Instructions

This tool is an interactive web application. It was created using R Shiny (https://shiny.rstudio.com/). You will be given a set of task-based scenarios to place the use of the tool in a context for problem solving and data analysis. Each task starts with an *italicized introduction* for context and includes step-by-step instructions for navigating the tool. After each task, you will be prompted to return to the survey page and to answer a series of questions. As you follow the instructions, please reflect on the water quality and resource management questions you may have for your own data, and consider the potential of a tool that could help to address your concerns.

Step 1: Download the files Ref_Sample.csv, CBM_Sample.csv, and projection.txt to an easily accessible location (ie Downloads, Desktop, etc)

Step 2: Open the survey page found at (insert final survey url)

**Step 3:** Click on the link to launch the tool: (insert final app url)

Step 4: Read the following prompt (in italics) and follow the instructions to upload the provided data files (Ref_Sample.csv and CBM_Sample.csv)

This online tool requires each user to provide the data they wish to work with. Two files have been provided to you: Ref_Sample and CBM_Sample. They are both comma-separated value files created in Microsoft Excel. In this task, you will upload both files to the appropriate locations in the app. You will also tell the app which columns in the files contain important information: the site names, the date, the latitude value, and the longitude value. You will also tell the app how to read the date column so that it may work properly with the data. These steps are necessary because they form the basis of the app's interaction with the data, and help to make the app robust against different column names, different date types, and different column orders.

Click on the Reference Dataset Tab and click on Data Upload and Properties.
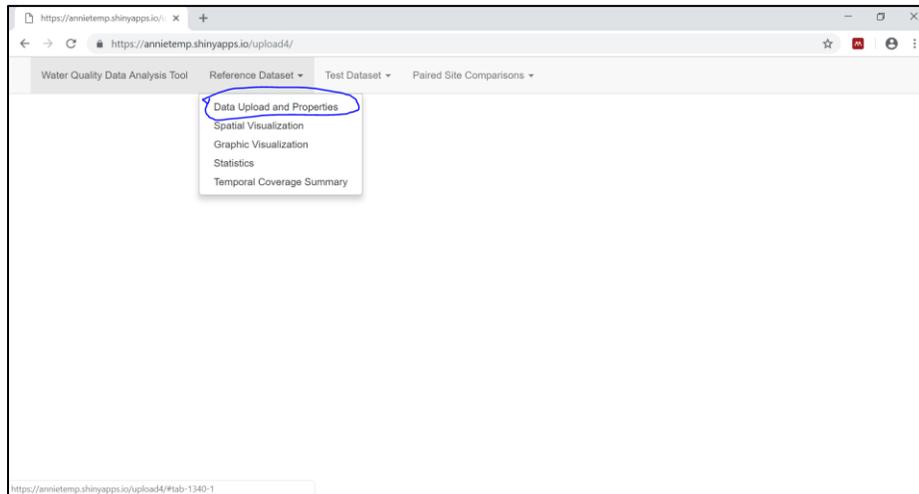
At the top of the leftmost column, under "Please Upload Reference Data CSV File", click on Browse.
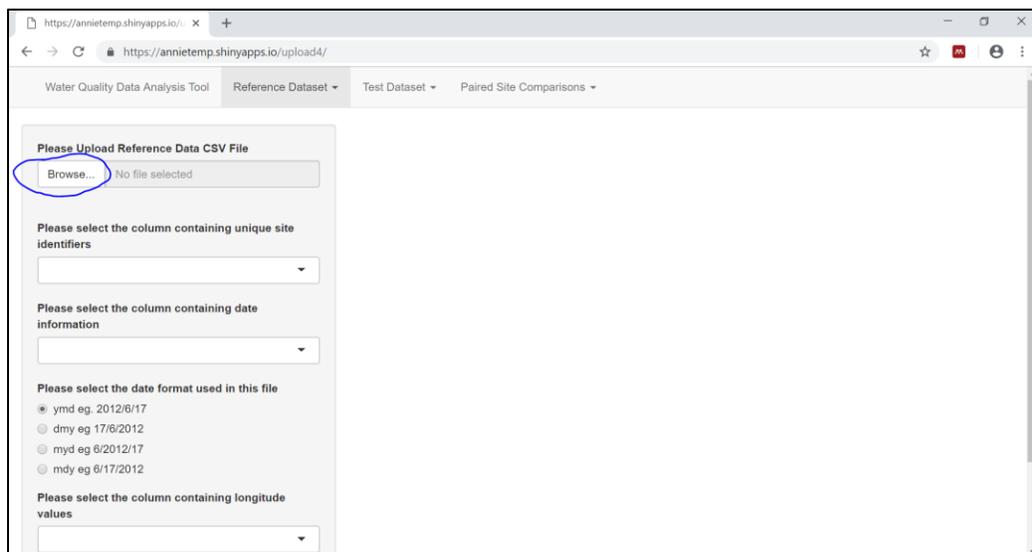
Navigate to the location where you saved the file Ref_Sample.csv

Select the file Ref_Sample.csv. Click Open

Once uploaded, your file will appear to the right of the grey column. The column names (headings) will appear in bold. Note that you can navigate through the file using the controls above and below the data.

In this file, each line has a site code (Site_UID), latitude and longitude coordinates (columns Y and X respectively), the date of observation (Date), and values for two water quality parameters, pH and DO (Dissolved Oxygen).

Use the first drop-down box to select Site_UID.

Use the second drop-down box (Please select the column containing date information) to select Date.

As the dates are formatted as day/month/year (check the Date column to verify), click on the dmy button.

Use the third drop-down box (Please select the column containing longitude values) to select the X column

Use the lowest drop-down box (Please select the column containing latitude values) to select the Y column.

Note that the options in each drop-down box are the names of the columns in the uploaded file.
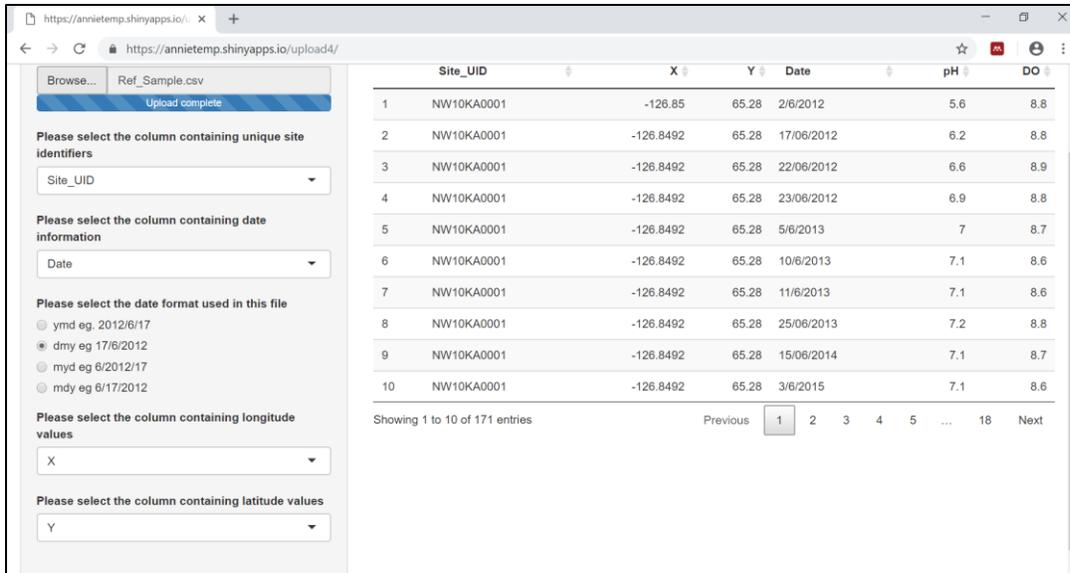
You have now successfully uploaded the first file. This data will serve as your reference/baseline, using sites monitored by Environment and Climate Change Canada.

Under the Test Dataset tab, click on Data Upload and Properties.
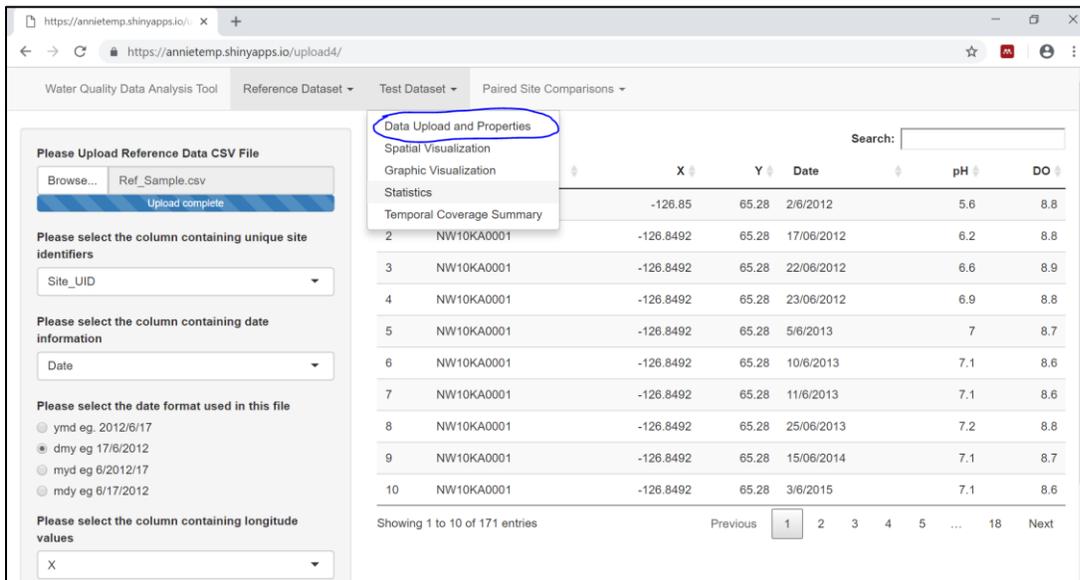
At the top of the leftmost column, under "Please Upload Reference Data CSV File", click on Browse.

Navigate to the location where you saved the file Ref_Sample.csv

Select the file CBM_Sample.csv. Click Open

Once uploaded, your file will appear to the right of the grey column. The column names (headings) will appear in bold. Note that you can navigate through the file using the controls above and below the data. In this file, each line has a site code (Site_UID), latitude and longitude coordinates (columns Y and X respectively), the date of observation (Date), and values for two water quality parameters, pH and DO (Dissolved Oxygen).

Use the first drop-down box (Please select the column containing unique site identifiers) to select Site_UID.

Use the second drop-down box (Please select the column containing date information) to select Date.

As the dates are formatted as day/month/year (check the Date column to verify), click on the dmy button.

Use the third drop-down box (Please select the column containing longitude values) to select the X column

Use the lowest drop-down box (Please select the column containing latitude values) to select the Y column

Note that this process is identical to that of the previous file.

You have now successfully uploaded the second file, CBM_Sample. This file will serve as the set of community-based water quality data that is being investigated and compared to the reference baseline.

Please return to the survey website and complete the questions in Section 1. Please return to this document when you have finished and proceed to Step 5 (do not exit the survey).

Step 5: Read the following prompt (in italics) and follow the instructions to see how different functions of the tool can be used together to identify data errors and to understand the data's structure

You have just compiled a new set of CBM data from the rivers around your community and are ready to begin analyzing your data. However, since the data was originally hand written and had to be entered into a spreadsheet, it is important to check for data entry errors for the parameters you are interested in analyzing, in this case Dissolved Oxygen (DO). Please ensure that you have already completed Task 1.

Under the Test Dataset tab, click on Spatial Visualization.

