

WQP Data Discovery Tool – Version 1.1

Release Date: February 3, 2017

Quick Start Guide

The WQP Data Discovery tool is a desktop application that provides an easy to use interface allowing users to query, summarize, QC and display data from the WQP. The tool uses open source R, a statistical programming language and several add-on packages, to visualize the data selected from the portal and to assist users in selecting data they may need for analysis.

INSTALLATION

WQP Data Discovery Tool Installation Package (Windows compatible). You can download the installation package from <https://www.epa.gov/waterdata/water-quality-portal-data-discovery-tool>. The *WQPdatadiscoverydesktoptoolv1.1.zip* file that you download from this site contains everything you need to run the tool. Note, the zip file is large and will take a few minutes to download and extract.

WQP Data Discovery Tool Installation Package (Mac compatible). Download the *WQPdatadiscoverytoolv1.1.zip* file to your computer and unzip it. Note the location where you stored these files. You will need to point R to this location to run the tool. Prior to running the tool, you will need to have R installed. If you have R installed skip down to the **Installation for Users with R section**. If you do not have R installed, proceed to the **Installation for Users without R** section below.

Installation for users without R. RTI recommends running R 3.3.1 or higher. Follow the steps below to install R.

1. Download R from the [R website, http://cran.cnr.berkeley.edu/](http://cran.cnr.berkeley.edu/)
2. Select the appropriate version of R for your computer and follow the directions below to download and install R:
 - a. Click on **Download R for Linux**
 - i. Chose the appropriate version
 - b. Click on **Download R for (Mac) OS X**
 - i. Chose version that is compatible with your current Operating System
 - c. Click on **Download R for Windows**
 - i. Click on **Base**
 - ii. Click on **Download R 3.2.3 for Windows**
3. Run the installer.

Installation for users with R. If you already have R installed, ensure that your version is 3.3.1 or higher. If you need to update R there are several options available:

- Install a newer version of R (see directions in Installation for users without R section)

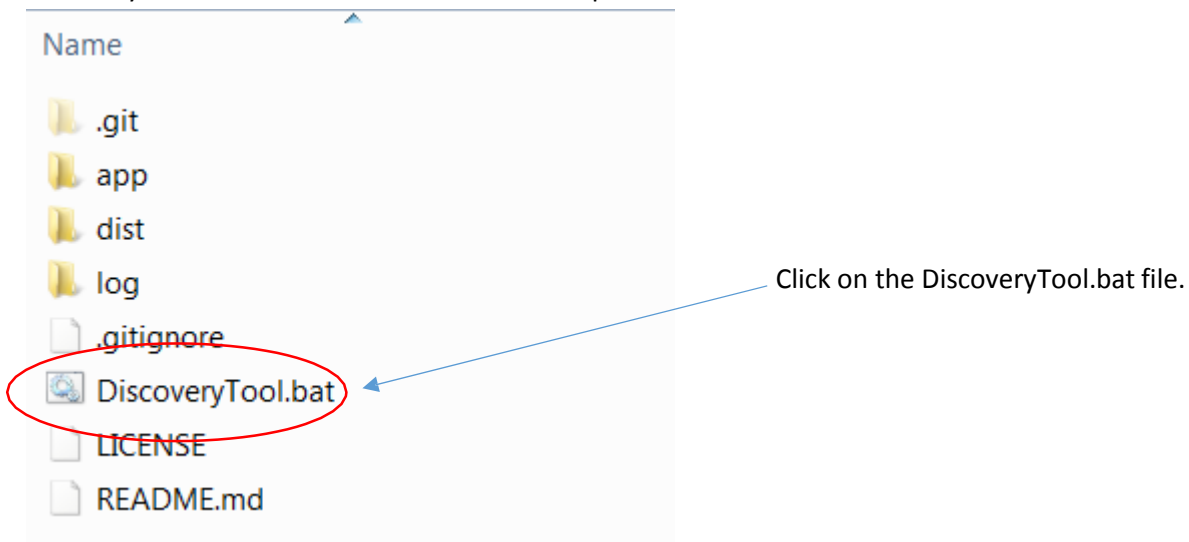
- Use the troubleshooting.R script available with this tool. To use this script, open the “Troubleshooting.pdf” document and follow the Instructions.
- For the experienced R user here is a link on how to update R: <http://www.r-statistics.com/2013/03/updating-r-from-r-on-windows-using-the-installr-package/>

You will also need to update your R packages. To update packages, visit this link: <http://www.inside-r.org/r-doc/utils/update.packages>.

LAUNCHING THE APP

The WQP Data Discovery Tool is launched via the R application and opens in your system's default internet browser.

Windows Users. To launch the app simply click on the DiscoveryTool.bat file. The first time you use the tool it may take a few minutes to launch. Subsequent sessions will load much faster.



