limited by the generalized nature of the curves as already mentioned. We can't apply these trends globally, because we can't be sure all parts of the ocean will have the same processes at work at the same depths.

7. Only after a thorough observation of the data should we start evaluating what it might mean. If two characteristics correlate in their data trends, there might be a relationship between the two characteristics. It might be that one causes the other, or it might mean that the cause of one is also the cause of the other. CORRELATION ≠ CAUSATION. But it can lead us toward the causes.

For example, in the graph, we see inverse correlation at the surface. Below 800 meters, there is a correlation for the same parameters, so we can't apply a simple causation model, because the data contradict that. However, there are processes at work at the surface of the ocean where sunlight is abundant that aren't at work at depth where it's dark. Photosynthesis will exist at the surface, and this process uses up carbon dioxide and produces oxygen. So that's a good hypothesis for explaining the surface variations. At depth, both gases are subjected to high pressures and cold temperatures, both of which lead to increased gas dissolution (like with carbonated beverages). Perhaps this explains why both gases are found in increasing amounts with depth. In the chlorophyll map, we see that high chlorophyll correlates with proximity to coastlines, and coastlines are the primary source of nutrients necessary for photosynthesizing autotrophs to grow. So perhaps the lack of chlorophyll in the central oceans and increase in chlorophyll in coastal areas are related to the availability of nutrients in the surface waters.

Are there more data we'd like to collect? Questions we'd to ask? If so, what?

For example, In the chlorophyll map, we might ask to look at the data by season or look at a small area in more detail. We might, as already discussed, want to see some other data sets alongside, such as nutrient content and sea surface temperatures. These will allow us to confirm our original hypothesis or transform it into something that fits the data better. In the graph, what's going on between 200 and 800 meters depth to oxygen? What process is at work there that seems to have such a big impact on oxygen, but not carbon dioxide? And what happens at depths greater than 1000 meters, where we no longer see any data? And what variations if any exist across latitudes and longitudes in the oceans?

We've only begun to explore the data and visualizations in these images. The more we look, the more questions we can ask. We want to apply skeptical and critical thinking to carefully evaluate what we see, determine if the data are robust and valid enough to allow us to generate some hypotheses, and then begin developing and testing those hypotheses so we can use these data points to learn more about the natural systems around us.

[end credits]

Reading Data Visualizations

Produced by Katryn Wiese
City College of San Francisco

All media produced by Katryn Wiese unless indicated below:
*Intro and exit music: used with permission © Alexis Harte (www.alexisharte.com)
*Quiz music: used with permission © Proceed with Caution
*Chlorophyll map (NOAA)
*Carbon Dioxide and Oxygen Variations with Depth generalized based on data from NOAA
*Hilo tsunami water level (NOAA)