Use the drop-down menu to select DO (Dissolved Oxygen).
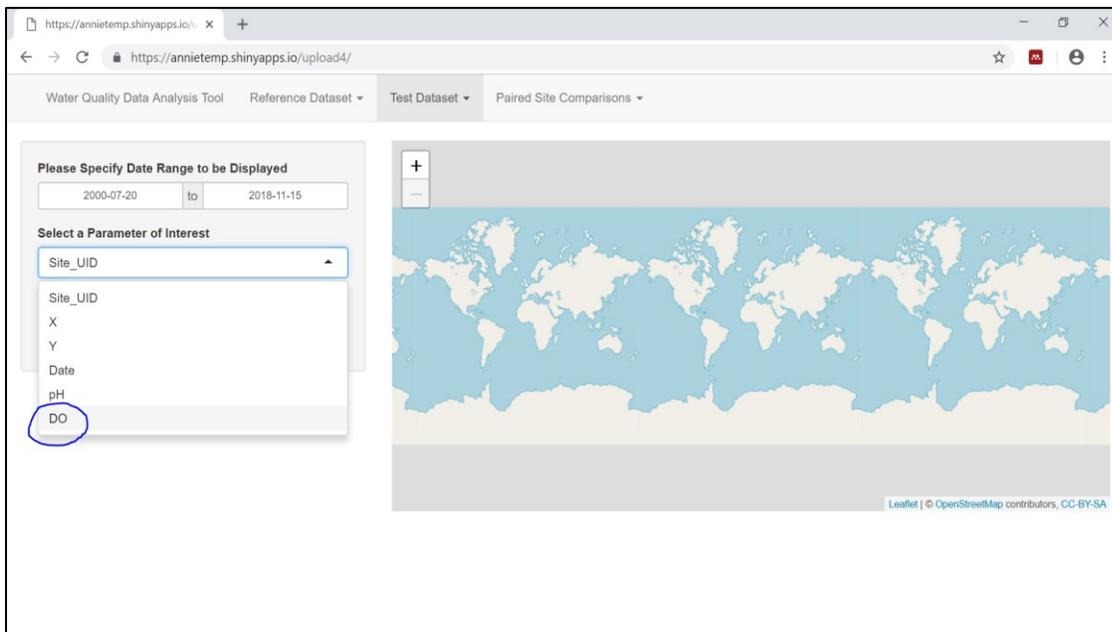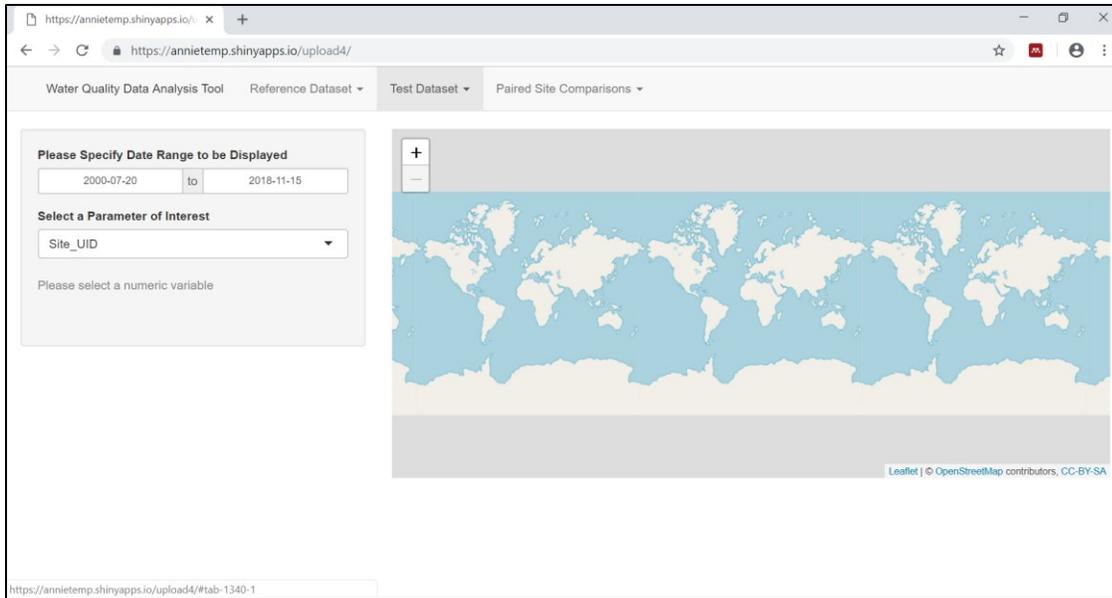
Note that, since DO is a numeric parameter, the app accepts this selection and shows two markers on the map. The slider shows a very high value on the right (96) – this could indicate a mistake in the data.
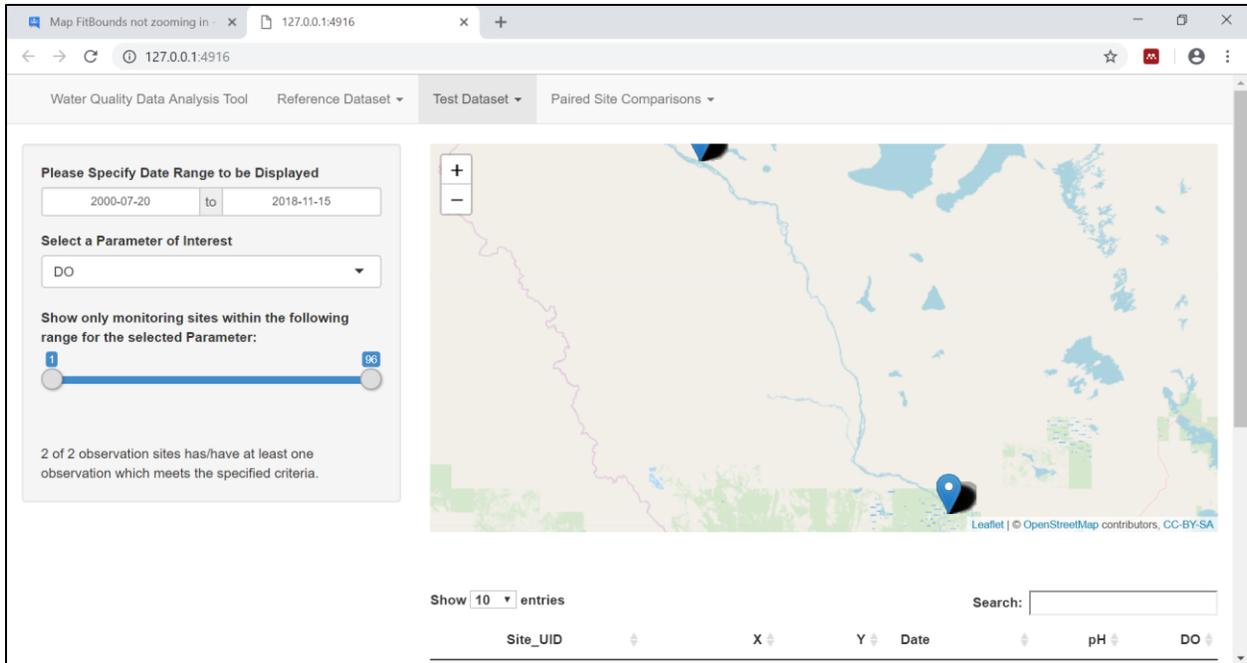
Use the slider to isolate this value by clicking on the left-most circle and dragging it toward the right end of the slider.
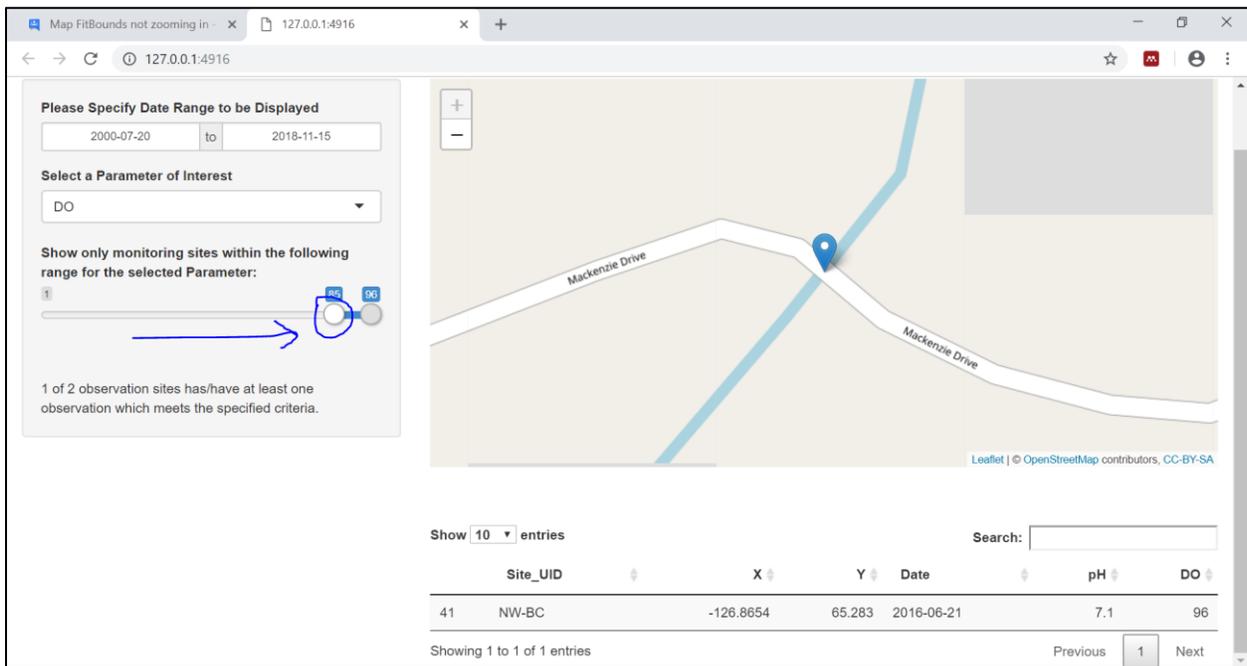
Determine which lines of code are relevant. Make note of the line number, the site code, and the date of the offending line (line 41, site NW-BC, 2016-06-21).

Even though this seems to be a large value, it is possible that it is still within a reasonable range, relative to the rest of the data.

