

Tools for Automated Data Analysis (TADA) Module 1: Water Quality Portal Data Discovery and Cleaning

- 1. Load
- 2. Overview
- 3. Flag
- 4. Filter
- 5. Censored Data
- 6. Harmonize and Calculate
- 7. Explore
- 8. Review

1. Synonym Harmonization

Use this section to harmonize characteristic-fraction-speciation synonyms. Click 'Compose Synonym Table' and the table will appear below. The table shows the characteristic-fraction-speciation combinations in your dataset (original columns highlighted blue), as well as any changes that will be made to TADA metadata to allow synonyms to be grouped appropriately (denoted by 'Target' and 'Conversion' columns). Many of these harmonization decisions have been made and documented by the TADA Team in the 'Assumptions' columns. Click the 'CSV' button at the top left corner of the table to download the synonym reference table for your dataset. You may edit manually and re-upload in the file upload widget next to the blue button. When you are ready to harmonize your dataset to the synonym table target elements, click 'Harmonize Data with Synonym Table'. This button only appears when a synonym table has been generated/loaded into this tab.

Compose Synonym Table

Alternative option: After running the 'Compose Synonym Table' button (above), download the TADAShiny-generated synonym table using the csv button at the top of the table, customize the table to meet your harmonization needs, and upload your csv here before clicking 'Harmonize Data with Synonym Table'.

Upload Custom Table (.csv only)

Browse... No file selected



Compose Synonym Table

Download Synonym Table (.csv)

Show 5 entries

Search:

Characteristic	Target Characteristic	Characteristic Assumptions	Fraction	Target Fraction	Fraction Assumptions	Speciation	Target Speciation	Speciation Assumptions	Speciation Conversion Factor (to AS N or AS P)	Harmonization Group
All	All	All	All	All	All	All	All	All	All	All
AMMONIA-NITROGEN	AMMONIA	Assumes AMMONIA-NITROGEN is synonymous to AMMONIA.	DISSOLVED	FILTERED	Assumes DISSOLVED and TOTAL SOLUBLE are equivalent to FILTERED. FILTERED is precise and preferred.		AS N			Nitrogen
AMMONIA-NITROGEN	AMMONIA	Assumes AMMONIA-NITROGEN is synonymous to AMMONIA.	TOTAL	UNFILTERED	Assumes AMMONIA-NITROGEN, TOTAL is synonymous to AMMONIA, UNFILTERED		AS N	Assumes AS N		Nitrogen
TEMPERATURE, WATER	TEMPERATURE									TEMPERATURE
KJELDAHL NITROGEN	TOTAL KJELDAHL NITROGEN (ORGANIC N & NH3)	Assumes KJELDAHL NITROGEN, TOTAL is synonymous to TOTAL KJELDAHL NITROGEN (ORGANIC N & NH3), UNFILTERED.	TOTAL	UNFILTERED	Assumes KJELDAHL NITROGEN, TOTAL is synonymous to TOTAL KJELDAHL NITROGEN (ORGANIC N & NH3), UNFILTERED.		AS N	Assumes AS N		Nitrogen
		Assumes NITROGEN is			Assumes NITROGEN,					

Alternative option: After running the 'Compose Synonym Table' button (above), download the TADAShiny-generated synonym table using the csv button at the top of the table, customize the table to meet your harmonization needs, and upload your csv here before clicking 'Harmonize Data with Synonym Table'.

Upload Custom Table (.csv only)

Browse... No file selected

2. Total Nitrogen and Phosphorus Summation

Data generators commonly monitor for several nutrient subspecies that, when added together, can be used to estimate a total nitrogen or phosphorus value. TADA uses the logic provided in ECHO's Nurient Aggregation page (see: <https://echo.epa.gov/trends/loading-tool/resources/nutrient-aggregation>) to rank and sum subspecies for a given day, location, depth, activity media subdivision, and unit. Total Nitrogen and Total Phosphorus values are added as new results in the dataset. Users may view the nutrient aggregation reference sheet by clicking 'See Summation Reference'. Once data are harmonized, the user may then summarize total N and P. **NOTE:** When two or more measurements of the same substance occur on the same day at the same location, the function uses the maximum of the group of values to calculate a total nutrient value.