Mac Users. Follow the directions below to launch the App:

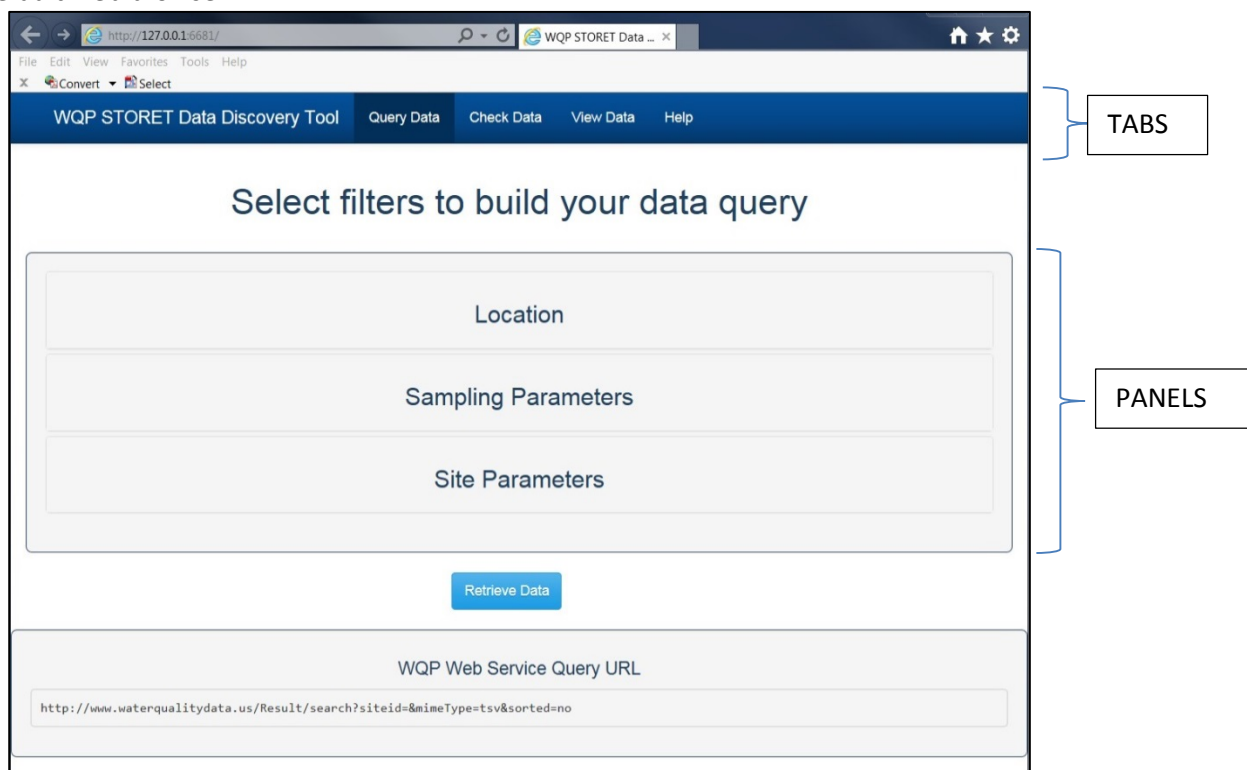
1. Open R by clicking on the R desktop icon or from your computers program files
2. Select File → Change dir...
 - a. Navigate to the location where you unzipped the **Installation Package**
 - b. Select the MISSION-WQP Folder and click OK
3. Select File → Source R Code...
 - a. Select Launch_visualization.R and click Open

The first time you run the app it will check to see if you have the necessary add-on R packages. If any packages are missing the tool will install them for you. It takes a few minutes for R to open and install all the packages needed to run the app. This step is skipped the next time you use the app.

4. The app will open in your system's default web browser. The tool is compatible with: Internet Explorer 11, Mozilla Firefox, and Google Chrome. If your default browser is incompatible you can copy the web link and paste it into a compatible browser.

DATA DISCOVERY TOOL

Once you have launched the Data Discovery Tool you will see the screen below open in your system's default web browser.



On the first tab, Query Data, you can build your query. The second tab, Check Data, allows you to view the data in tabular form and QA/QC it. The third tab, View Data, allows you to map, filter, and graph the data you've imported. The fourth tab, Help, contains help information and tool tips.

QUERY DATA TAB

The '**Query Data**' tab allows users to select filters to build a query. This tab sorts the web query into three panels: Location, Sampling Parameters and Site Parameters. The Data Discovery Tool will only allow you to import data when your query returns less than 200,000 records.

Generating a Query. To begin writing a query, click on a panel label to expand it and view the query parameters. When you are done making your selections in a panel, you can click on the label again to close it. Hovering the cursor over a field will produce a pop-up description of the field. This feature is available for all data query fields. When you are done building your query, click on the Retrieve Data Button.

WQP STORET Data Discovery Tool

Query Data

Check Data

View Data

Help

Select filters to build your data query

Click here to expand and close a panel.

Location

Sampling Parameters

Site Parameters

Click here to run the query.