Under the Test Dataset tab, click on Statistics.
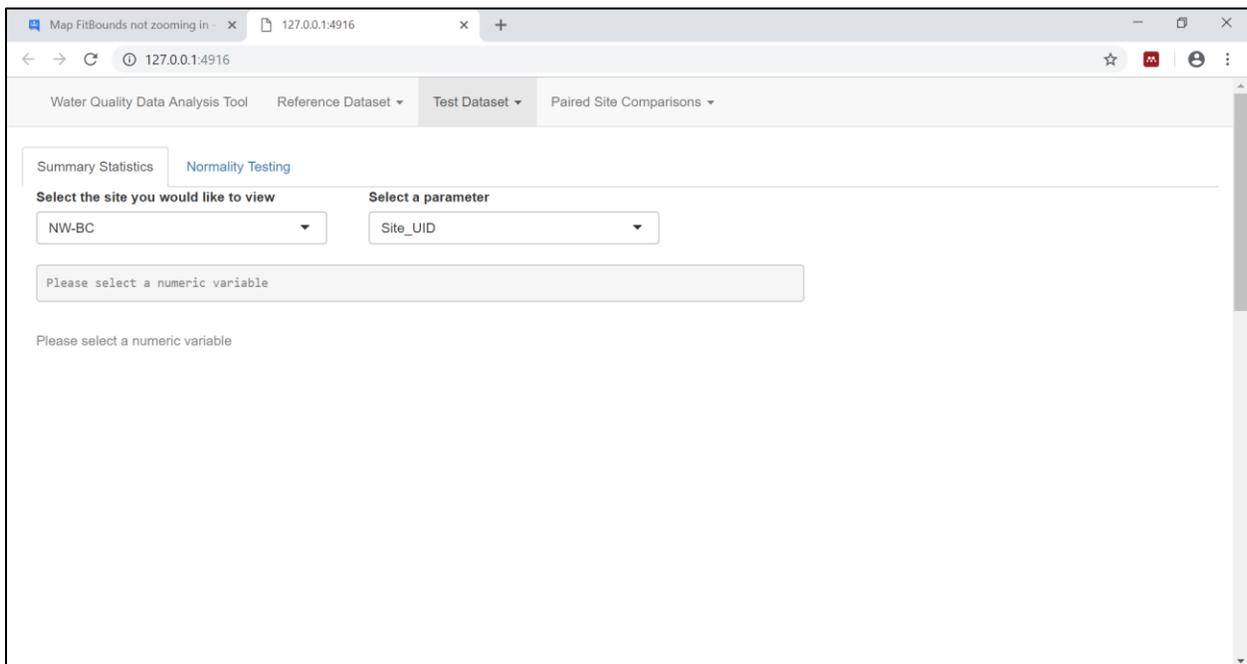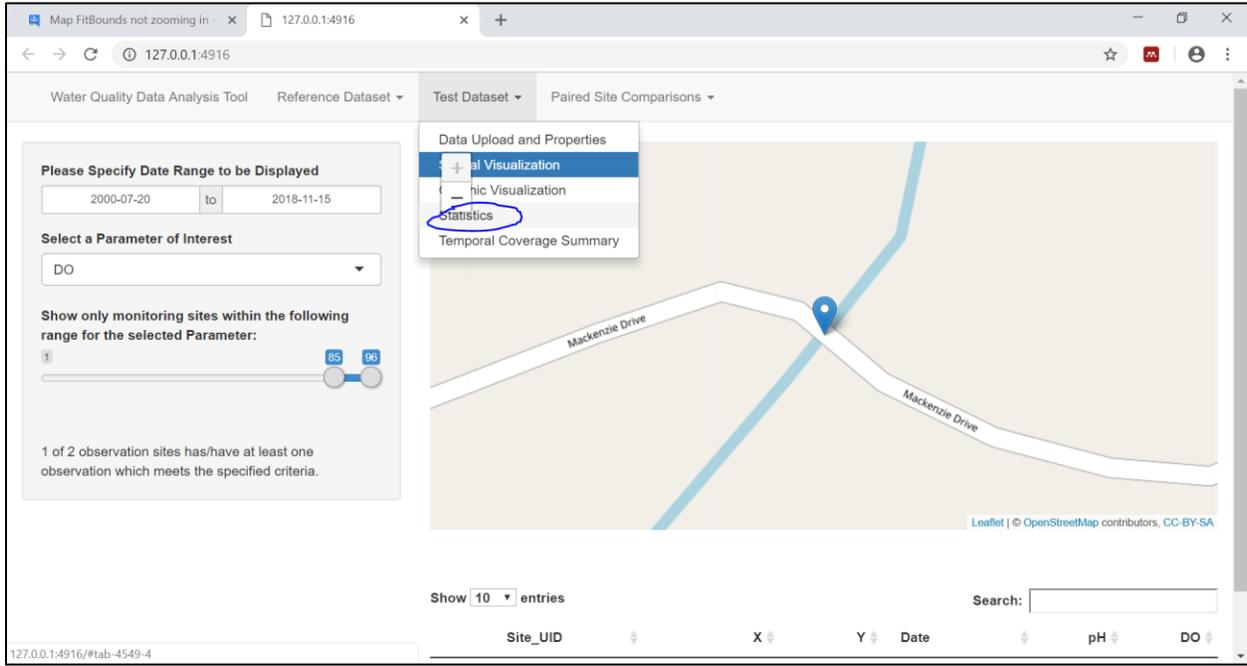
Ensure that site NW-BC is selected. Use the drop-down menu under "Select a parameter" to select DO.
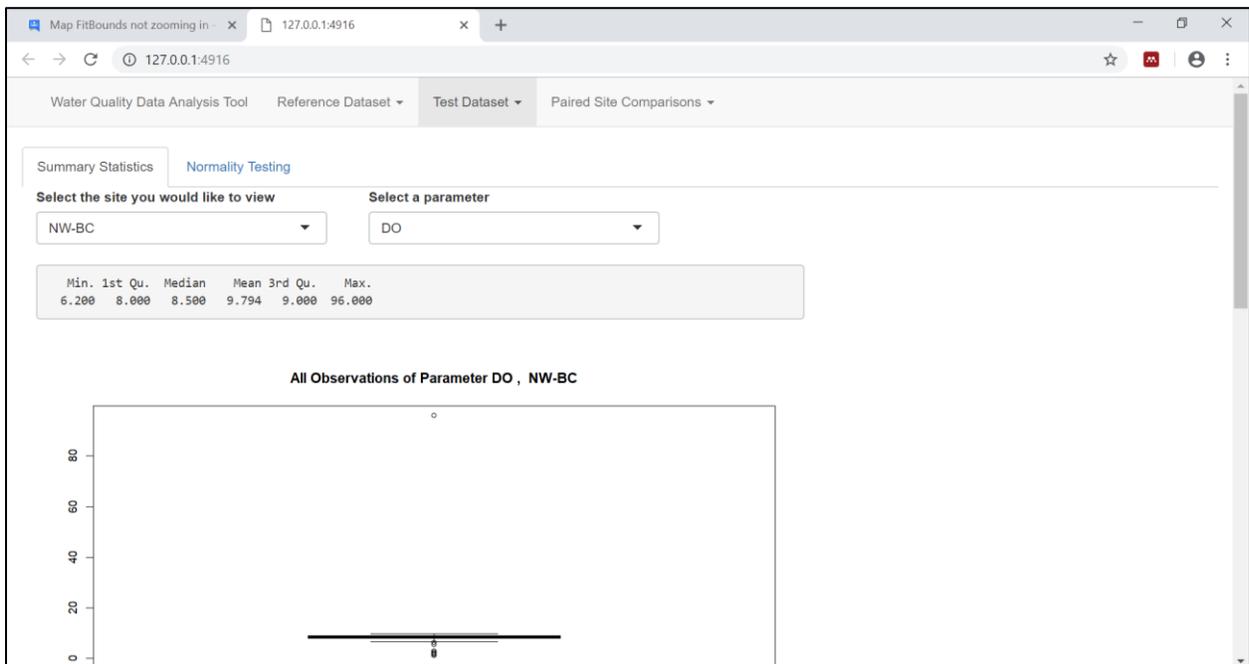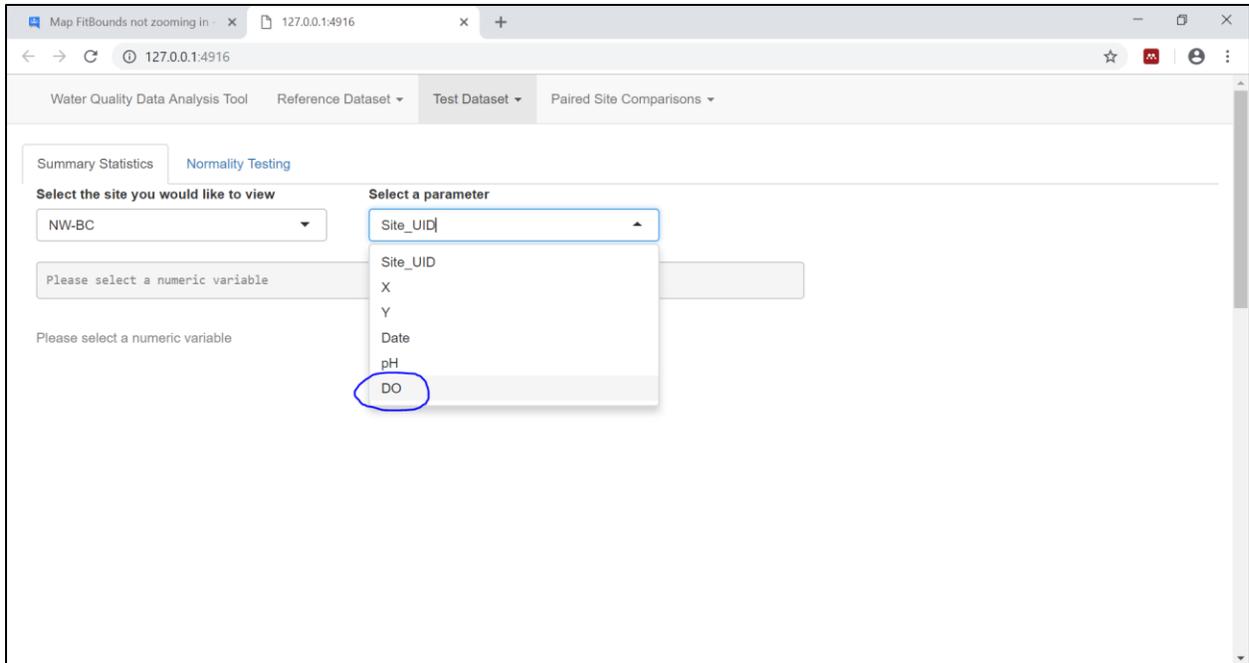
Three graphs will appear on the screen (scroll down if needed). The graphs show boxplots. In this case, it is not necessary to know everything about how to interpret boxplots. However, it is clear that there is one single point that does not fall anywhere near the other lines and points on all three of the graphs. This shows that the value of 96 is not similar to the rest of the data, and is likely a data entry mistake.

Return to the Spatial Visualization page under the Test Dataset tab. The page will have remained the same since you last viewed it.
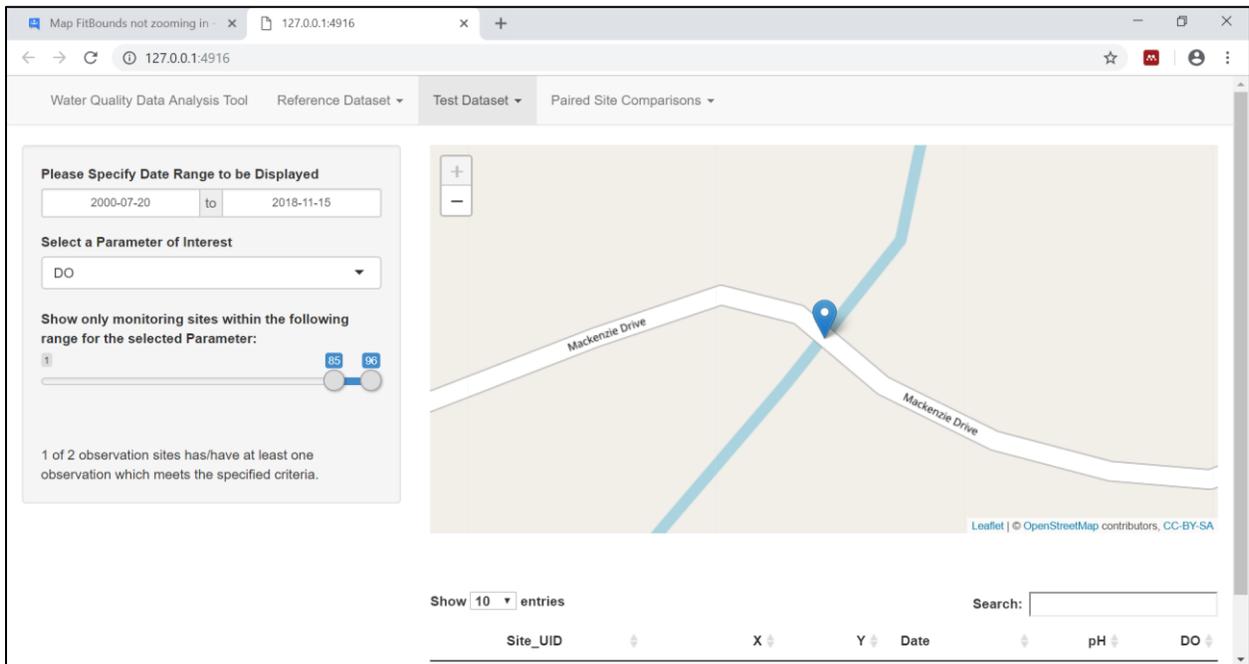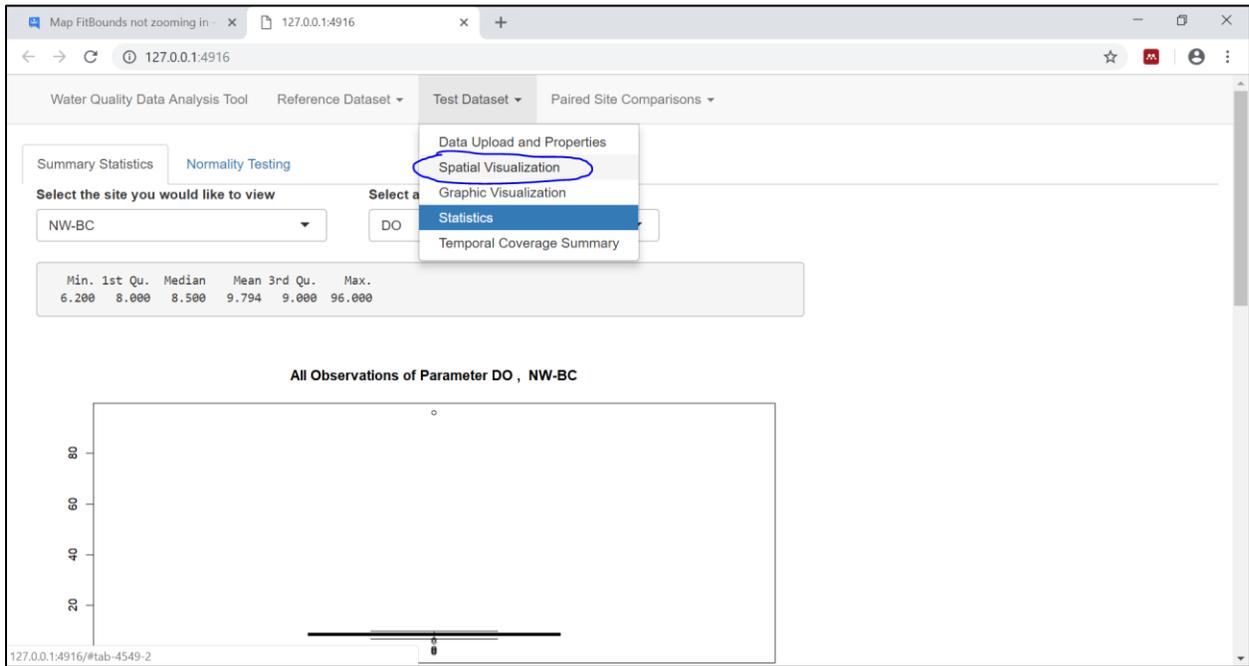
Recall that line 41 holds the extremely high value. Change the slider to include all of the data by clicking the leftmost circle and dragging it to the left end of the slider.

