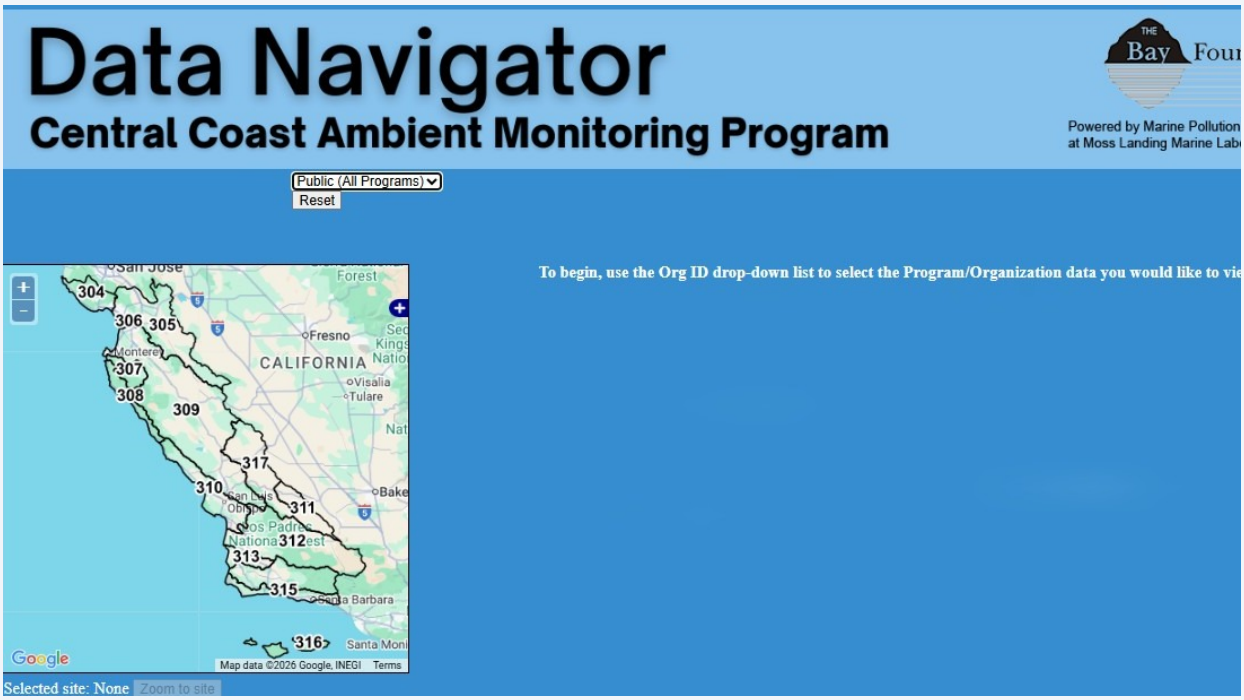
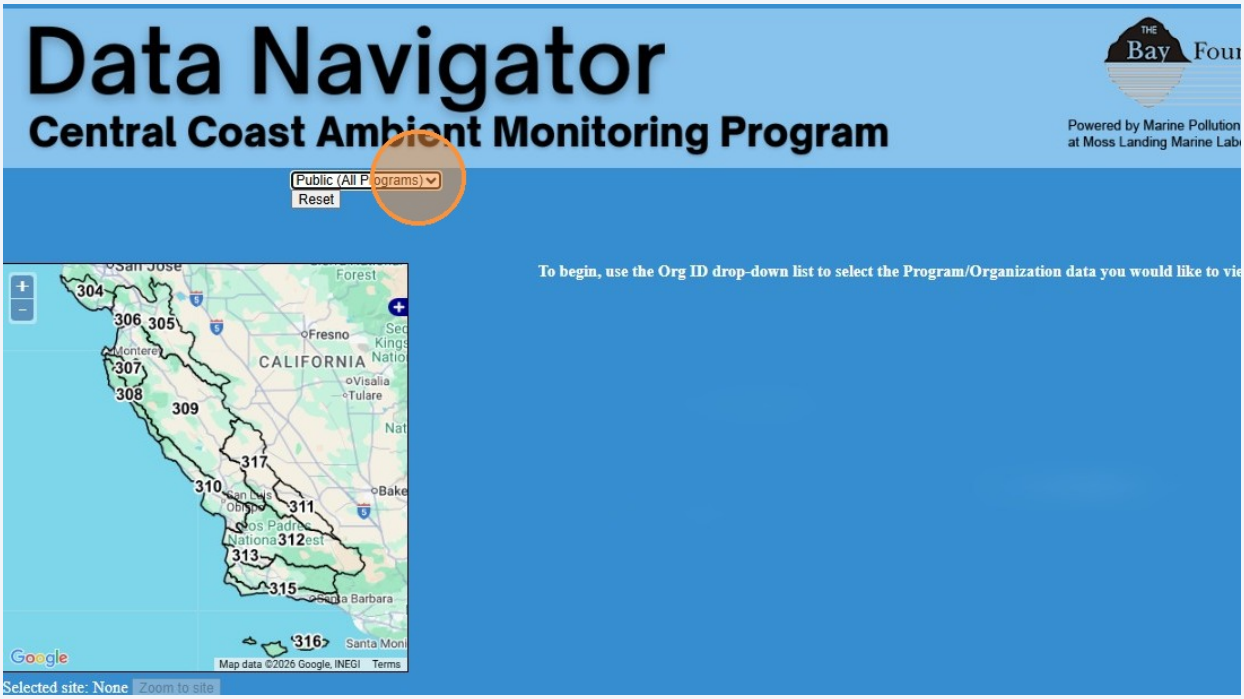
Data Navigator Guide for CCAMP Statistical Analysis