Retrieve Data

WQP Web Service Query URL

`http://www.waterqualitydata.us/Result/search?siteid=&mimeType=tsv&sorted=no`

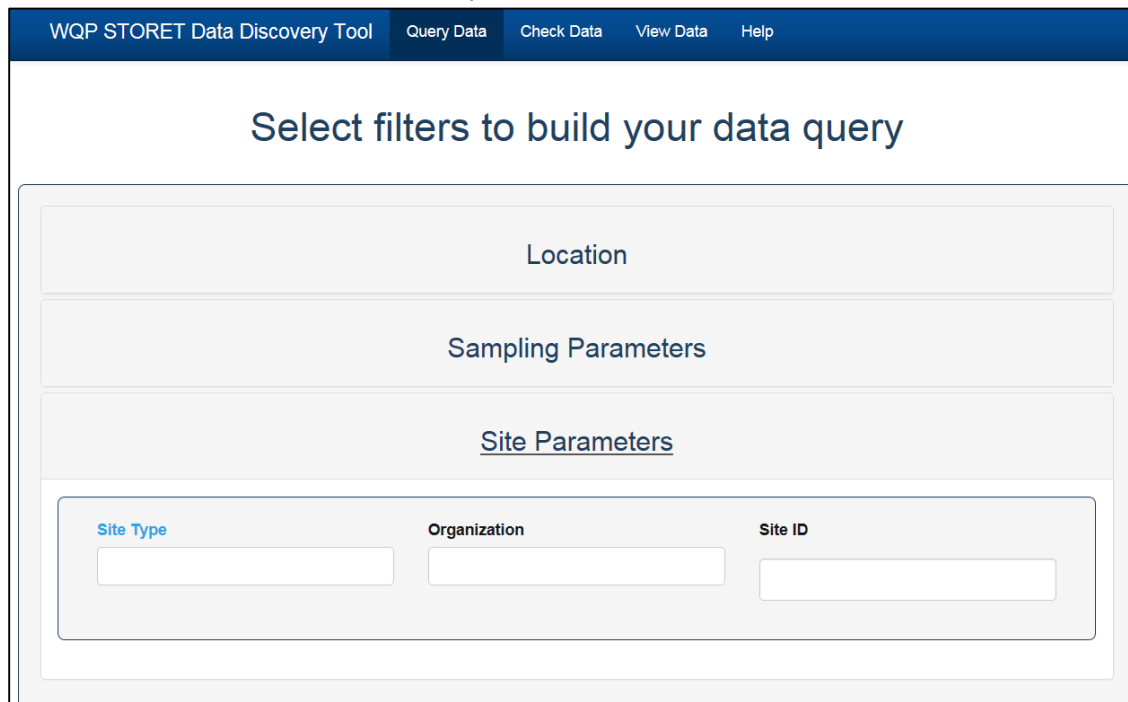
Location. State and county can be selected from a drop-down menu. HUC, Point Location and Bounding box must be entered manually.

The screenshot shows the 'WQP STORET Data Discovery Tool' interface. The top navigation bar includes 'Query Data', 'Check Data', 'View Data', and 'Help'. The main heading is 'Select filters to build your data query'. Below this, the 'Location' section is highlighted. It contains three main input areas: 'Place' with 'Choose a State' and 'Choose a County' dropdowns, and an 'HUC' input field with a placeholder 'Enter a huc code (ex. 03020201)'. The 'Point Location' section includes 'Latitude (decimal degrees)', 'Longitude (decimal degrees)', and 'Distance from point (miles)' input fields. The 'Bounding Box' section includes 'North', 'South', 'East', and 'West' input fields. All input fields currently contain the number '0'.

Sampling. Sample Media can be selected from a drop-down menu. To enter a characteristic group or characteristic begin typing the first few letters in the dialog box and then select the appropriate choice from the drop-down list. Alternatively, you can type in the name and click Enter.

The screenshot shows the 'WQP STORET Data Discovery Tool' interface. The top navigation bar includes 'Query Data', 'Check Data', 'View Data', and 'Help'. The main heading is 'Select filters to build your data query'. Below this, the 'Sampling Parameters' section is highlighted. It contains a 'Date Range (MM-DD-YYYY)' section with 'From' and 'To' input fields. To the right of the date range is a 'Sample Media' dropdown menu. Further right are 'Characteristic Group' and 'Characteristic' input fields. All input fields are currently empty.

Site. Site Type can be selected from a drop-down list. To enter an organization begin typing the first few letters and then select it from the drop-down list. Alternatively, you can type in the name and click Enter. Site ID must be entered manually.



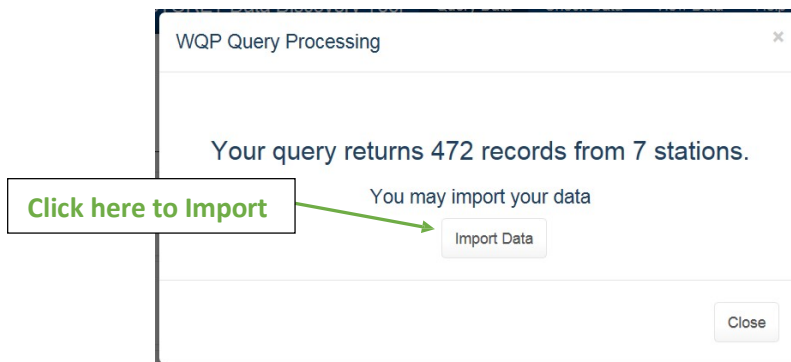
The screenshot shows the 'WQP STORET Data Discovery Tool' interface. At the top is a navigation bar with tabs: 'Query Data', 'Check Data', 'View Data', and 'Help'. The main heading is 'Select filters to build your data query'. Below this are three stacked filter sections: 'Location', 'Sampling Parameters', and 'Site Parameters'. The 'Site Parameters' section is expanded, showing three input fields: 'Site Type' (with a blue label), 'Organization', and 'Site ID'.