Click on the Show entries dropdown and select 50.

Scroll down to line 41. We can see that the day before the 96 was recorded, the DO at the same location was 9.6.

36	NW-BC	-126.8654	65.283	2016-06-05	7	6.2
37	NW-BC	-126.8654	65.283	2016-06-06	7	6.3
38	NW-BC	-126.8654	65.283	2016-06-10	7	8.7
39	NW-BC	-126.8654	65.283	2016-06-13	7	9.6
40	NW-BC	-126.8654	65.283	2016-06-20	7	9.6
41	NW-BC	-126.8654	65.283	2016-06-21	7.1	96
42	NW-BC	-126.8654	65.283	2016-07-02	7.3	9.4
43	NW-BC	-126.8654	65.283	2016-07-03	7.2	9.2
44	NW-BC	-126.8654	65.283	2016-07-08	7.2	9.1
45	NW-BC	-126.8654	65.283	2016-07-11	7.2	9.1
46	NW-BC	-126.8654	65.283	2016-07-25	7.2	9.1
47	NW-BC	-126.8654	65.283	2016-07-26	7.2	9
48	NW-BC	-126.8654	65.283	2016-07-26	7.2	9.3
49	NW-BC	-126.8654	65.283	2017-08-25	7.7	9
50	NW-BC	-126.8654	65.283	2017-08-24	7.8	8.9

Showing 1 to 50 of 140 entries   Previous 1 2 3 Next

It is a likely guess, then, that a period was accidentally removed, changing what should have been a 9.6 to a 96. To be certain though, it is a good idea to check with the individual or individuals who originally recorded the data.

Take some time to navigate around and experiment with the Spatial Visualization and Statistics pages.

Please return to the survey website and complete the questions in Section 2. Please return to this document when you have finished and proceed to Step 6 (do not exit the survey).

Step 6: Read the following prompt and follow the instructions to see how the Temporal Coverage Summary page can be used to investigate data gaps

Your community/organization has been successfully collecting local water quality data for a period of years. This year, you have been approached by several enthusiastic new volunteers. They ask you when and where (if at all) they should focus their data-collecting efforts. You decide to check if any months or sites are being under-sampled, in order to improve the temporal and spatial coverage of your dataset.

Under the Test Dataset tab, click on Temporal Coverage Summary.
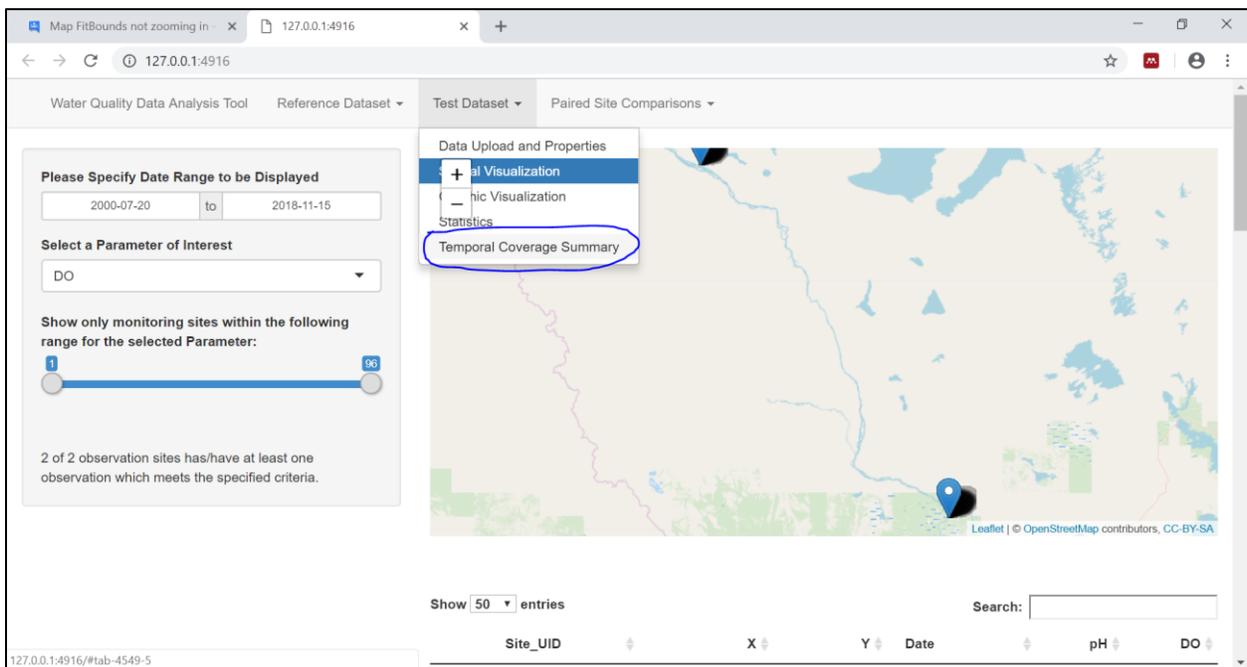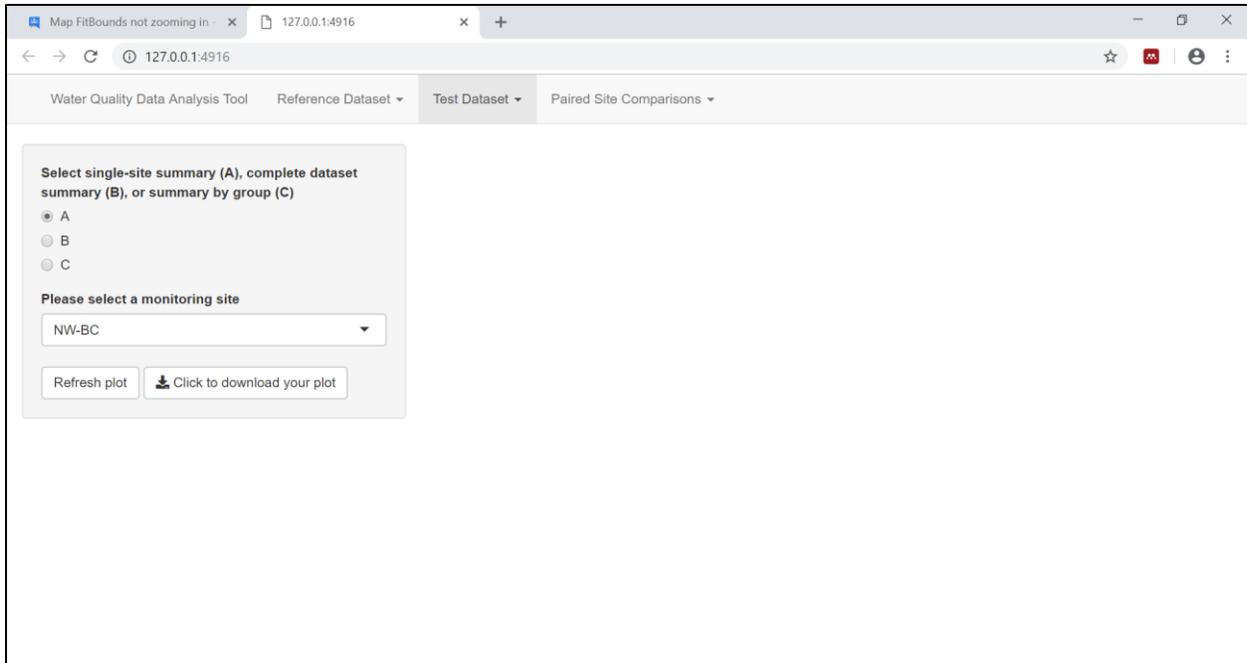
This page counts the number of observations by month and sorts the data by year. This allows a user to see which months and years are most frequently sampled, and to compare sample sizes between years.

Select circle B, to view the entire dataset. Click Refresh plot.
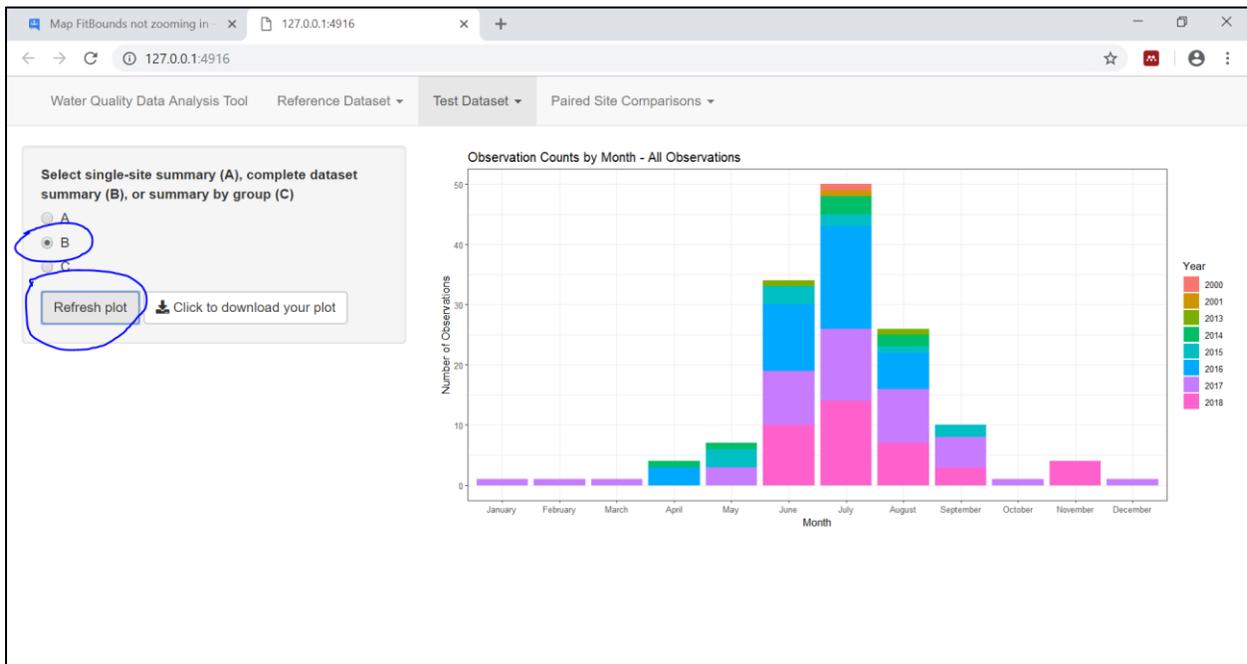
The number of observations are displayed along the y (vertical) axis. Month names are displayed along the bottom x (horizontal) axis. One line in the uploaded file corresponds to one observation. Notice that