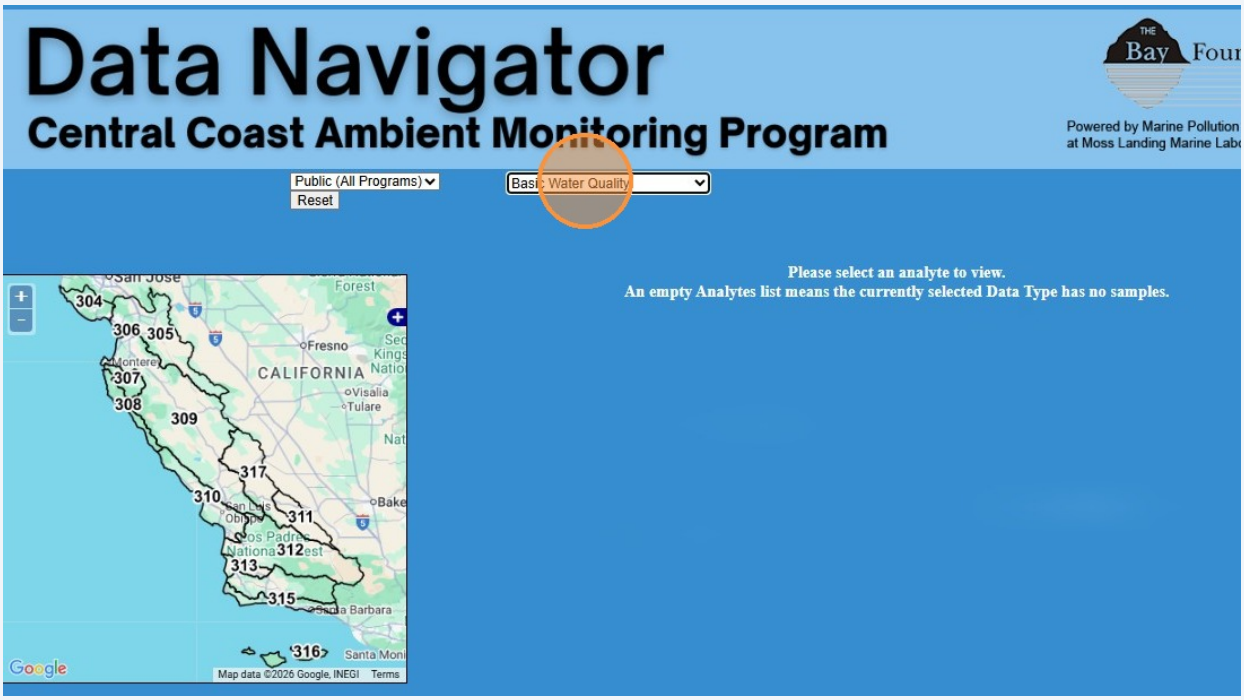
1 Navigate to www.ccamp.org/datanavigator