WQP Web Service Query URL. At the base of the 'Query Data' tab the web service URL is displayed. As you apply and modify query parameters the query URL will automatically update.



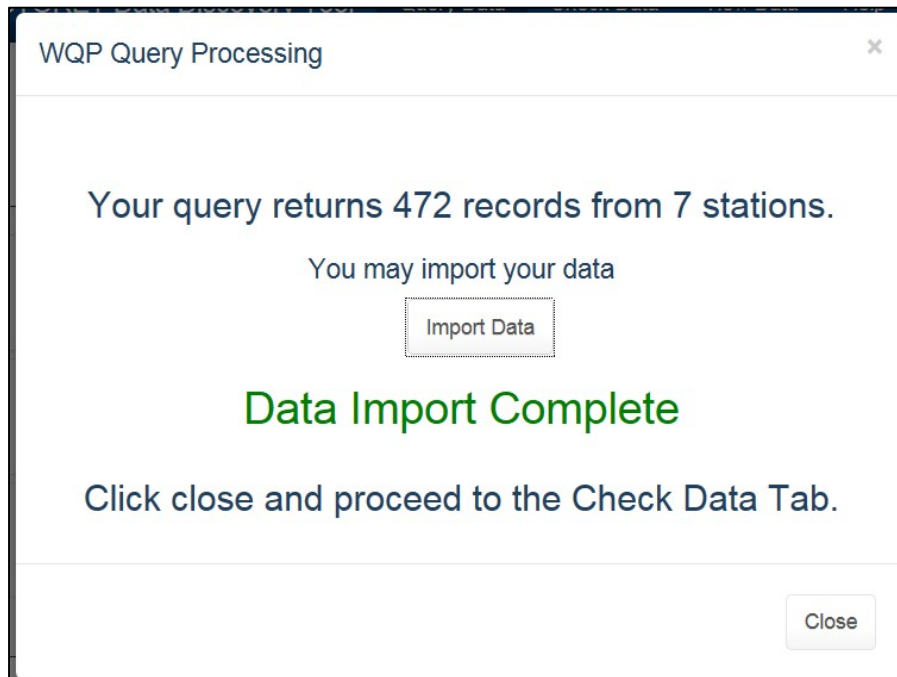
The screenshot shows a box titled 'WQP Web Service Query URL'. Inside the box, the URL 'http://www.waterqualitydata.us/Result/search?siteid=&mimeType=tsv&sorted=no' is displayed.

Retrieving/Importing Data. Once you click the Retrieve Data button the tool will begin gathering data from the Portal and check the number of records and stations. A separate screen will open and show the number of records/stations for the query. Click on Import Data, or refine query to get fewer results if necessary.



The screenshot shows a dialog box titled 'WQP Query Processing'. It contains the text 'Your query returns 472 records from 7 stations.' Below this, it says 'You may import your data' and there is an 'Import Data' button. A green callout box with the text 'Click here to Import' has an arrow pointing to the 'Import Data' button. At the bottom right of the dialog is a 'Close' button.

A **Data Import Complete** message will display when the download is finished. Note: depending on the number of records the download may take a few minutes. The data are now available for use in the tool.



CHECK DATA TAB

This tab is used to examine the data imported from the Portal. The Home tab provides a **Data Summary**, a menu to **select a method to deal with Non-Detections** and a description of **Available Data Sets**.

WQP STORET Data Discovery Tool
Query Data
Check Data
View Data
Help

Home
All Data
Non Detects
W/O Units
W/O Methods
Duplicates
Filtered Data
Summary

Data Summary

The total imported data set contains **472** records from **7** stations representing **11** parameters with, **118** non-detects. **3** records without units **0** records without methods **1** duplicate records

Date: 2016-03-01 15:12:55

Web Query:

http://www.waterqualitydata.us/Result/search?organization=YAN_WQX&siteid=&huc=&mimeType=tsv&sorted=no

Please select a method to deal with Non-Detections

☒ Ignore Non-Detections - remove from data set
☐ Set Non-Detections equal to zero
☐ Set Non-Detections equal to the Limit of Detection
☐ Set Non-Detections equal to the 1/2 times the Limit of Detection

Available Data Sets

All Data:
This table displays all of the raw data records imported from the Water Quality Portal

Non Detects:
These are the records with values for the 'Result Detection Condition Text' field equal to 'Not Detected' or 'Present below Quantitation Limit'

W/O Units:
These data records have no data entered in either the 'Result Measure - Measure Unit Code' or the 'Quantitation Limit Measure - Measure Unit Code' fields.

W/O Methods:
There are 14 Activity Type Codes which do not require a sample to have a specified method. These data records do not match those 14 Activity Type Codes AND have no data entered in the 'Result Analytical Method - Method Identifier' field.