*the months of June, July, and August have the most observations. Depending on your locale and the dominant weather/climate conditions, this could make sense in terms of ice coverage and accessibility.*

Select circle A, to isolate a single site. The first site, NW-BC, will automatically populate the drop-down menu. Click Refresh plot.
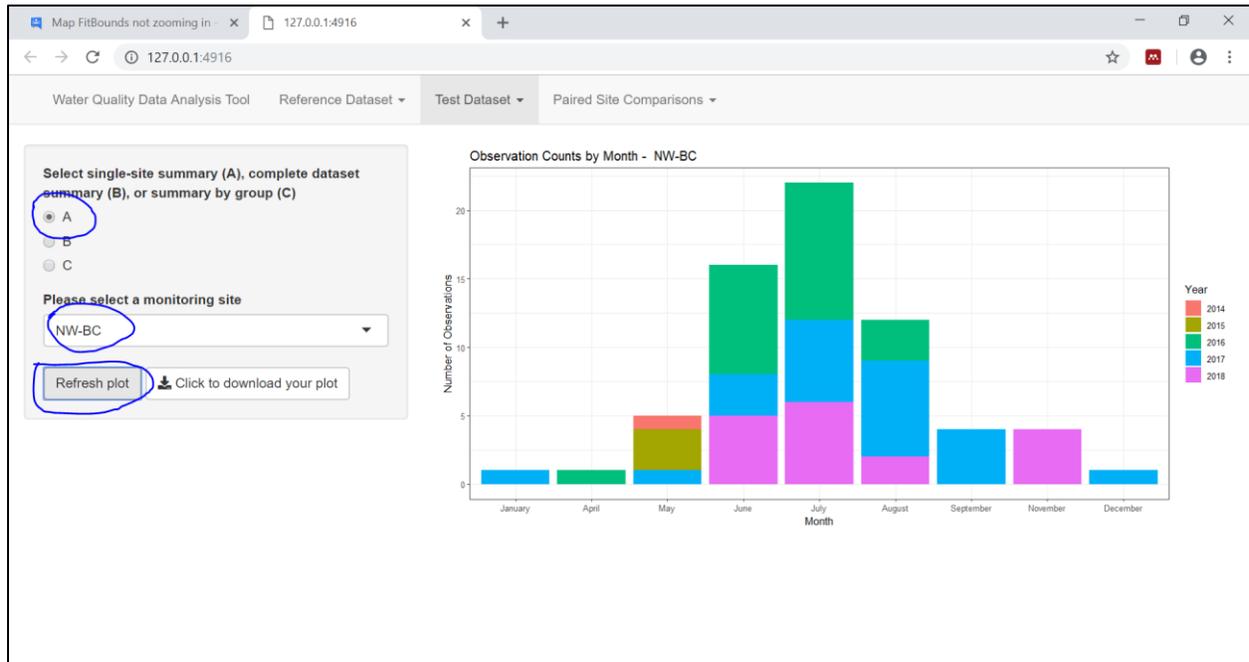
Take a moment to examine the new graph. Note the observation counts and range of years.

Use the drop-down menu to select the other site, SIM-LR. Click Refresh plot.
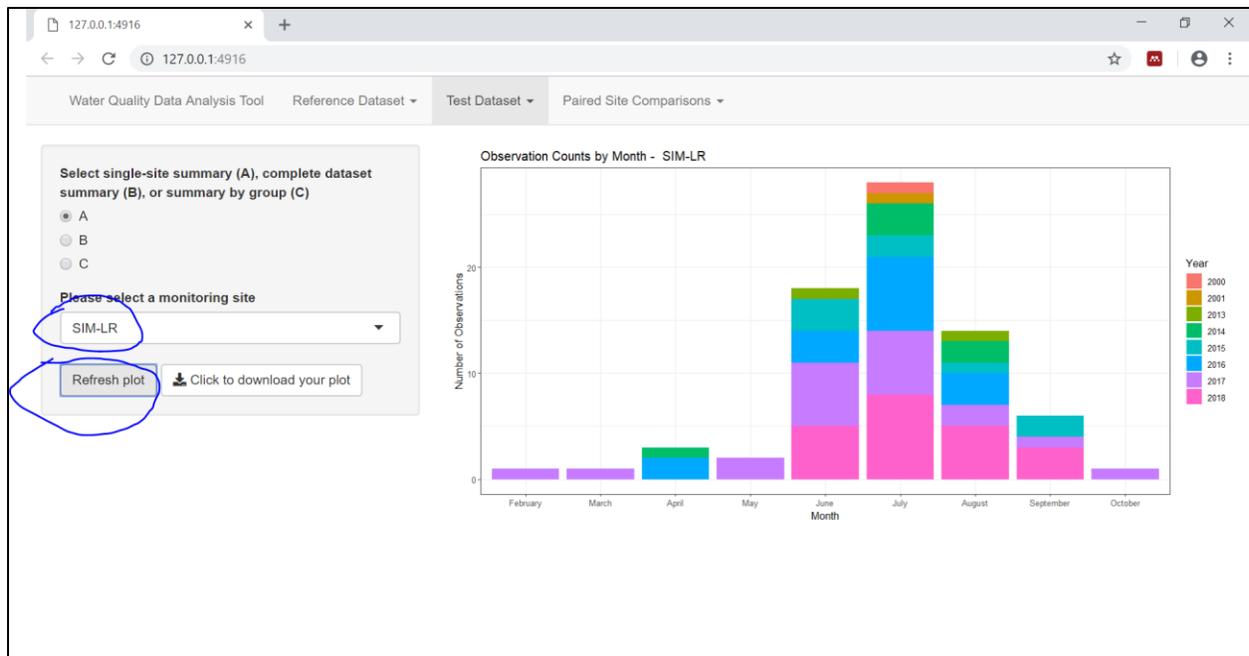
Take a moment to examine the new graph.

Note that the observation numbers are mostly even between the two sites. Thus, no site necessarily needs to be more sampled than the other.

Each image can be downloaded by clicking on the button labelled "Click to download your plot".

Take some time to experiment with the temporal coverage summary page. Note that by selecting Summary by Group, you are prompted to specify a column with group names, and then to select a group of interest from that column. This can be used if each site in your dataset falls within a larger group or geographic extent, ie a region, town, municipality, or other aggregate area.
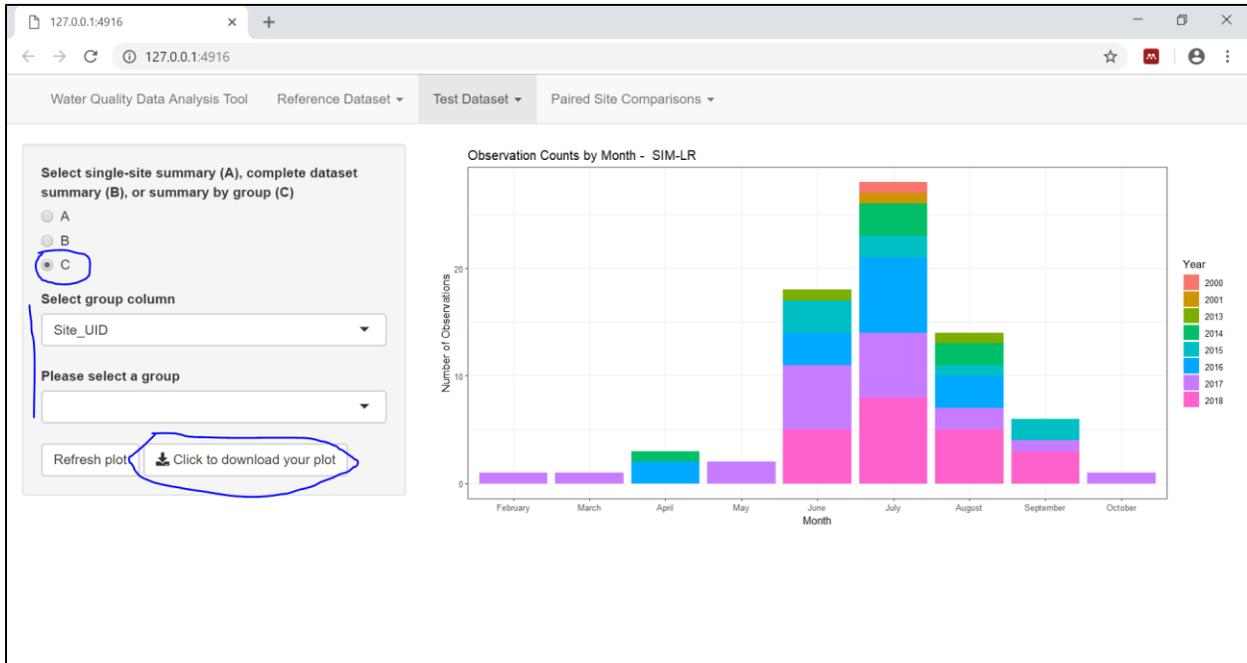
Please return to the survey website and complete the questions in Section 3. Please return to this document when you have finished and proceed to Step 7 (do not exit the survey).

Step 7: Read the following prompt and follow the instructions to use the Graphic Visualization page to connect data trends to specific weather events.

In early June of 2018, there was an unusually intense rainstorm that lasted for a number of days. Nonetheless, some dedicated volunteers decided to collect water quality samples from the site NW-BC during and after the rain event. Concerned that such intense rainstorms may be becoming more frequent, the volunteers wanted to see if there was any noticeable change to the water's pH at that site in response to the rainstorm – something to watch for with future storms.

Under the Test Dataset tab, click on Graphic Visualization.
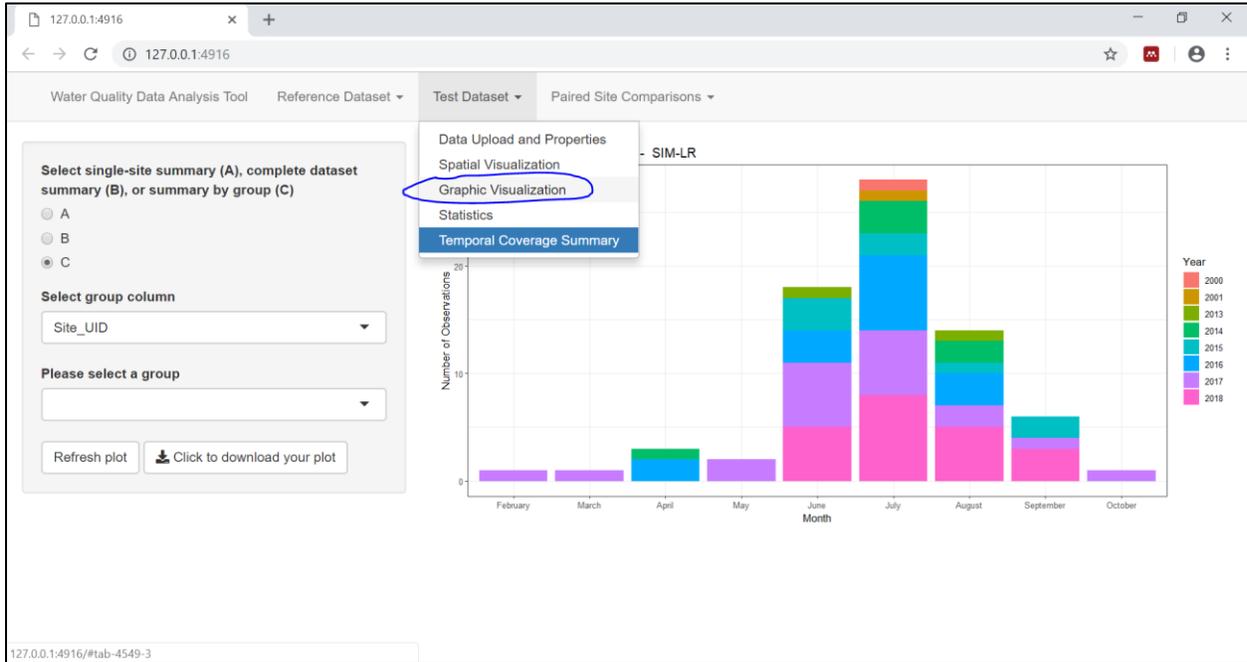
The first drop-down menu allows you to choose between visualizing a single variable or two variables. Since we are interested in pH data throughout the month of June 2017, we need to visualize two variables: pH and Date.

