A Guide to Statistics and Figures Used in Scientific Analyses

Understanding Significance and p-Values:

Statistical tests are used to determine if relationships between sites and variables are the result of random occurrences or if they are patterns with an underlying explanation. The key value that comes out of a statistical test is the p-value, which ranges from 0 to 1 and represents the likelihood that a relationship is the result of random variation. A p-value of 0.05 and below is considered statistically significant; there is a 5% chance or lower that the observed relationship is caused by randomness, meaning that there is likely a pattern with an explanation. For example, if a statistical test is run to compare water temperature at two sites and the p-value is 0.02, there is only a 2% chance that the difference between the sites is due to random variation. Rather, some other phenomenon is leading to statistically significant differences in water temperature at the two sites. Conversely, a p-value of 0.75 would indicate that there is a 75% chance that any difference between sites is caused merely by randomness and that there are no significant differences in water temperature between sites.

How to Read Boxplots:

A boxplot is a chart that visualizes the distribution of all measurements of a specific variable and allows for comparisons between sample sites. Boxplots are made up of two parts, a box and whiskers (Figure 1a). The box represents the spread of the middle half of data points and shows the median. Whiskers extend to maximum and minimum values. Data points outside the whiskers can be considered outlier values. The size of the box represents the spread of the data: a small box indicates that most values are similar, while a large box signifies high variability.

Letters above boxplots denote statistical relationships between sites; sites that share a letter have statistically similar means (or averages) and sites that do not share any letters have significantly different means at the $p < 0.05$ level (Figure 1b). If there are no letters, there are no statistically significant differences between any sample sites. For example, in Figure 1b, which shows dissolved oxygen, all sites are labeled with either an ‘a’, ‘b’, or ‘ab.’ DCWM1 and Crum Creek Main Stem Upstream (CC2) are labeled with an ‘a,’ meaning they are statistically similar to each other and to all sites that have an ‘ab.’ The only site that DCWM1 and CC2 are significantly different from is Ridley Creek at Okehocking Preserve (RCOK1), which is only labeled with a ‘b.’ However, RCOK1 is also statistically similar to all the sites that are labeled with an ‘ab,’ since these sites all share a ‘b.’