Duplicates:
These data records are duplicated within the imported data set. This means these records match all fields of another record in the data set except for the 'Activity Type' and 'Activity ID' fields

Filtered Data:
The Filtered Dataset includes only results with units and methods. Duplicate records have been removed. This is the data set passed to the map and table on the 'View Data' page.

Summary:
This table shows summary statistics of all unique combinations of station, media, characteristic, unit, and sample fraction.

Select method for Non-Detections. Users can use this menu to decide how non-detections are shown in the data set. By default, the tool is set to 'Ignore Non-Detections - remove from data set'. Users can also select: Set Non-Detections equal to zero; Set Non-Detections equal to the limit of detection or Set Non-Detections equal to the ½ times the Limit of Detection.

Please select a method to deal with Non-Detections

☒ Ignore Non-Detections - remove from data set

☐ Set Non-Detections equal to zero

☐ Set Non-Detections equal to the Limit of Detection

☐ Set Non-Detections equal to the 1/2 times the Limit of Detection

Click a Radio button to change how Non-Detections are handled in the dataset.

Available Data Sets. Preliminary QA/QC measures are applied to the data to generate the six different data sets and a summary. The data sets are presented in separate tabs.

On the dataset tabs Users can:

- **Search** data by typing in the Search Box
- **Column visibility.** Not all the available columns are displayed by default. Clicking on the column visibility button will bring up a menu of column names. Users can enable or disable columns by checking/unchecking the boxes.
- **Save Data** by clicking on the Save Data button.

WQP STORET Data Discovery Tool							
Query Data		Check Data		View Data		Help	
Home	All Data	Non Detects	W/O Units	W/O Methods	Duplicates	Filtered Data	Summary
All Imported Data Records							
This table displays all of the raw data records imported from the Water Quality Portal							
<div>Save Data</div>							
Show 50 entries		Search:		Show / hide columns			
Station	Name	ActivityStartDate	Characteristic	Result	ActivityMediaSubdivisionName	OrganizationFormalName	ActivityTypeCode
YAN_WQX-BCR06	Beaver Creek	2013-05-13	Conductivity	488.20		Yavapai-Apache Nation	Sample-Routine
YAN_WQX-BCR06	Beaver Creek	2013-08-27	Conductivity	561.00		Yavapai-Apache Nation	Sample-Routine
YAN_WQX-BCR06	Beaver Creek	2013-10-29	Conductivity	459.10		Yavapai-Apache Nation	Sample-Routine

Summary Tab. In addition to the six datasets, there is also a summary Tab. This tab provides summary statistics of all unique combinations of station, media, characteristic, unit and sample fraction.

WQP STORET Data Discovery Tool Query Data Check Data View Data Help

Home All Data Non Detects W/O Units W/O Methods Duplicates Filtered Data **Summary**

Click the button below to run a summary of the data

[Click here to summarize data](#) → Summarize Data

WQP STORET Data Discovery Tool Query Data Check Data View Data Help

Home All Data Non Detects W/O Units W/O Methods Duplicates Filtered Data **Summary**

Click the button below to run a summary of the data

Summarize Data

This table shows summary statistics of all unique combinations of station, media, characteristic, unit, and sample fraction.

Save Data

Show 50 entries Search:

Station	Name	ActivityMediaName	Characteristic	Unit	ResultSampleFractionText	Minimum	Maximum	Average	Count
YAN_WQX-BCR06	Beaver Creek	Water	pH	None	Total	8.100	8.300	8.2000000	7
YAN_WQX-BCR06	Beaver Creek	Water	Inorganic nitrogen (nitrate and nitrite)						6

VIEW DATA TAB

The view data tab shows only the filtered data from the Check Data Tab. The left side of the screen contains Data Filters, and the right side has separate tabs for Map, Table, Station Summary and Parameter/Unit Summary.

WQP STORET Data Discovery Tool Query Data Check Data **View Data** Help

Map **Table** Station Summary Parameter/Unit Summary

Data filters

Submit!

Filter by Organization

Filter by Station

Filter by Sample Media

Filter by Sample Fraction

Filter by Parameter

Filter by Units

Filter by Methods

Filter by Result Qualifier

Select value range:

Minimum: 0 Maximum: 1160

Select a Date Range:

2015-01-06 to 2016-12-30

Map Displays Filtered Data

* To populate the tabs 'Station Summary' and 'Parameter/Unit Summary', please select a station using the panel located on the map.

Data filters

Submit!

Filter by Organization

Filter by Station

Filter by Sample Media

Filter by Sample Fraction

Filter by Parameter

☒ Select All
☐ Deselect All

Select parameter:

pH Dissolved oxygen (DO) Temperature, water
Escherichia coli
Inorganic nitrogen (nitrate and nitrite) Nitrogen
Conductivity
Suspended Sediment Concentration (SSC)
Phosphorus Turbidity Kjeldahl nitrogen

Data Filters. This panel provides users with several options for refining their data set. These filters drive the data available for display in the Interactive Map and the Table tab.

To apply a data filter, click on the panel to expand it. In each panel filter you can choose to Select all or Deselect All items. Items can also be removed from the selection one at a time by clicking on it and then hitting the delete key. The removed item will move to drop-down box below the **Select [filter name]:** box. It can be added back by clicking on it.