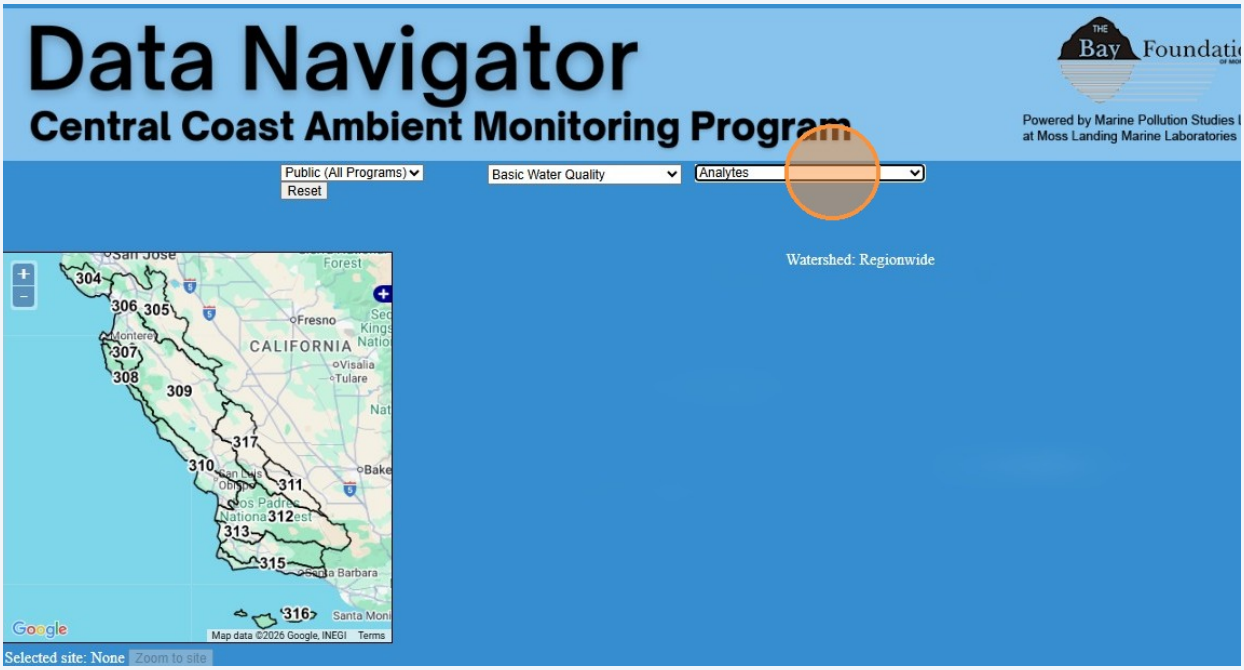
2 Click the "Org ID" drop down menu and select the "Public (All Programs)" option.



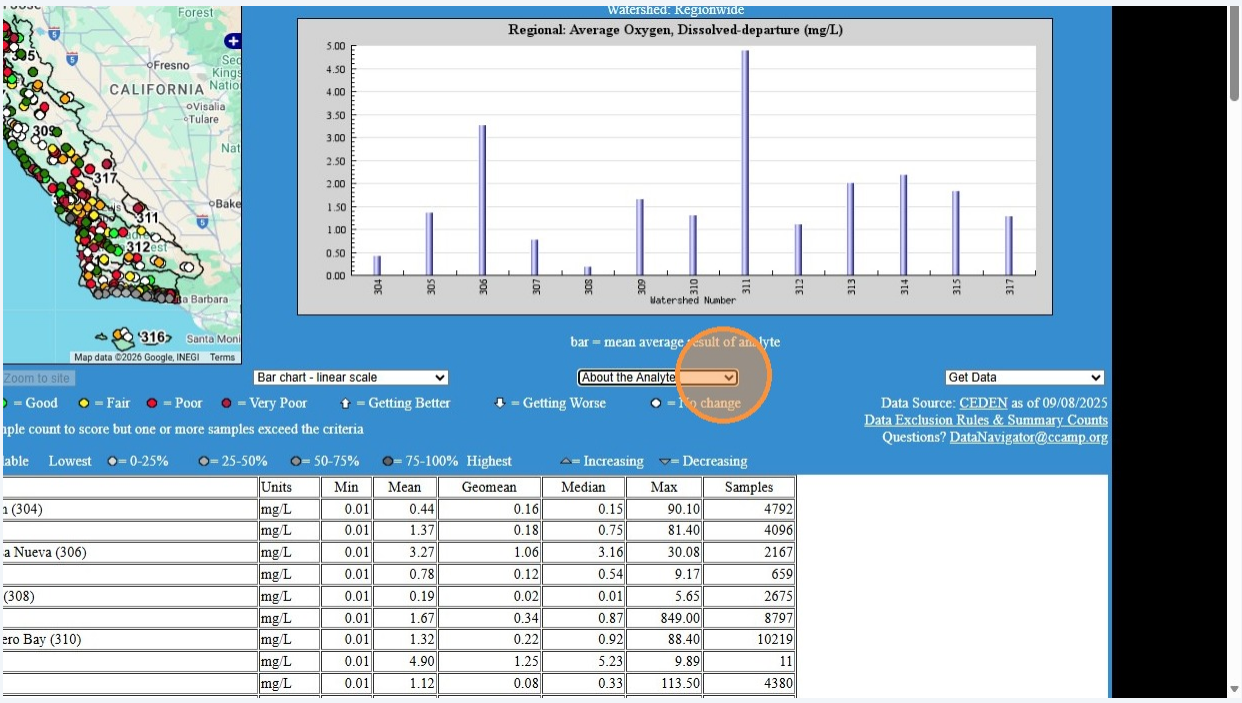
3 Click the Data Type drop down menu and select the "Basic Water Quality" option.