See Summation Reference (.csv)

TADA Working Summary

Total Results in Dataset: 134,050

Total Results Flagged for Removal: 2,575

Total Results Retained: 131,475



Tools for Automated Data Analysis (TADA) Module 1: Water Quality Portal Data Discovery and Cleaning

- 1. Load
- 2. Overview
- 3. Flag
- 4. Filter
- 5. Censored Data
- 6. Harmonize and Calculate
- 7. Explore
- 8. Review

1. Determine Characteristic Groups of Interest

First, pick the columns you'd like to use to group your data. You can stick with 'TADA.ComparableDataIdentifier' (where each group of data represents one unique characteristic-unit-fraction-speciation combination), or you can add other columns that will help you divide and review your data most effectively. For example, you may also want to divide up your data based on monitoring location type, or hydrologic condition. Once you've decided on your grouping columns, click 'Generate Groups' below to group data. It is **optional** to add more grouping columns.

Select grouping columns


 Generate Groups



1. Determine Characteristic Groups of Interest

First, pick the columns you'd like to use to group your data. You can stick with 'TADA.ComparableDataIdentifier' (where each group of data represents one unique characteristic-unit-fraction-speciation combination), or you can add other columns that will help you divide and review your data most effectively. For example, you may also want to divide up your data based on monitoring location type, or hydrologic condition. Once you've decided on your grouping columns, click 'Generate Groups' below to group data. It is **optional** to add more grouping columns.

Select grouping columns

 Generate Groups

2. Pick Groups to Map and Plot

Use the drop down to pick the characteristic group(s) (max of 2) you'd like to map/plot and then click 'Generate Map'. If you have **one** group selected, the map will display site markers with radii that reflect the mean result measure value at each site: larger site markers correspond to higher mean result measure values. If you have **two** results selected, the map will display which sites have data for one or both of the selected groups.

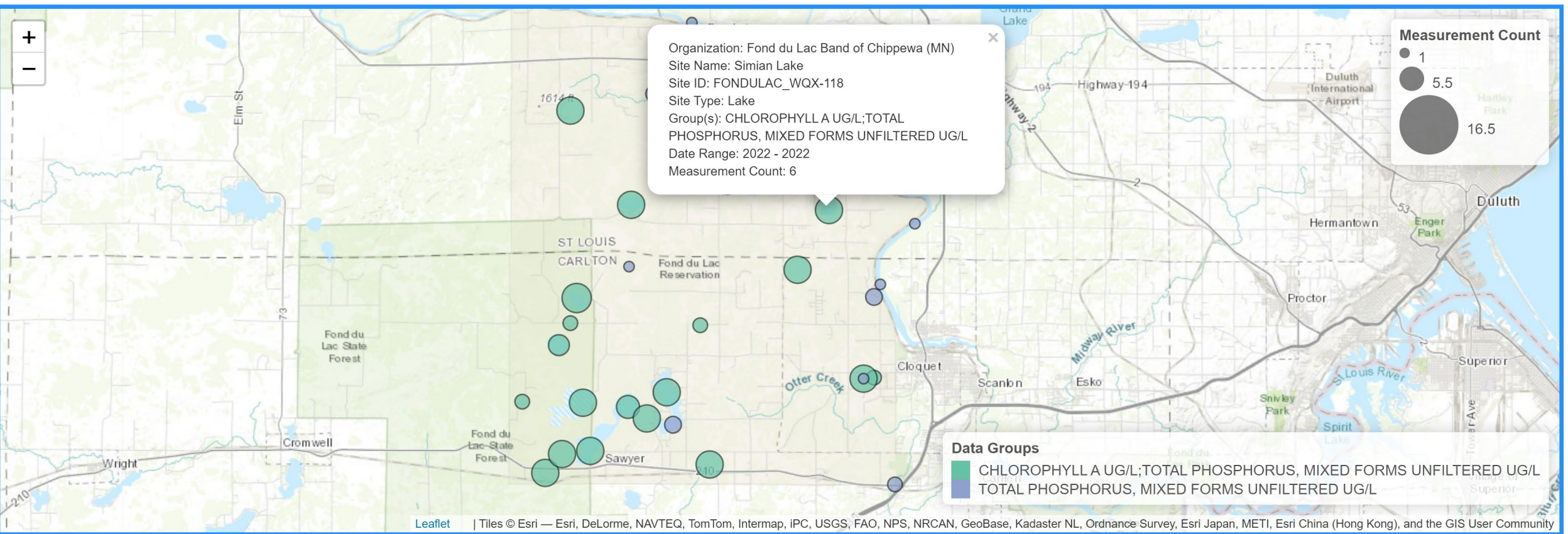
Select up to TWO groups

 Generate Map



Select up to TWO groups

TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L | CHLOROPHYLL A UG/L | [Generate Map](#)