Note: A filter will only be applied if an item is selected. If you chose Deselect all, that filter will not be applied to the query.

Once all filters are set, click on the **Submit!** Button to update the Map and Table tabs.

Data filters

Submit!

Filter by Organization

Filter by Station

Filter by Sample Media

Filter by Sample Fraction

Filter by Parameter

Filter by Units

Filter by Methods

Filter by Result Qualifier

Select value range:

Minimum: 0.01 Maximum: 1920

Select a Date Range:

2012-11-08 to 2014-07-30

Select value range: Users can refine the value range for the data set by typing in a minimum and maximum value. By default, the tool shows the minimum and maximum range of the imported filtered dataset (the range does not change when selecting filters). **Note:** **Users must enter both a minimum and maximum value for this filter to be applied. The tool will fail if just one value (either min or max) is used.**

Select a Data Range: Users can refine the date range for the data set. Clicking on a date will bring up a calendar that can be used to select a date. By default, the tool shows the date range for the filtered dataset.

Interactive Map. The map shows circles for all the stations currently displayed in the **Interactive Data Table**. The size of the circle marking each station is determined by the number of results available. The circles will resize as filters are applied. Clicking on any station marker will reveal a popup displaying the select this Location button, latitude, longitude, station ID, station name, number of samples, and number of results. The **Station Summary** tab can be updated by clicking on the select this Location button in the station marker popup.

WQP STORET Data Discovery Tool Query Data Check Data View Data Help

Map Table Station Summary Parameter/Unit Summary

Data filters

Submit!

Filter by Organization

Filter by Station

Filter by Sample Media

Filter by Sample Fraction

Filter by Parameter

Filter by Units

Filter by Methods

Filter by Result Qualifier

Select value range:

Minimum: 0 Maximum: 1160

Select a Date Range: 2015-01-06 to 2016-12-30

Map Displays Filtered Data

Latitude 35.844708, Longitude -78.711774
Station ID: 21NC03WQ-JF24, Station Name: Crabtree Cr, 20 sample(s), 20 results

Select this Location

* To populate the tabs 'Station Summary' and 'Parameter/Unit Summary', please select a station using the panel located on the map.

Table Tab. This tab provides users with a tabular view of the filtered data shown on the map. On this tab users can:

- **Sort** data by clicking on the column heading.
- **Search** data by typing in the Search Box.
- **Save Data** by clicking on the Save Data button.

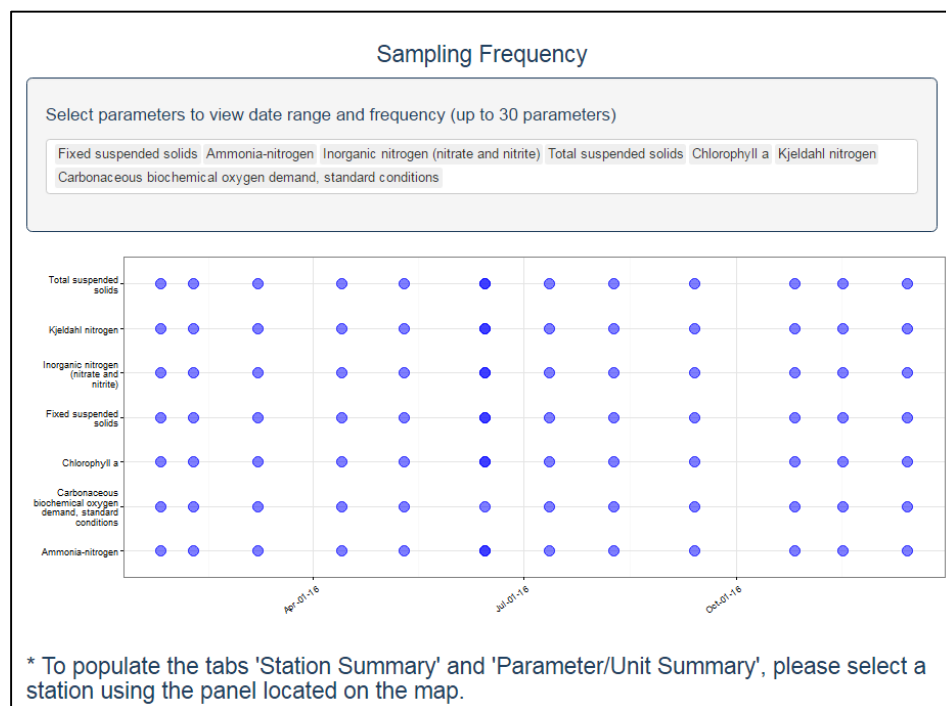
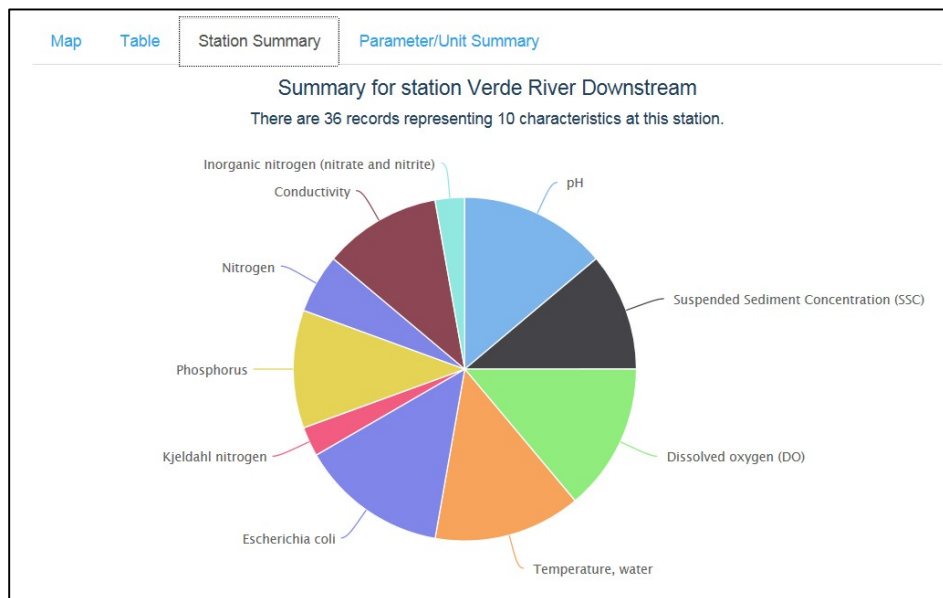
Map Table Station Summary Parameter/Unit Summary