4 Click the "Analytes" drop down menu and choose an analyte (ex. Oxygen, Dissolved-departure).

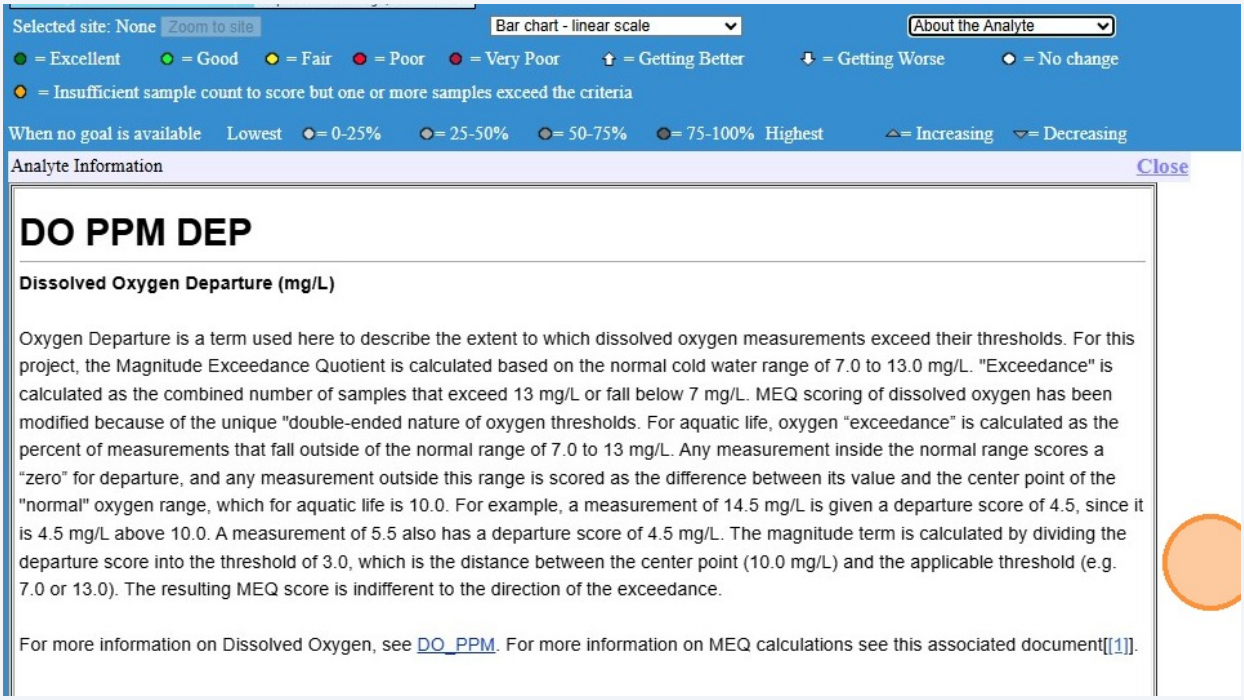


5 Click on the "More Information" drop down menu beneath the graph, then select the "About the Analyte" option.

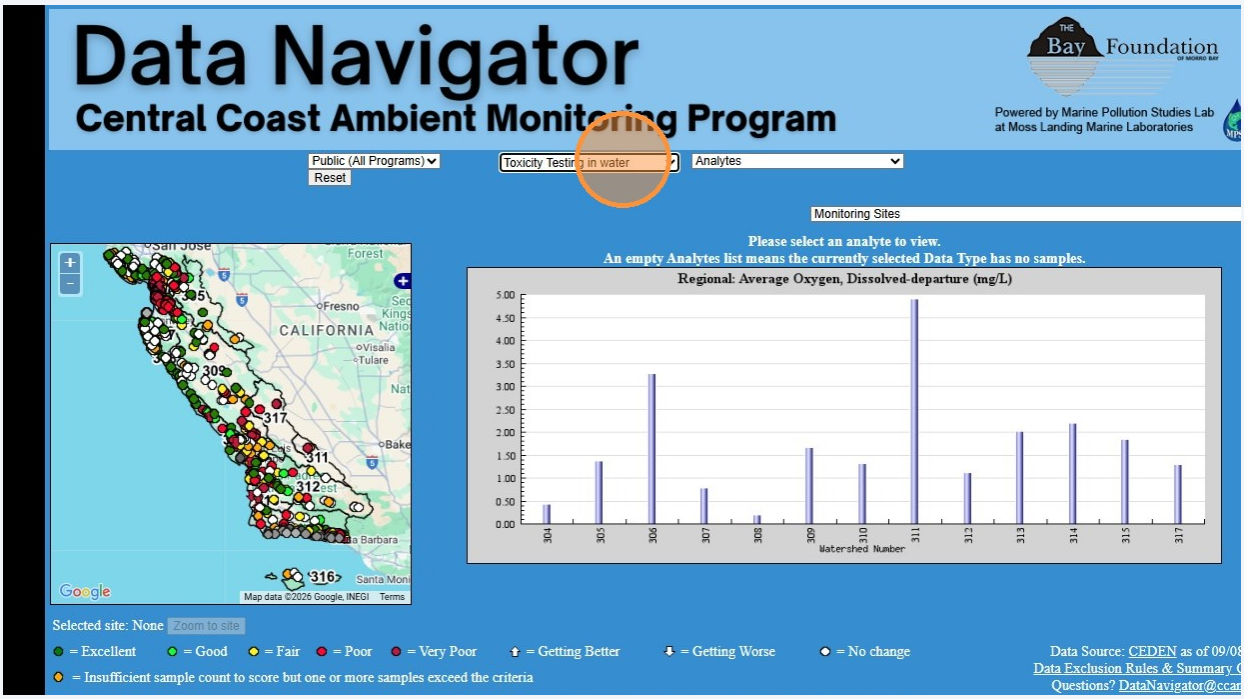


	Units	Min	Mean	Geomean	Median	Max	Samples
n (304)	mg/L	0.01	0.44	0.16	0.15	90.10	4792
	mg/L	0.01	1.37	0.18	0.75	81.40	4096
a Nueva (306)	mg/L	0.01	3.27	1.06	3.16	30.08	2167
	mg/L	0.01	0.78	0.12	0.54	9.17	659
(308)	mg/L	0.01	0.19	0.02	0.01	5.65	2675
	mg/L	0.01	1.67	0.34	0.87	849.00	8797
ero Bay (310)	mg/L	0.01	1.32	0.22	0.92	88.40	10219
	mg/L	0.01	4.90	1.25	5.23	9.89	11
	mg/L	0.01	1.12	0.08	0.33	113.50	4380