3. Select Specific Sites (Optional)

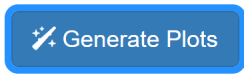
Use the drop down to pick the sites you'd like to include in the plots below and then click 'Generate Plots'. Defaults to all sites in the dataset. **NOTE:** Currently, the single-characteristic scatterplot, histogram, and boxplot show the first characteristic from the drop down above the map: **TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L.**

Select sites

3. Select Specific Sites (Optional)

Use the drop down to pick the sites you'd like to include in the plots below and then click 'Generate Plots'. Defaults to all sites in the dataset. **NOTE:** Currently, the single-characteristic scatterplot, histogram, and boxplot show the first characteristic from the drop down above the map: **TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L.**

Select sites

 Generate Plots

Single Characteristic Figures

Two-Characteristic Scatterplot

Benchmark values appear as horizontal lines on the single-characteristic scatterplot figure, below. Benchmarks may reflect established or proposed water quality criteria, important physical/chemical/biological thresholds, or any other values relevant to your use case.

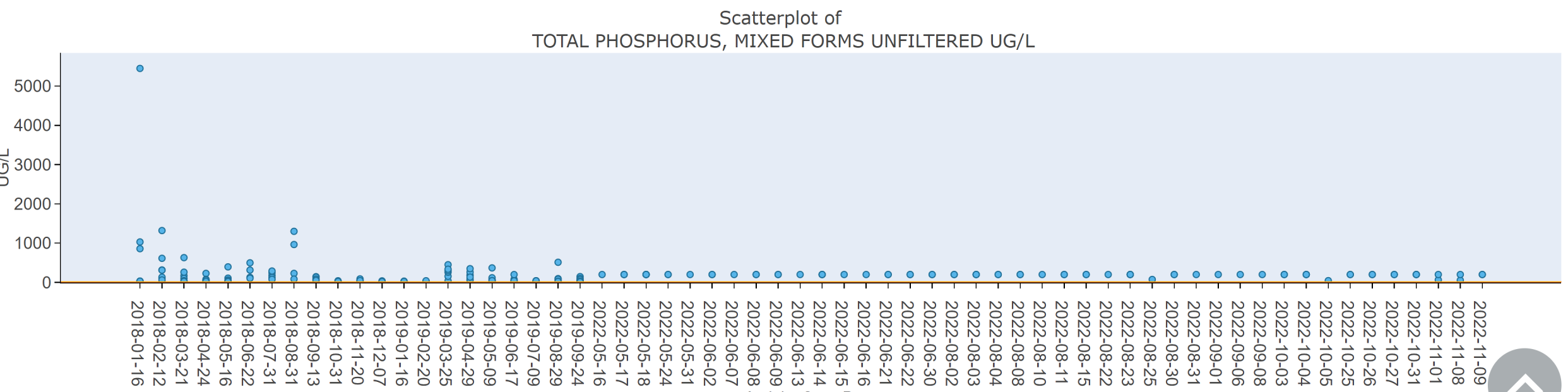


Single Characteristic Figures | **Two-Characteristic Scatterplot**

Benchmark values appear as horizontal lines on the single-characteristic scatterplot figure, below. Benchmarks may reflect established or proposed water quality criteria, important physical/chemical/biological thresholds, or any other values relevant to your use case.

Benchmark 1 **Benchmark 2**

204 out of 204 measurements (100%) exceed benchmark 1, while 204 out of 204 measurements (100%) exceed benchmark 2.