Save Data

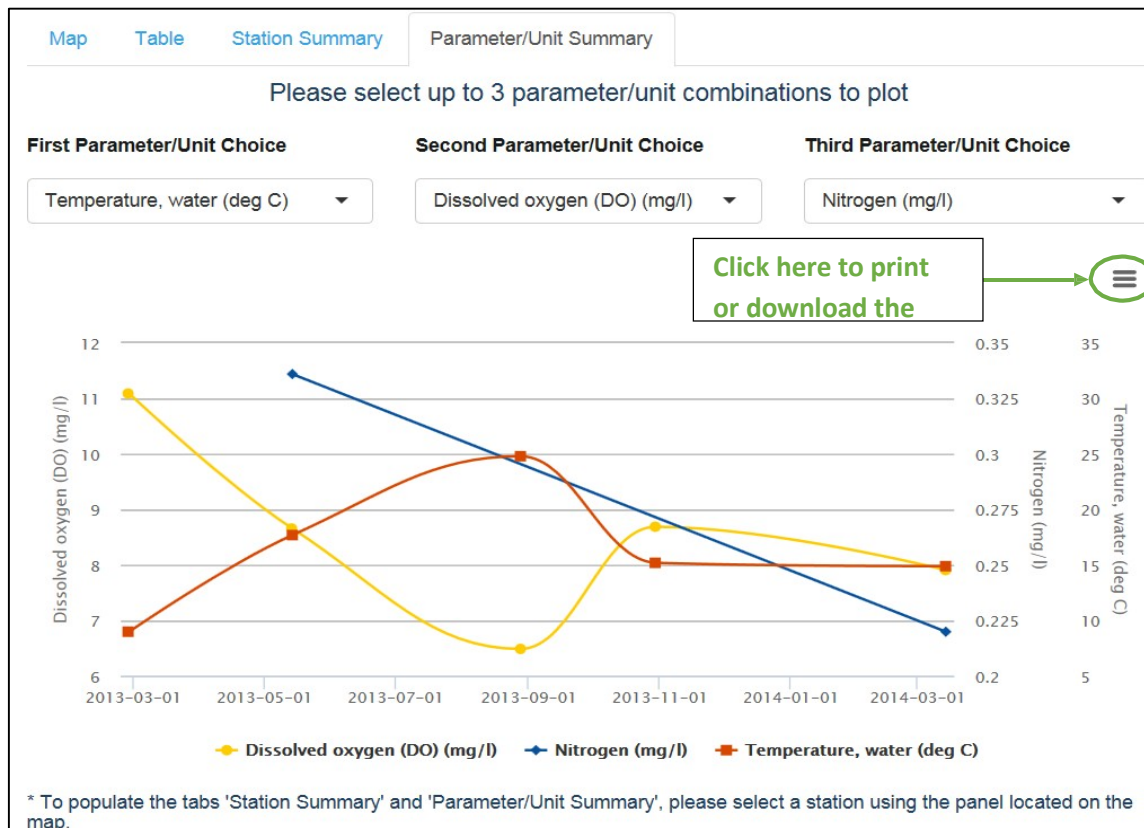
Show 100 entries Search:

Station	Name	Organization	Characteristic	Result	Unit	Method	ActivityStartDate
21NC03WQ-LC01	FALLS OF THE NEUSE RESERVOIR	21NC03WQ	Carbonaceous biochemical oxygen demand, standard conditions	2.3	mg/l	Unknown	2016-04-13
21NC03WQ-LC01	FALLS OF THE NEUSE RESERVOIR	21NC03WQ	Volatile suspended solids	0	mg/l	Fixed and Volatile Solids in Water	2016-04-13
21NC03WQ-LC01	FALLS OF THE NEUSE RESERVOIR	21NC03WQ	Total solids	73	mg/l	Total Solids Dried 103-105C in Water	2016-04-13
21NC03WQ-LC01	FALLS OF THE NEUSE RESERVOIR	21NC03WQ	Organic carbon	8	mg/l	Total Organic Carbon by Combustion-Infrared Method	2016-04-13

Station Summary Tab. The top pie chart shows the breakdown of the total results for the station by characteristic. The **Sampling Frequency** chart shows the timing of sample collection and associated results for the station. Note that parameters can be added to and removed from the sampling frequency chart by clicking inside the **Select parameters to view data range and frequency** box. Note: This tab will only display after selecting a station on the Map and clicking the **Select this Location** button.