Select 2 from the first drop-down menu.

The second drop-down menu is for selecting the site or sites you are interested in. Select NW-BC by clicking in the second box.

Since we are only interested in June 2017, we have to filter out the rest of the data. Use the third box to select June.

Use the fourth box to select 2017.

Use the next box ("Select variable to display (Y)") to select pH, our first variable.

Use the next box to select Date.

To make any trends easier to see, select Line under "Choose plot type".

Finally, you can add a title of your choice by typing in the box labelled "Enter the name of the graph".

Click the "refresh map and update settings" button and return to the top of the page to view your new graph.
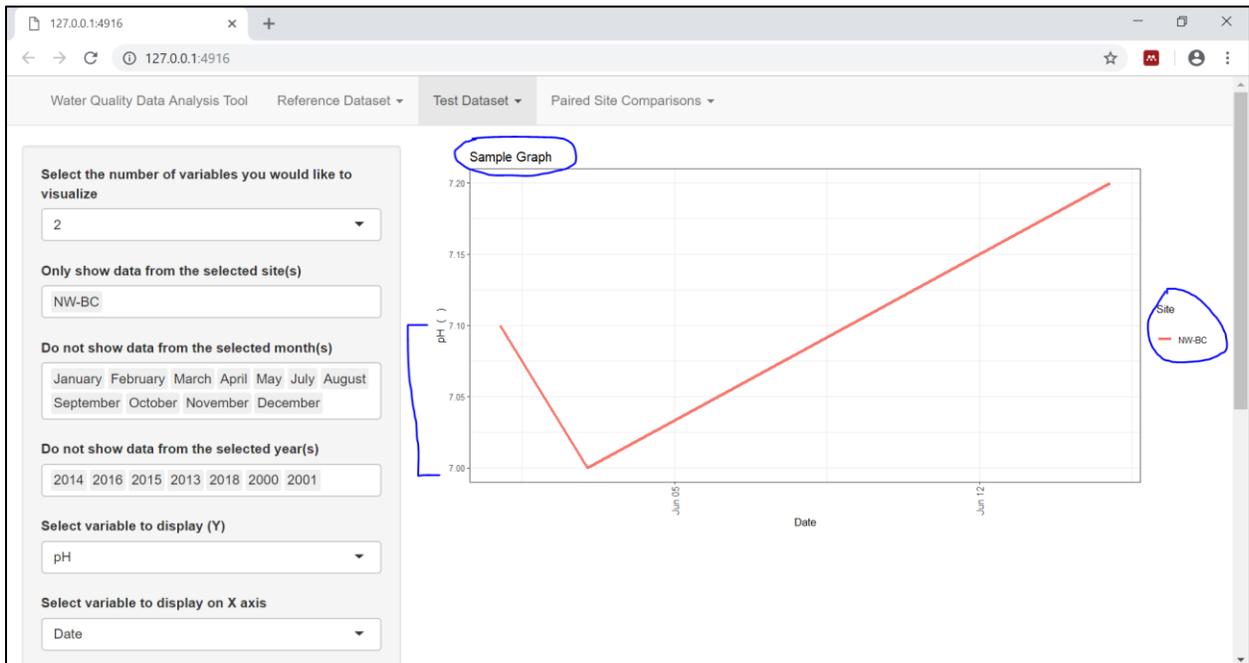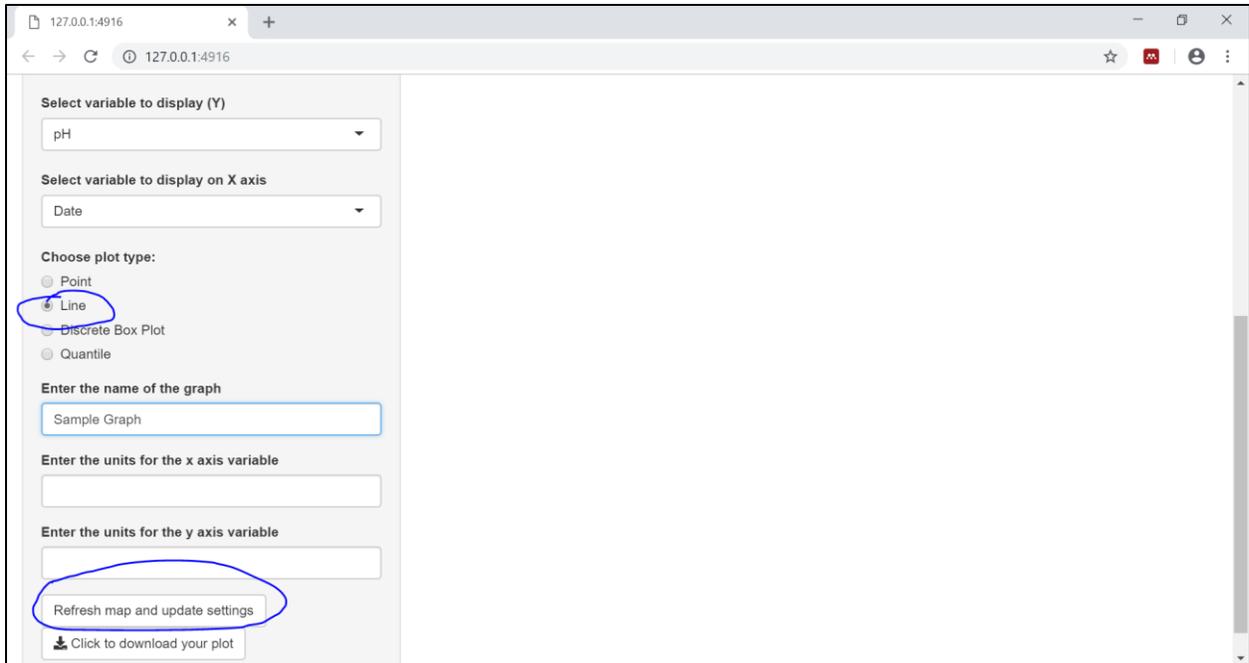
There is a noticeable drop in pH in the early part of the month. This could potentially be a result of the excess runoff caused by the storm event. Note, however, that while the line seems to drop substantially, by observing the numbers along the vertical access it can be seen that the data did not change by much.

Take some time to navigate around and to experiment with the Graphic Visualization page using different inputs and selections. Make note of the function of the various user inputs (Title, units, etc) and of the options offered when the number of variables to be displayed is 1. Remember to click on "Refresh map and update settings" to view each new graphic. Some example results are shown below (insert image).
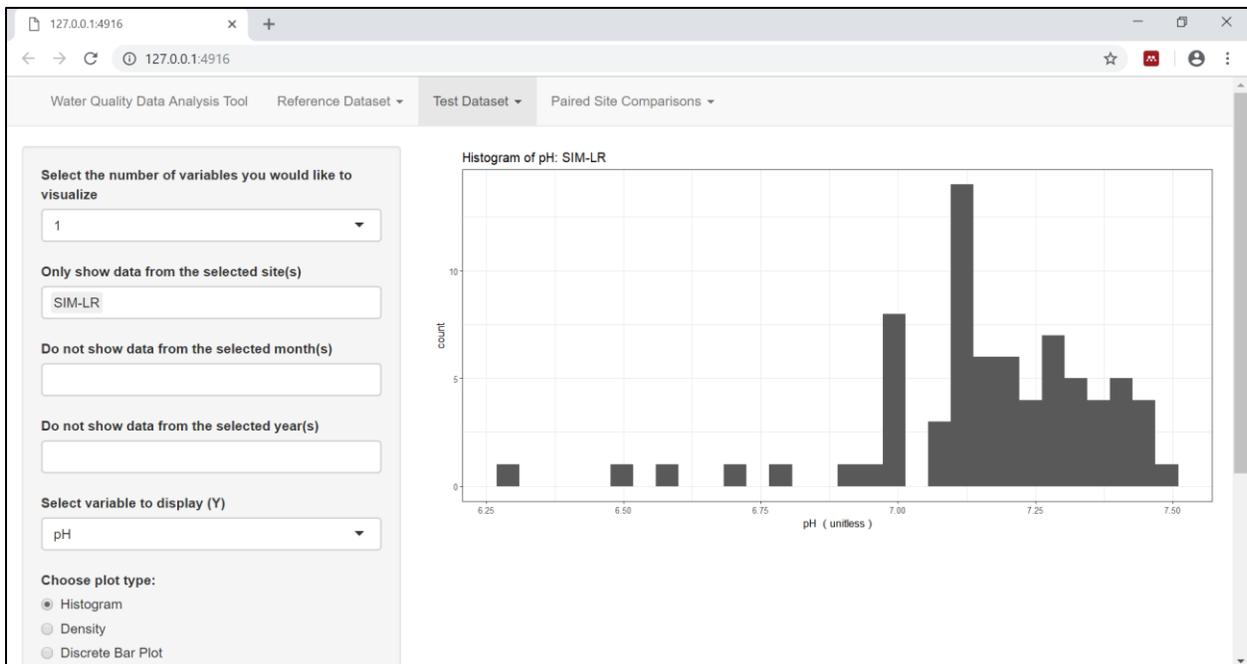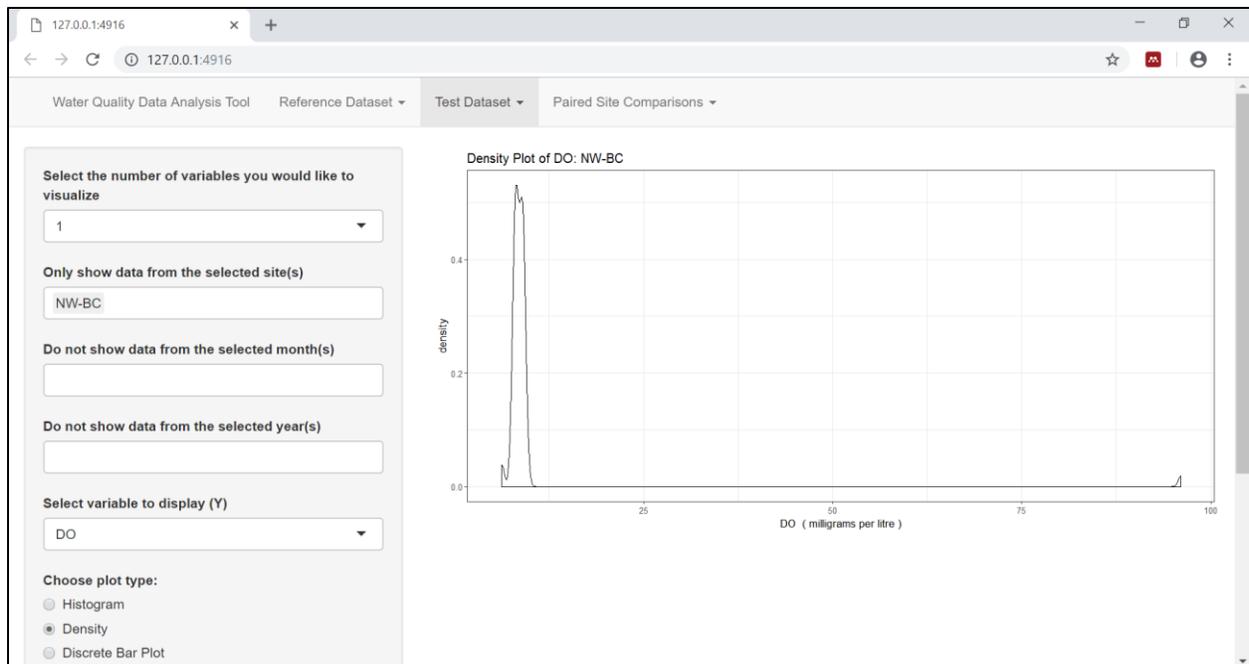
Please return to the survey website and complete the questions in Section 4. Please return to this document when you have finished and proceed to Step 8 (do not exit the survey).