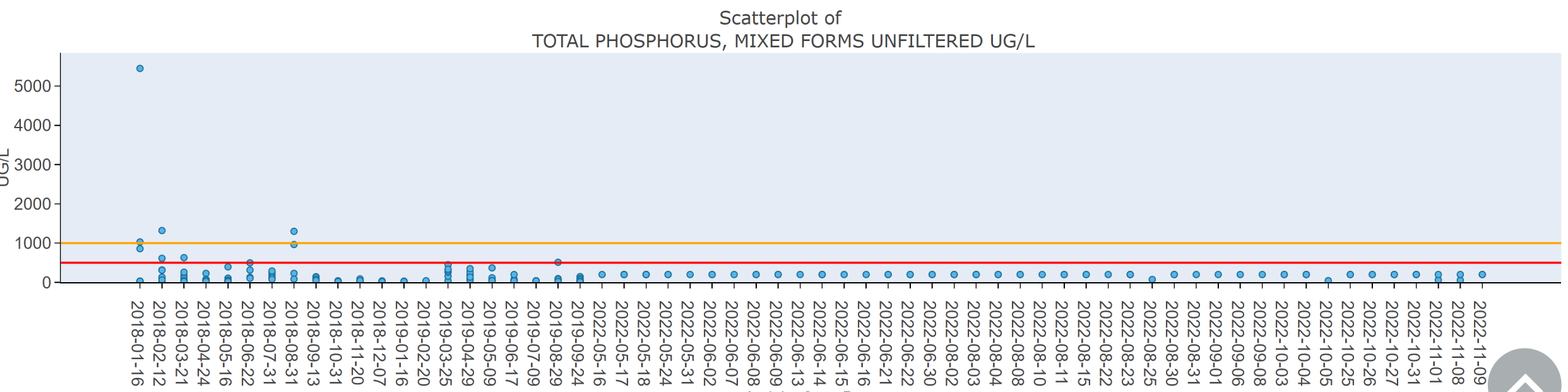


Single Characteristic Figures | **Two-Characteristic Scatterplot**

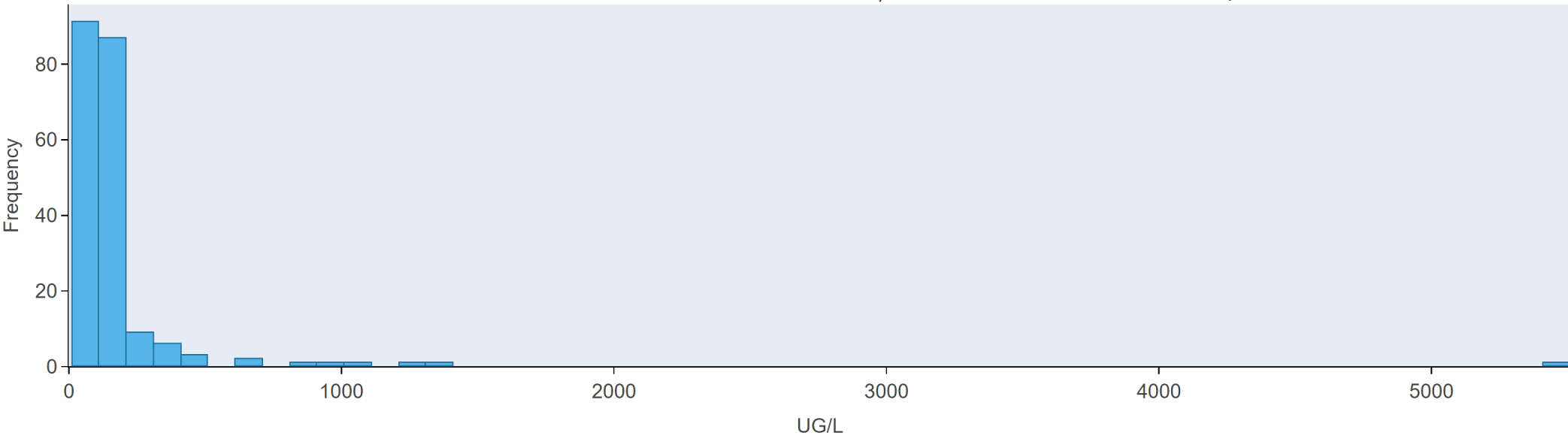
Benchmark values appear as horizontal lines on the single-characteristic scatterplot figure, below. Benchmarks may reflect established or proposed water quality criteria, important physical/chemical/biological thresholds, or any other values relevant to your use case.

Benchmark 1 **Benchmark 2**

9 out of 204 measurements (4.4%) exceed benchmark 1, while 4 out of 204 measurements (2%) exceed benchmark 2.