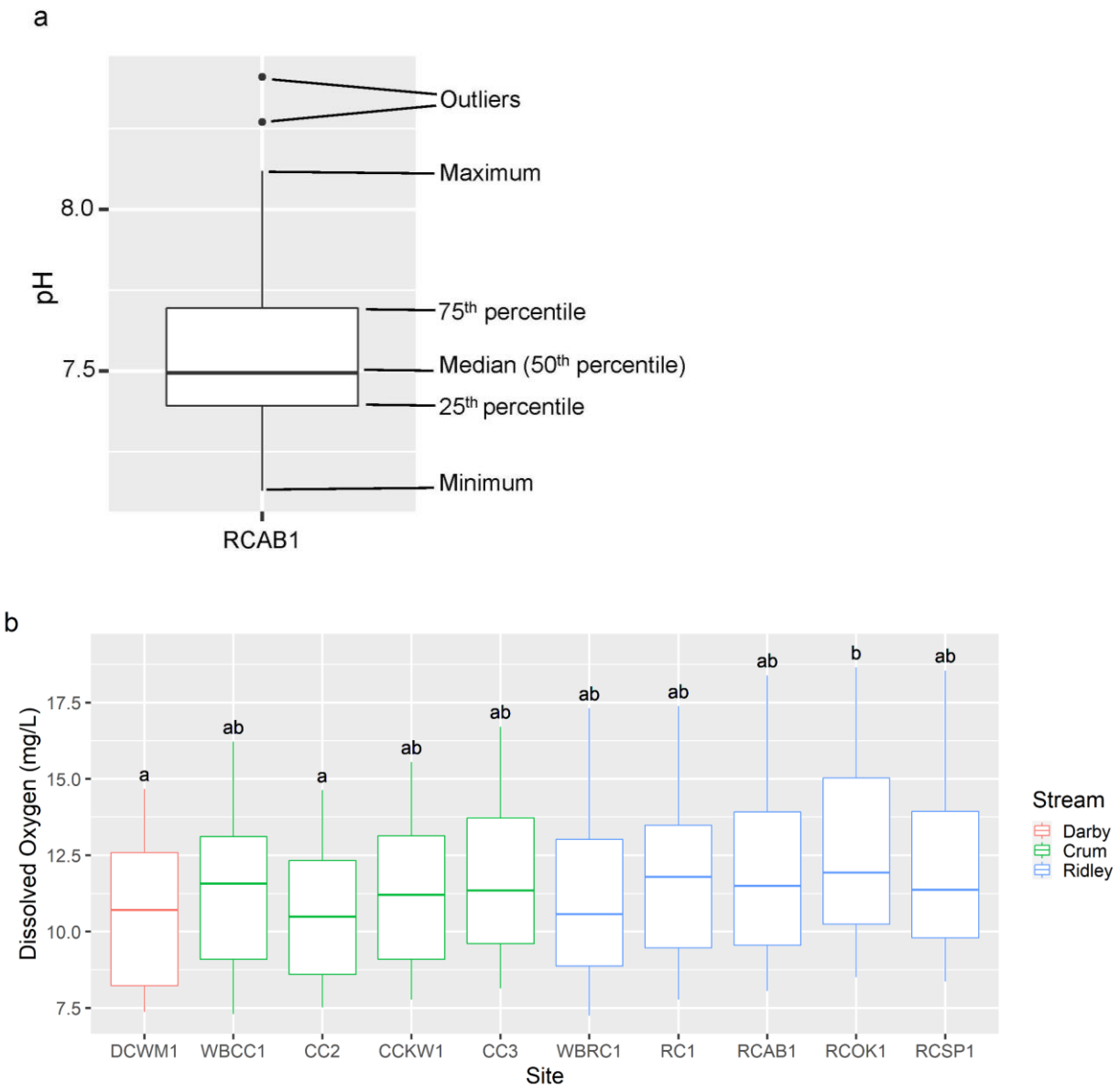

Figure 1. (a) Sample boxplot of the distribution of pH measurements at a sample site. The box represents the interquartile, or IQ, range (the spread of the data from the 25th percentile to the 75th percentile) and the line within the box is the median (the halfway point of the data). Whiskers extend to the maximum and minimum values and any dots outside the whiskers can be considered outliers. (b) Dissolved oxygen across sample sites with significance letters.

How to Read Scatterplots:

Scatterplots are graphs that show individual data points as dots (Figure 2). Each dot represents one data point — in this case, a measurement at a sample site — that is plotted against two variables, one on the horizontal axis (x-axis) and one on the vertical axis (y-axis). Scatterplots are useful for visualizing changes in a specific variable (such as water temperature) over time or for comparing two different variables. Many scatterplots in this document that show the change in a variable over time have a large gap with no data. This gap reflects a pause in sampling due to the COVID-19 pandemic.

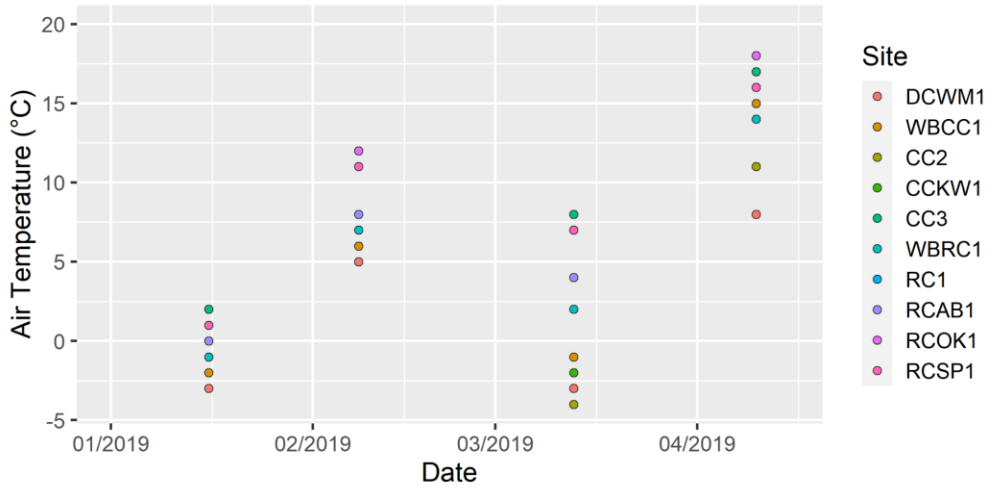

Figure 2. A sample scatterplot of air temperature from January 2019 through April 2019. Each stack of data points represents a sampling day and each sample site is represented by a dot. Dots are colored by site.

Funding

This report was made possible through a grant from the **William Penn Foundation**. The William Penn Foundation, founded in 1945 by Otto and Phoebe Haas, is dedicated to improving the quality of life in the Greater Philadelphia region through efforts that increase educational opportunities for children from low-income families, ensure a sustainable environment, foster creativity that enhances civic life, and advance philanthropy in the Philadelphia region. In 2021, the Foundation will grant more than $117 million to support vital efforts in the region.
The opinions expressed in this report are those of the author(s) and do not necessarily reflect the views of the William Penn Foundation.