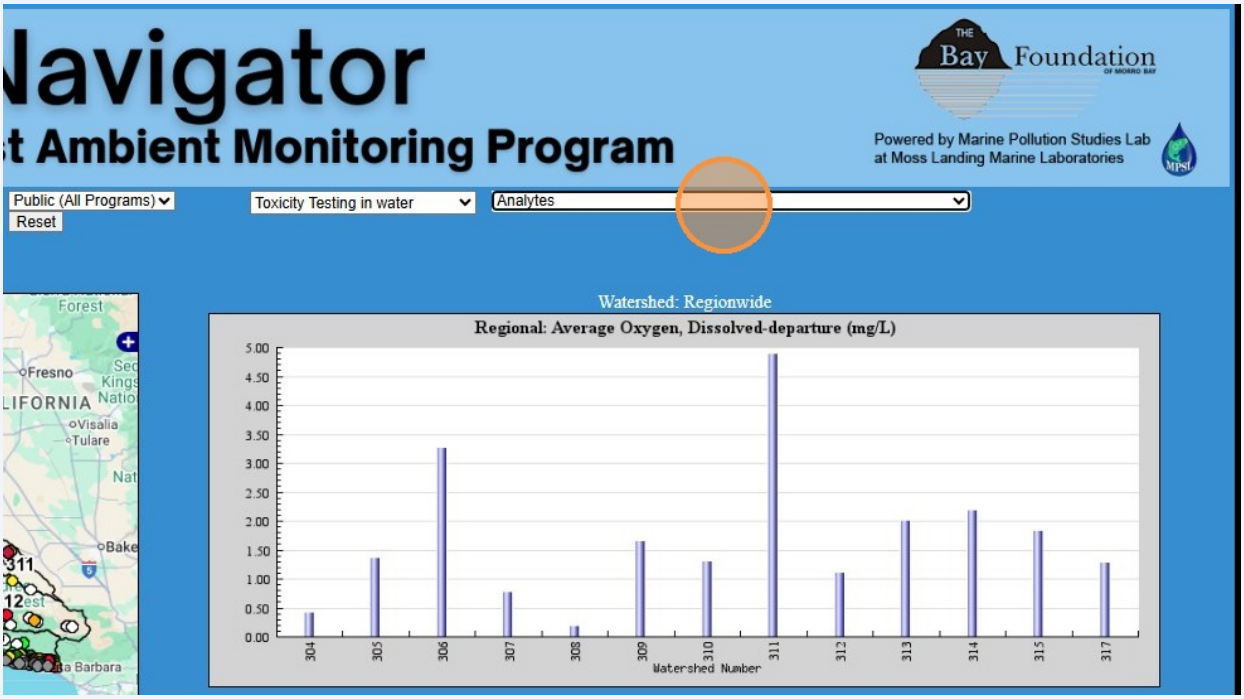
6 Below the key, an explanation of the analyte will be provided [Task 1.4].



7 To select a different analyte, click the 'Data Type' drop down menu, and select the "Toxicity Testing in water" option.



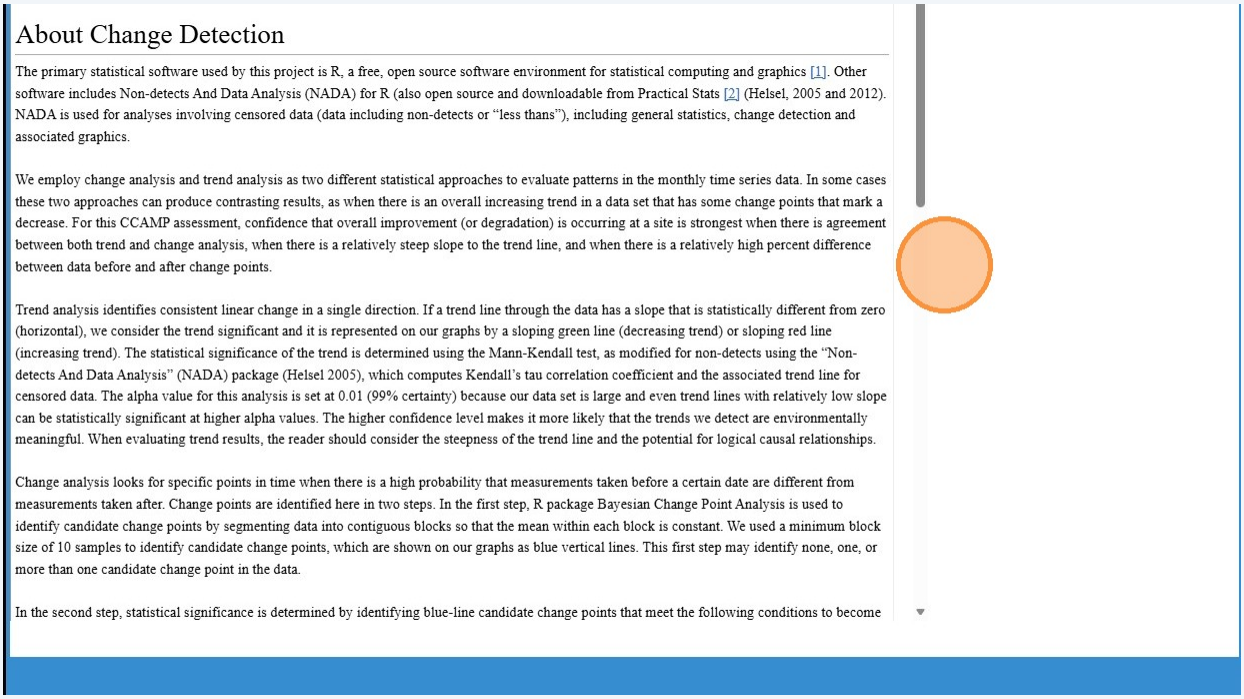
8 Click the "Analytes" option, and select an analyte (ex. Fish Survival).