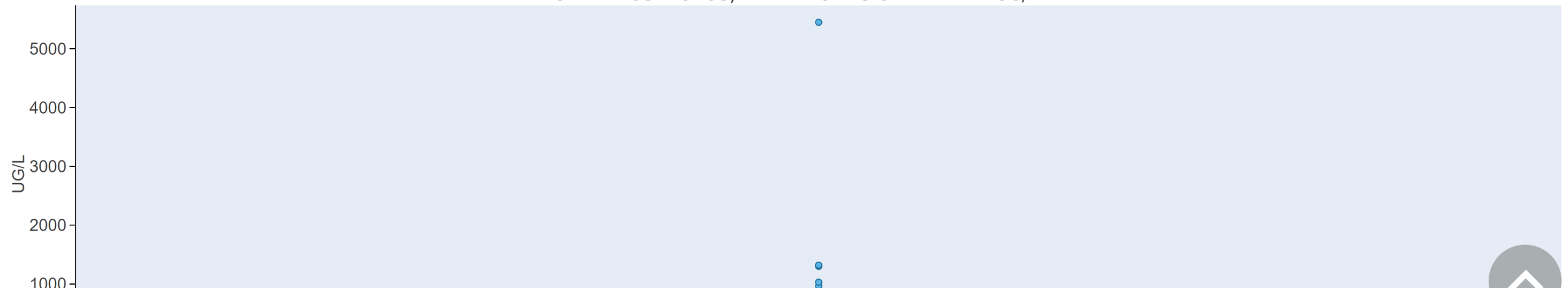
Histogram of TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L



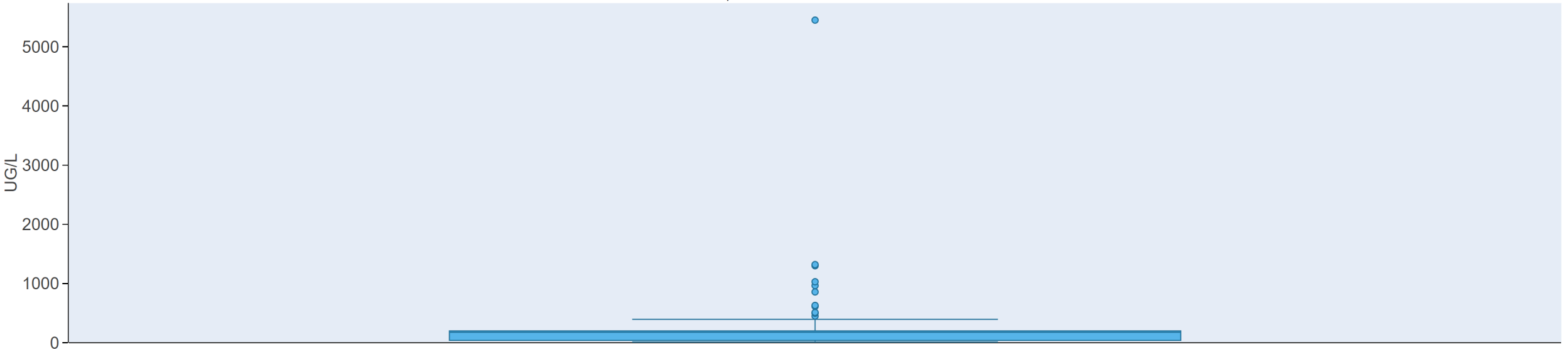
Select 'Outliers Removed' and Deselect 'All Data' to View a Subset of the Data

- All Data
- Outliers Removed
 - Upper Threshold: 395
 - Lower Threshold: 12

Boxplot of TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L



Boxplot of TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L



TADA Working Summary

Total Results in Dataset: 134,050
Total Results Flagged for Removal: 2,575
Total Results Retained: 131,475