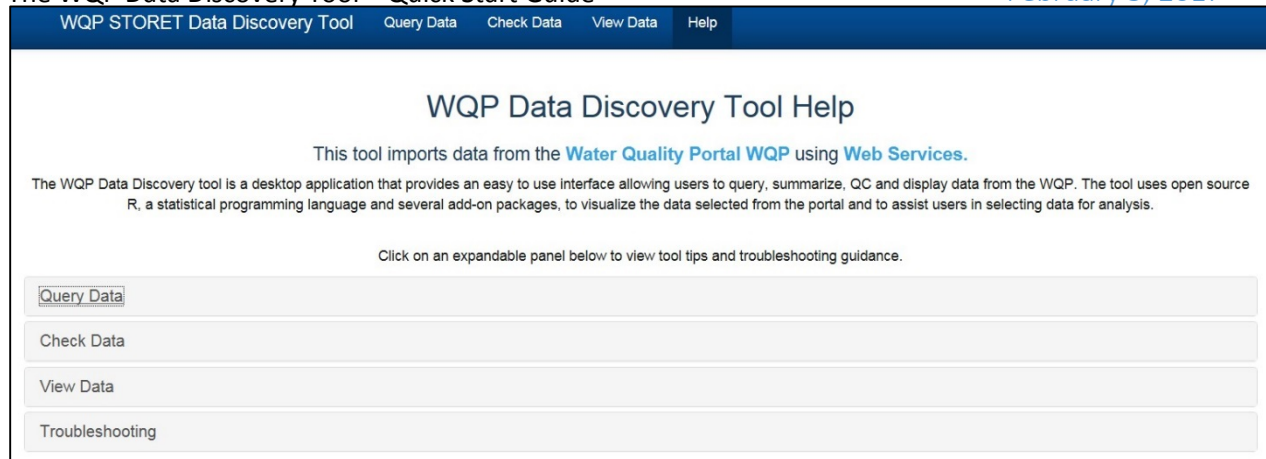


Parameter/Unit Summary Tab. This tab will allow you to select up to three parameters for the selected station to graph over time. User can zoom into a data point on the graph by clicking on the chart and drawing a box around the area that want to zoom into. A **reset zoom** button will appear and users can click this button to return to the original view. Users can also print the graphic or export it as an image file. Note: This tab will only display after selecting a station on the Map and clicking the **Select this Location** button.



HELP TAB

The Help Tab provides users with information on how to use the Data Discovery tool. Click on a panel to expand it and view the information.



ENDING YOUR SESSION IF YOU ARE NOT USING THE R-PORTABLE PACKAGE

When you're finished with your session, navigate to the R terminal that you used to launch the tool and Hit the Escape key. You may now close R and the tab in your internet browser where the tool was running. If R asks you, you do not need to save any information about the R session.

ADDITIONAL INFORMATION

R Software and Supplementary Packages. R is a free statistical software package and many people contribute to it in the form of "packages", much like people contribute "apps" to mobile phones. These packages simply extend the capabilities of R beyond the basic installation. These are all free as well. They undergo a vetting process before they are made available to the public, so you can be confident that they're safe. The **WQP Data Discovery Tool**, utilizes packages that are popular and well known to be safe.

Installing R packages if you don't have admin rights. If you do not have administrative rights on your machine, then you will need to install the packages in a local directory. R will prompt you with a dialog box that says "Would you like to use a personal library instead?", click Yes. R will then identify a location on your computer to store the files. You will receive a dialogue box that says "Would you like to create a personal library 'C:\Users\JDoe\Documents\R\Win-library\3.2' to install packages into?", click Yes.

RStudio. If you are an R Studio user, the app can be launched via RSTUDIO. An R GUI that can be downloaded at <http://www.rstudio.com/products/rstudio/download/>. Once the app is installed click "File" >> "Open File ..." and navigate to where the app is and open the "Data Visualization Tool" folder. You will see two files named 'server.R' and 'ui.R'. You can open either of them. After opening an arrow with the text "Run App" beside it will appear on the top right of the window displaying the file. Clicking this arrow launches the app. Ignore any warnings or messages in the R terminal. If the last line of text that appears in the R terminal is "Listening on <http://127.0.0.1:7088>", you are ready to use the app. When you open the App in RStudio, it will default to opening in the RStudio Viewer. While you can use the app in the RStudio Viewer, you will be unable to save the data to your computer. If you open the App using RStudio, the first thing you need to do is select "Open in Browser" in the upper left corner.