9 Beneath the graph, click the "More Information" drop down menu, and select 'About Change Detection' [Task 2.2].

	Units	Min	Mean	Geomean	Median	Max	Samples
a (304)	% control	66.3	93.9	93.6	95.2	100.0	28
	% control	42.1	96.7	96.4	99.6	100.0	351
a Nueva (306)	% control	85.0	93.8	93.6	95.0	100.0	6
	% control	100.0	100.0	100.0	100.0	100.0	4
(308)	% control	98.0	-	-	-	100.0	2
	% control	0.0	93.7	94.2	98.9	100.0	563
ero Bay (310)	% control	72.2	96.8	96.6	99.0	100.0	167

10 Scroll down to learn more about Change Point Analysis.



About Change Detection

The primary statistical software used by this project is R, a free, open source software environment for statistical computing and graphics [1]. Other software includes Non-detects And Data Analysis (NADA) for R (also open source and downloadable from Practical Stats [2] (Helsel, 2005 and 2012). NADA is used for analyses involving censored data (data including non-detects or "less thans"), including general statistics, change detection and associated graphics.

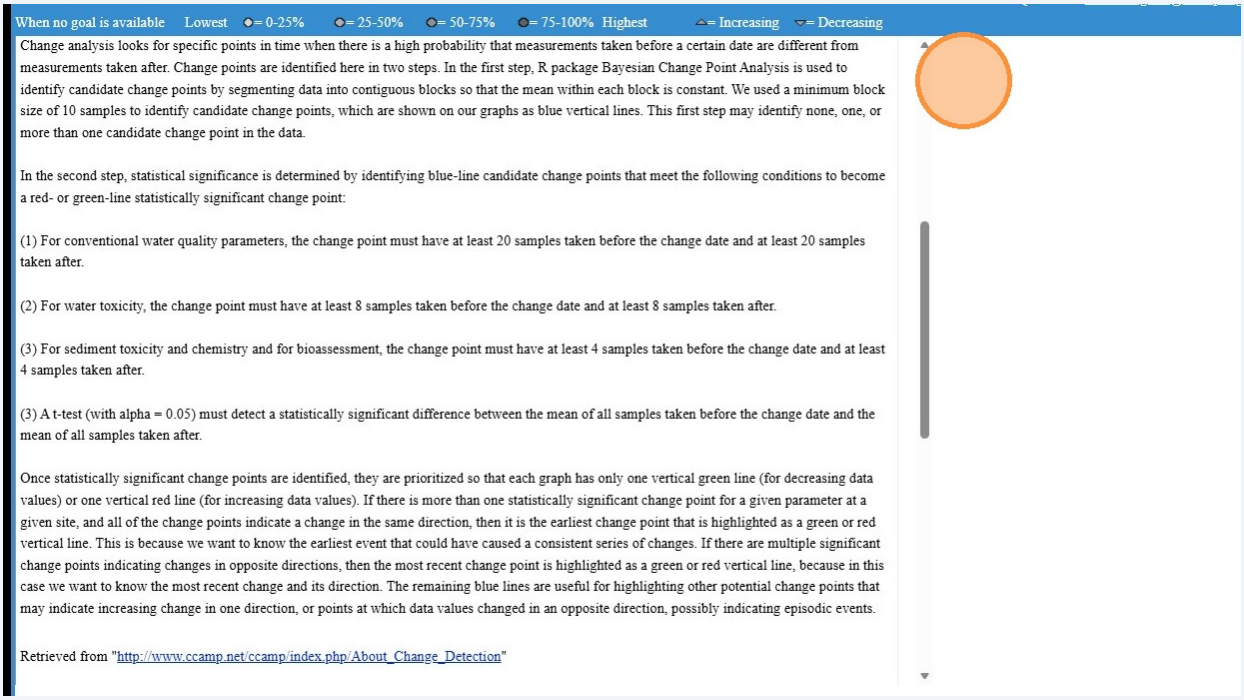
We employ change analysis and trend analysis as two different statistical approaches to evaluate patterns in the monthly time series data. In some cases these two approaches can produce contrasting results, as when there is an overall increasing trend in a data set that has some change points that mark a decrease. For this CCAMP assessment, confidence that overall improvement (or degradation) is occurring at a site is strongest when there is agreement between both trend and change analysis, when there is a relatively steep slope to the trend line, and when there is a relatively high percent difference between data before and after change points.