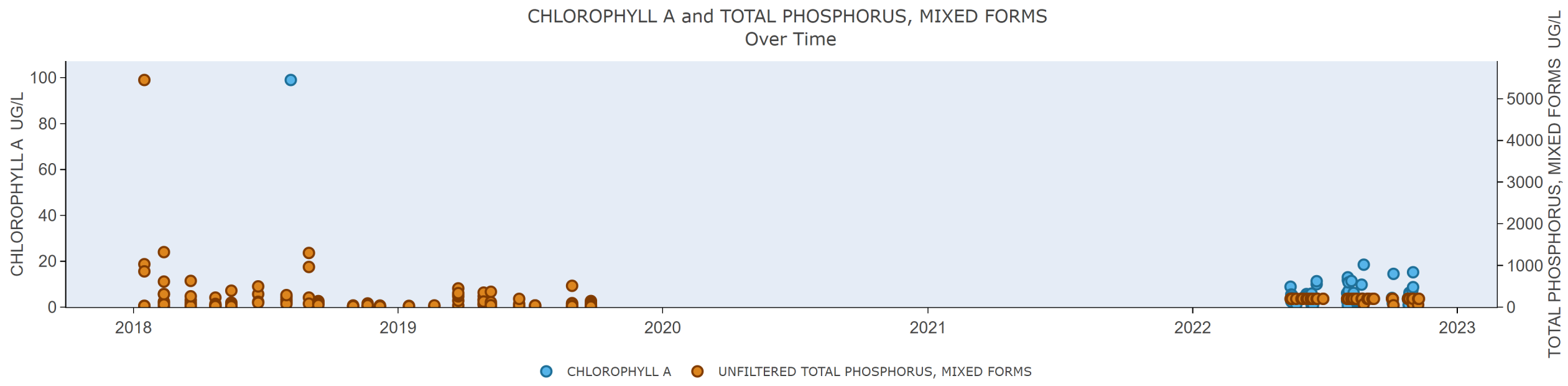
Total Sites in Dataset: 227
Total Sites Flagged for Removal: 0



3. Select Specific Sites (Optional)

Use the drop down to pick the sites you'd like to include in the plots below and then click 'Generate Plots'. Defaults to all sites in the dataset. **NOTE:** Currently, the single-characteristic scatterplot, histogram, and boxplot show the first characteristic from the drop down above the map: **TOTAL PHOSPHORUS, MIXED FORMS UNFILTERED UG/L.**

Select sites



EPA development environment: The content on this page is not production ready. This site is being used for **development** and/or **testing** purposes only.

 An official website of the United States government [Here's how you know](#)



- [Environmental Topics](#)
- [Laws & Regulations](#)
- [Report a Violation](#)
- [About EPA](#)

Tools for Automated Data Analysis (TADA) Module 1: Water Quality Portal Data Discovery and Cleaning

- 1. Load
- 2. Overview
- 3. Flag
- 4. Filter
- 5. Censored Data
- 6. Harmonize and Calculate
- 7. Explore
- 8. Review

Removal Review

Use this tab to review flagging and filter decisions and explore the filtered dataset. Click the button below to begin.

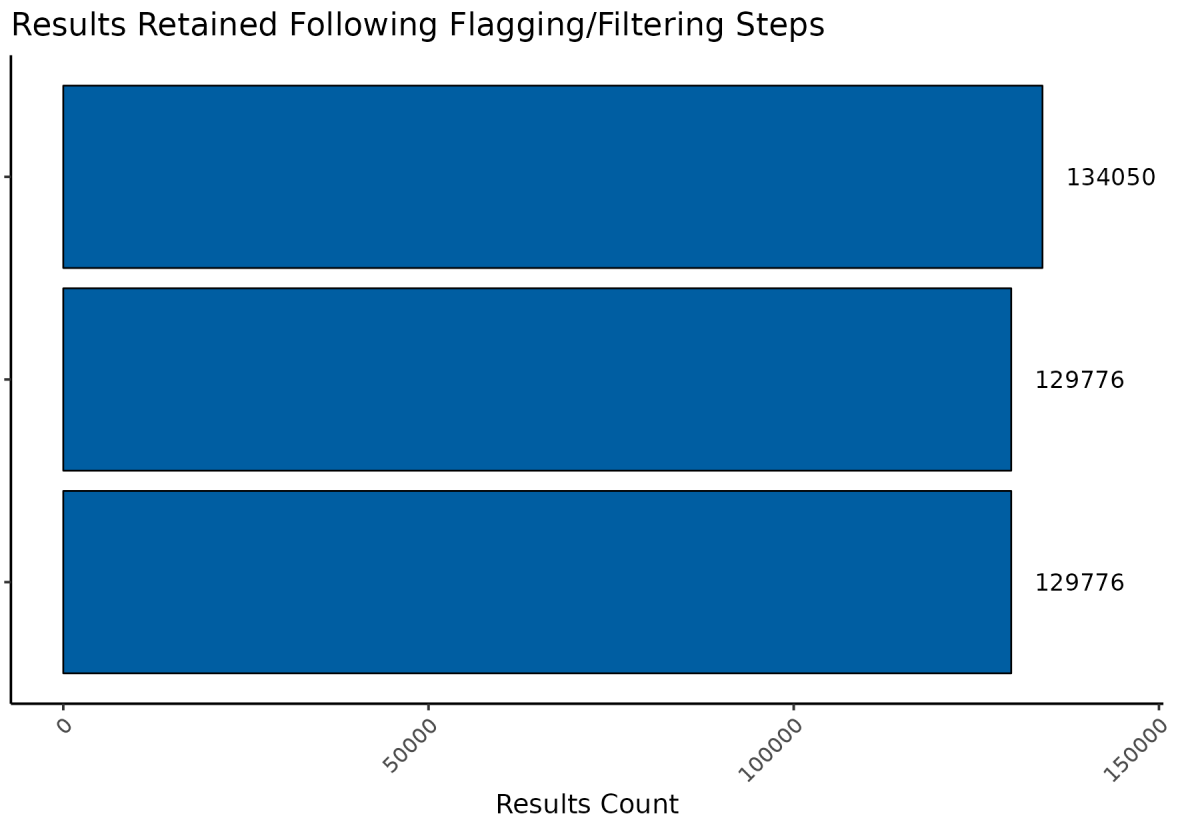
[Load Review Data](#)