Step 8: Read the following prompt and follow the instructions to see how the tool can facilitate data sharing and collaboration within the CBWQM community.

Your community/organization has been working hard to collect water quality data in your area. A nearby community contacts you asking for help. They say that they have noticed something strange with the June 2017 pH data at one of their sites, which happens to be fairly close to one of your sites(SIM-LR). They ask if you would be willing to share with them what your data shows for that site, so that they may compare the two sites and try to determine the problem with their data, if there is one. You decide that you are willing to help but would prefer not to send them your entire dataset. You now need to share with them a portion of your data.

A good first step is to verify that you have data for the site and time in question – SIM-LR in June of 2017.

Under the Test Dataset tab, click on Temporal Coverage Summary. Ensure that the circle for A (single-site summary) is selected. Use the drop-down menu to select the site SIM-LR. Click on Refresh Plot.
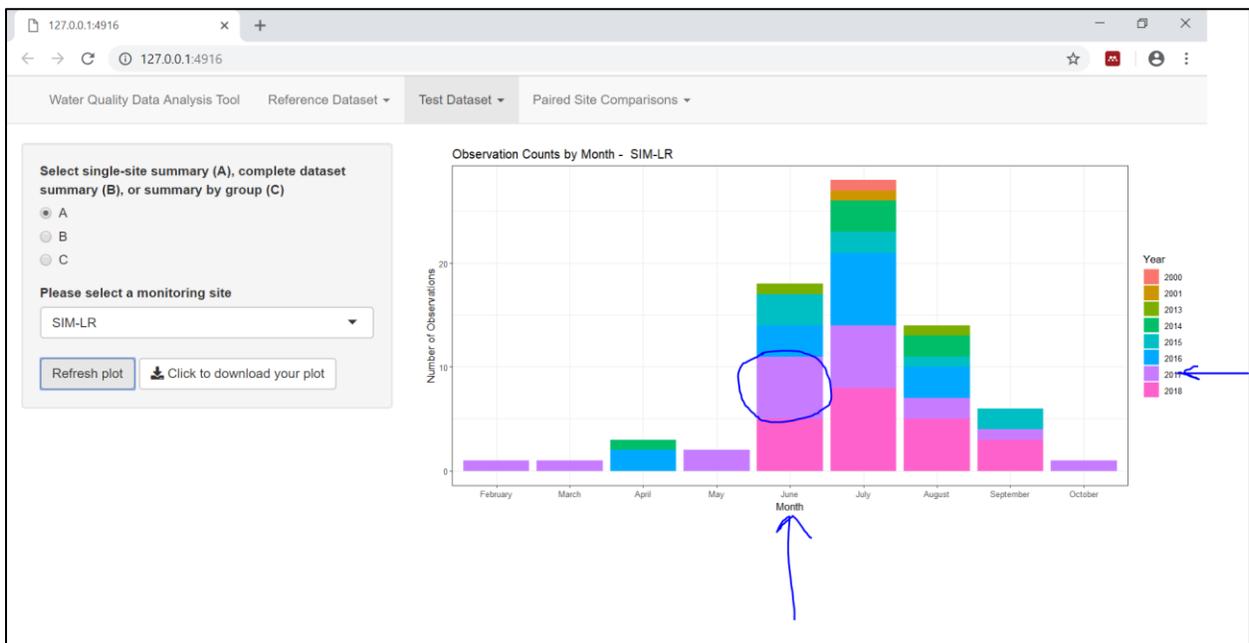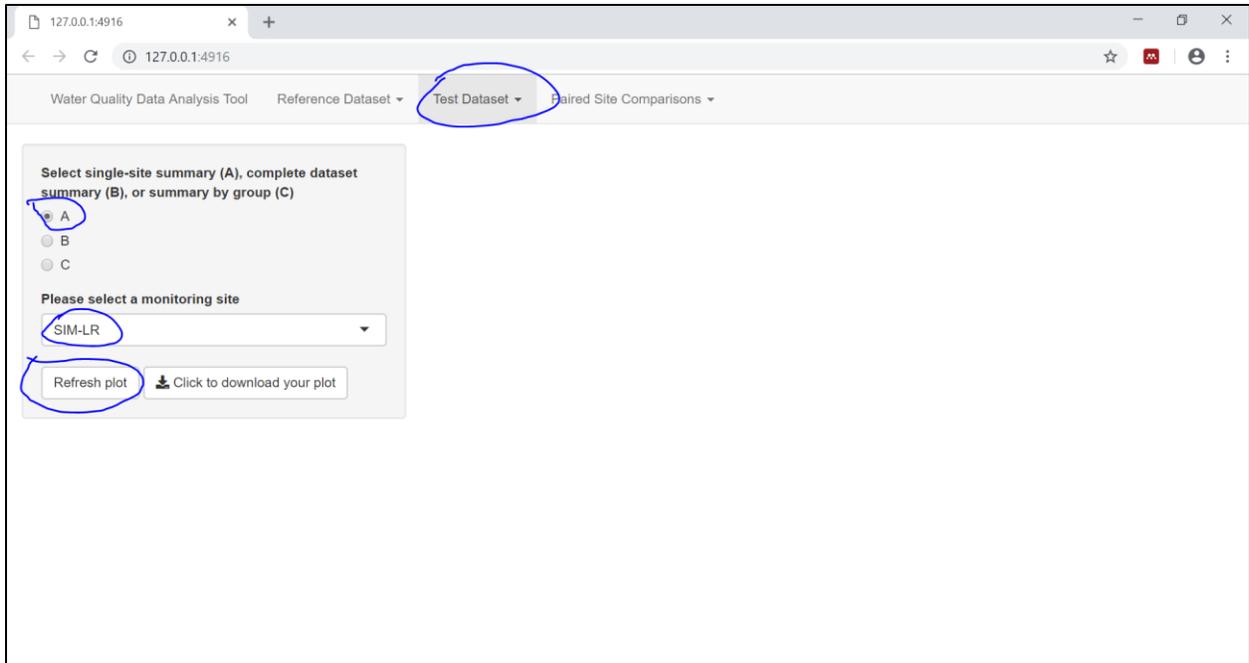
Observe that the year 2017 is displayed in a purple colour (second from the bottom in the legend, found to the right of the graph). Find the column for the month of June and see if the purple colour of 2017 is in this column.

It is, so you know that you have some data to show to the other community/organization.

Under the Test Dataset tab, click on Graphic Visualization.

Since we are interested in pH data throughout the month of June 2017, we need to visualize two variables: pH and Date.

Select 2 from the first drop-down menu.

Select SIM-LR by clicking in the second box.

Since we are only interested in June 2017, we have to filter out the rest of the data.

Use the third box to select all months but June.

Use the fourth box to select all years but 2017.

Use the next box ("Select variable to display (Y)") to select pH, our first variable.

Use the next box to select Date.
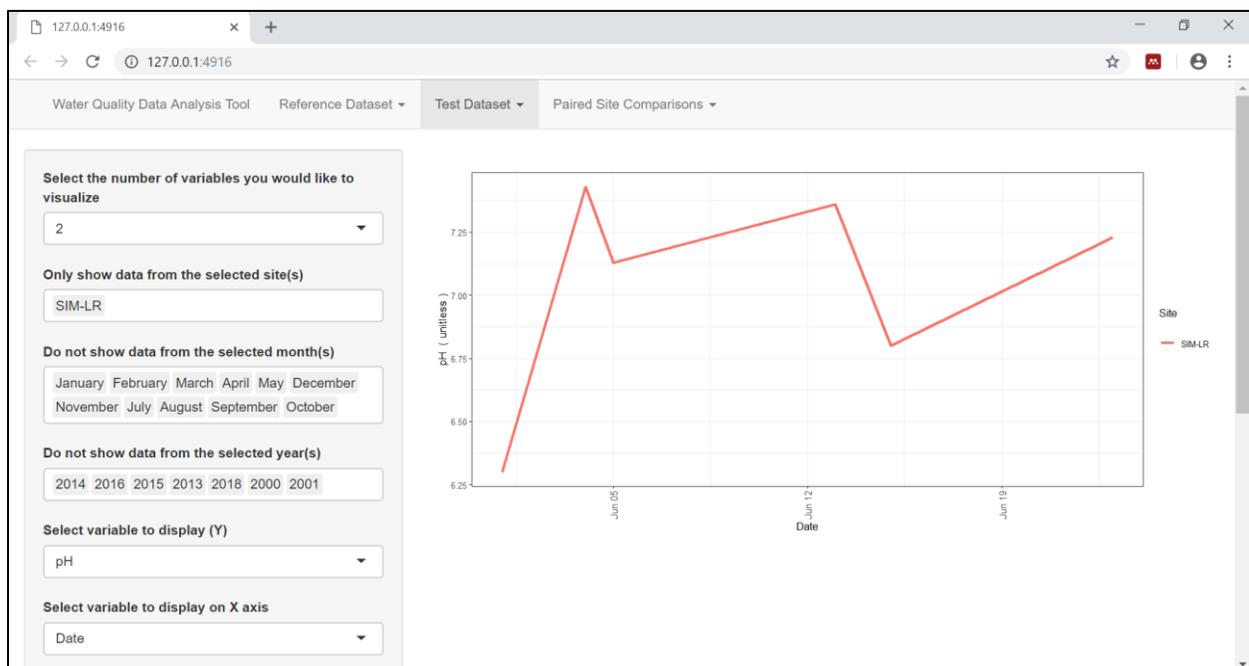
To make any trends easier to see, select Line under "Choose plot type".