Trend analysis identifies consistent linear change in a single direction. If a trend line through the data has a slope that is statistically different from zero (horizontal), we consider the trend significant and it is represented on our graphs by a sloping green line (decreasing trend) or sloping red line (increasing trend). The statistical significance of the trend is determined using the Mann-Kendall test, as modified for non-detects using the "Non-detects And Data Analysis" (NADA) package (Helsel 2005), which computes Kendall's tau correlation coefficient and the associated trend line for censored data. The alpha value for this analysis is set at 0.01 (99% certainty) because our data set is large and even trend lines with relatively low slope can be statistically significant at higher alpha values. The higher confidence level makes it more likely that the trends we detect are environmentally meaningful. When evaluating trend results, the reader should consider the steepness of the trend line and the potential for logical causal relationships.

Change analysis looks for specific points in time when there is a high probability that measurements taken before a certain date are different from measurements taken after. Change points are identified here in two steps. In the first step, R package Bayesian Change Point Analysis is used to identify candidate change points by segmenting data into contiguous blocks so that the mean within each block is constant. We used a minimum block size of 10 samples to identify candidate change points, which are shown on our graphs as blue vertical lines. This first step may identify none, one, or more than one candidate change point in the data.

In the second step, statistical significance is determined by identifying blue-line candidate change points that meet the following conditions to become

11



When no goal is available Lowest = 0-25% = 25-50% = 50-75% = 75-100% Highest = Increasing = Decreasing

Change analysis looks for specific points in time when there is a high probability that measurements taken before a certain date are different from measurements taken after. Change points are identified here in two steps. In the first step, R package Bayesian Change Point Analysis is used to identify candidate change points by segmenting data into contiguous blocks so that the mean within each block is constant. We used a minimum block size of 10 samples to identify candidate change points, which are shown on our graphs as blue vertical lines. This first step may identify none, one, or more than one candidate change point in the data.

In the second step, statistical significance is determined by identifying blue-line candidate change points that meet the following conditions to become a red- or green-line statistically significant change point:

(1) For conventional water quality parameters, the change point must have at least 20 samples taken before the change date and at least 20 samples taken after.

(2) For water toxicity, the change point must have at least 8 samples taken before the change date and at least 8 samples taken after.

(3) For sediment toxicity and chemistry and for bioassessment, the change point must have at least 4 samples taken before the change date and at least 4 samples taken after.

(3) A t-test (with alpha = 0.05) must detect a statistically significant difference between the mean of all samples taken before the change date and the mean of all samples taken after.

Once statistically significant change points are identified, they are prioritized so that each graph has only one vertical green line (for decreasing data values) or one vertical red line (for increasing data values). If there is more than one statistically significant change point for a given parameter at a given site, and all of the change points indicate a change in the same direction, then it is the earliest change point that is highlighted as a green or red vertical line. This is because we want to know the earliest event that could have caused a consistent series of changes. If there are multiple significant change points indicating changes in opposite directions, then the most recent change point is highlighted as a green or red vertical line, because in this case we want to know the most recent change and its direction. The remaining blue lines are useful for highlighting other potential change points that may indicate increasing change in one direction, or points at which data values changed in an opposite direction, possibly indicating episodic events.

Retrieved from "http://www.ccamp.net/ccamp/index.php/About_Change_Detection"