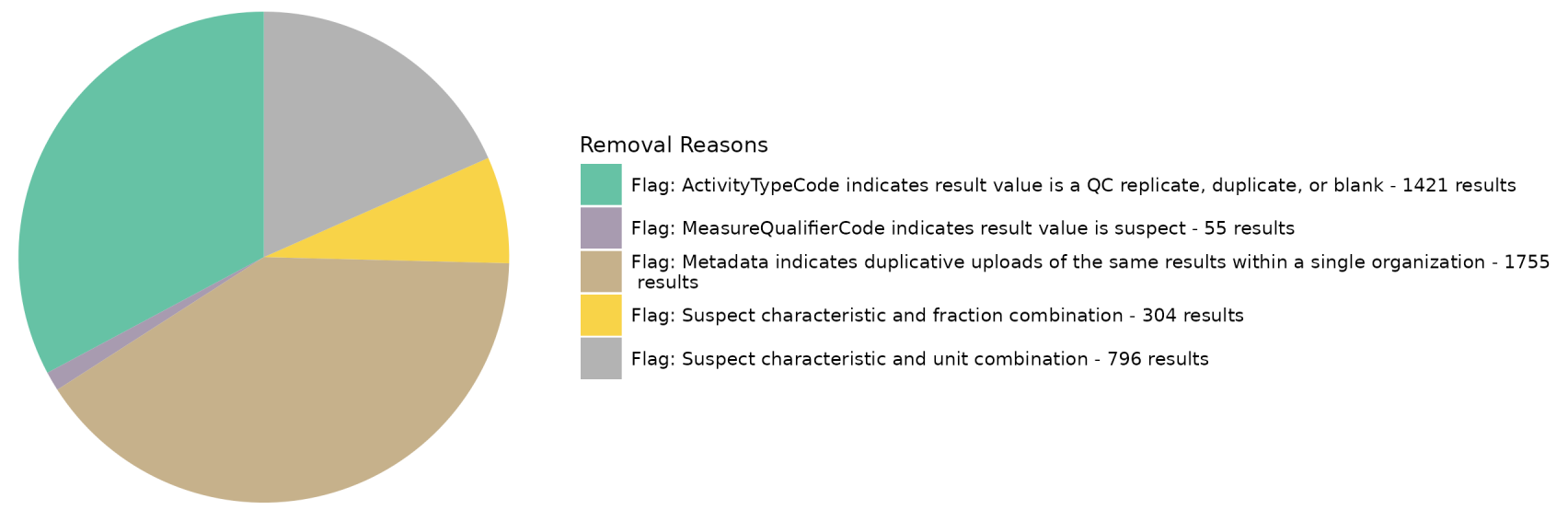
Removal Review

Use this tab to review flagging and filter decisions and explore the filtered dataset. Click the button below to begin.

[Load Review Data](#)



50000
100000
150000
Results Count



Note: This pie chart shows the number of results flagged/filtered for each reason. Some results may be removed for multiple reasons. Because of this, the total number of flagged results in this pie chart is equal to or greater than the number of unique results removed.

TADA Working Summary

Total Results in Dataset: 134,050

Total Results Flagged for Removal: 4,274

Total Results Retained: 129,776