Finally, you can add a title of your choice by typing in the box labelled "Enter the name of the graph".
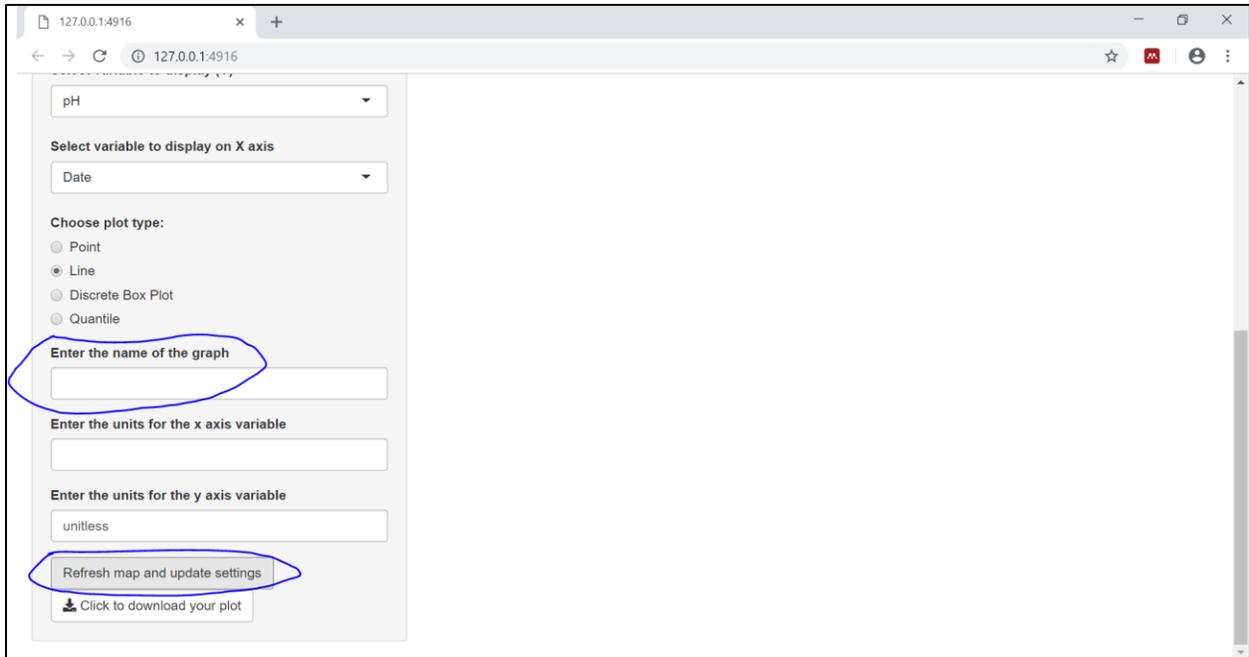
Click the "refresh map and update settings" button and return to the top of the page.
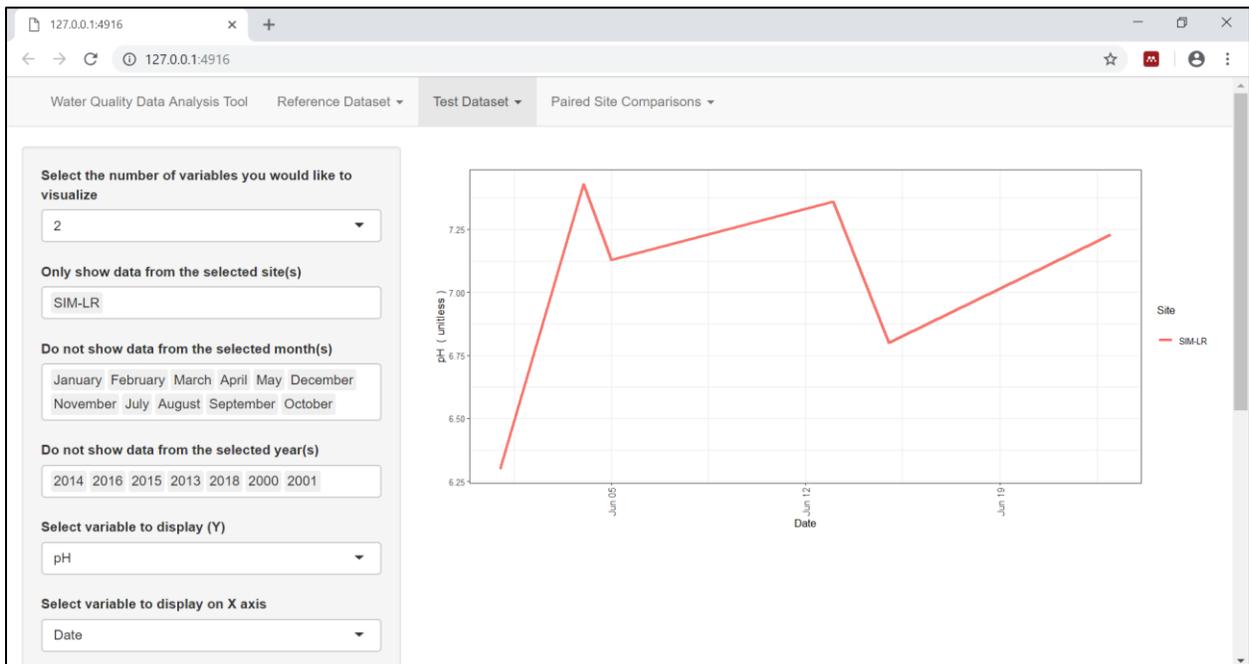
The last step is to save this image, so it can be sent away. Scroll down and click on the button labelled "Click to download your plot".

Please return to the survey website and complete the questions in Section 5. Please return to this document when you have finished and proceed to Step 9 (do not exit the survey).

Step 9: Read the following prompt and follow the instructions to see how the tool can be used to analyze CBM data by comparing it to a set of long-term, baseline data.

Your community/organization wishes to use a set of community-collected data to create a new pamphlet promoting water quality awareness and resource management. Some individuals express concern over the validity and accuracy of the CBM data. They explain that one way to see if the CBM data is realistic is to compare it to another data set, from a regulated monitoring system. You agree that a comparison would be a good idea – you might learn something about the CBM data as well!

Previously, you uploaded two files: Ref_Sample and CBM_Sample. The data contained in Ref_Sample will serve as the regulated baseline.

Under the Paired Sites Comparison tab, click on Site Pairings.
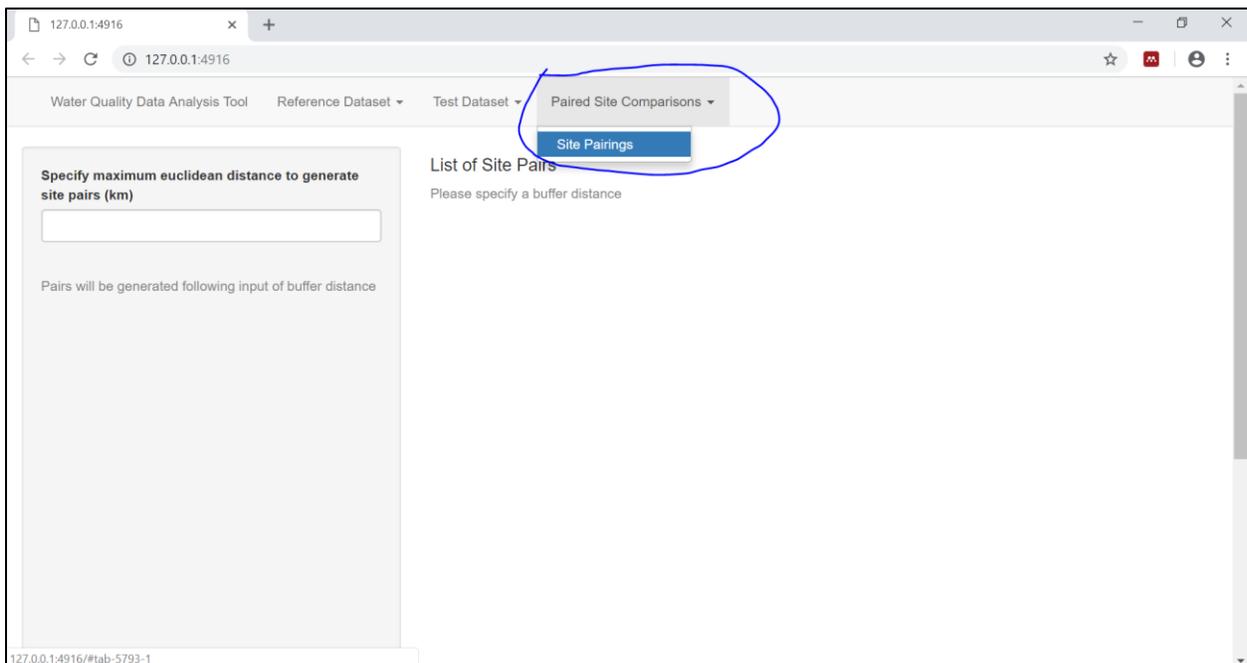
To compare two data sets, one site from each set can be compared to each other. Sites closer in proximity are more likely to experience similar weather and hydrologic conditions. This page asks for a pairing distance – how close together must two sites be (one from the reference file, the other from the CBM file), to be considered a 'pair'?

Type the number 2 into the text box and observe the result.

Take a moment and carefully examine the different boxes, graphics, and text that now populate the Site Pairings page.

The top centre of the page is a table (List of Site Pairs), which enumerates all of the site pairs whose two sites are within the specified distance (2 kilometers) of each other. The measured distance between the sites in each pair is also given, and each pairing is assigned a number.

Just below the table are three drop-down boxes. The first box on the left, labelled "Select the site pair you would like to view", allows the user to switch between the pairs in the list. Whichever pair is selected will appear on the map to the left.

Use the first drop-down box to select site pair 2 and observe the resulting change to the map.

The remaining drop-down boxes function to select the parameter of interest, and to specify what data from the test dataset will be tested against the baseline. Scroll down and note the grey text. For the statistical comparison to be successful, the parameters being used must both be numeric (represented by numbers). For example, Dissolved Oxygen (DO) is a numeric parameter, but Date is not.

Use the drop-down box labelled "Select a parameter from the Reference Dataset" to select pH. Repeat for the drop-down box labelled "Select a parameter from the Test Dataset". Note that, when comparing between two data sets, you must be certain that the units are consistent.

Separate selections are done to account for the possibility that one physical parameter could be referred to in different ways by each file. For example, "calcium carbonate" and "alkalinity" are sometimes used interchangeably.

Select the month of May from the appropriate drop-down box (labelled "Select the month you would like to view"). Click on the drop-down labelled "Select the test year" and observe that the only option is 2017. This is because, in the test (CBM) dataset, the only year with at least one observation in May is 2017. Press the button below the map on the left side of the page to run the test. You will now see the following text:

With 95% confidence, the test data does not fall outside the critical effect size (normal range) of the reference data

This statement declares that the CBM observation for May 2017 is realistic, when compared against all reference observations in May from 2012-2017 (per the slider labelled "Years included in the slider's range will be used as a long-term baseline").

Below the text, two graphs display the specified data used in the comparison process.

With 95% confidence, the test data does not fall outside the critical effect size (normal range) of the reference data

Test Data - May 2017

Reference Data - May
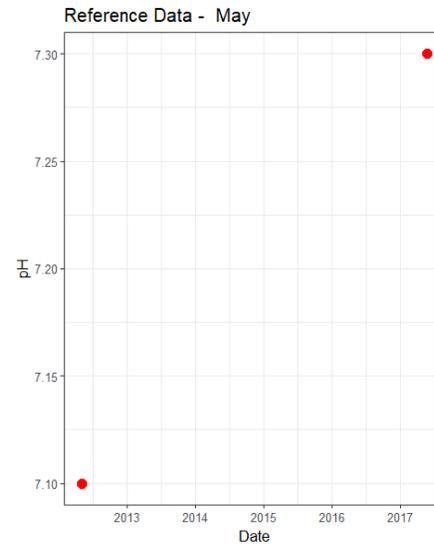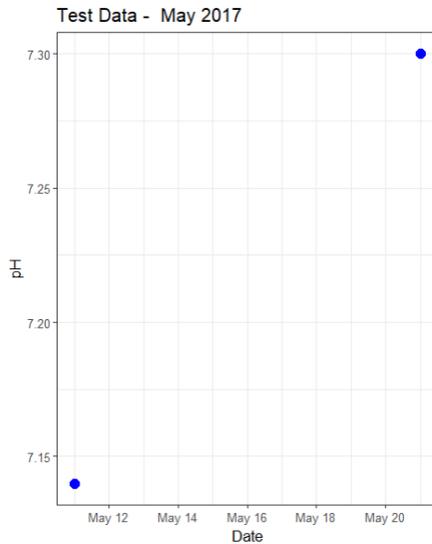
Experiment with different sites, parameters, and dates to get a sense of how the comparison process works. Remember to only compare pH to pH or DO to DO, as those are the only numeric water quality parameters provided in the reference and test datasets.

Change the box selections to compare the CBM Dissolved Oxygen (DO) data from June 2017 with the baseline data. In this case, the CBM data is statistically different from the reference data.

(Picture once equation confirmed).

By working through the various months and years of the test (CBM) data, it is possible to get a sense of the data's statistical validity, with respect to the more regulated and controlled baseline data.

Please return to the survey website and complete the remaining questions.