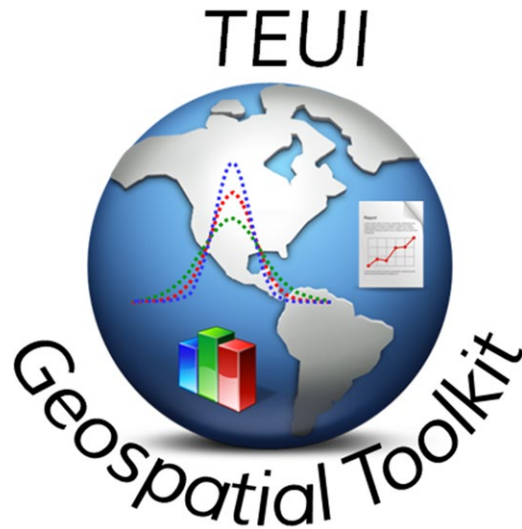# Data Management and Statistics

Welcome to the **Data Management and Statistics** training module for the TEUI Geospatial Toolkit! In this module you will learn the basics of setting up a Toolkit project, selecting your analysis data, running statistics, and checking the map unit symbology.

You will learn how to:

- Create a project
- Add and remove geospatial data
- Select a map unit (if desired)
- Run statistics
- Check map unit symbology

Required data:

- **(c:\...\TEUI_Training\Analysis_Features)**
  - **soilmu_a_wa680.shp** (polygons)
  - **soilmu_a_wa680_raster.img** (raster)
  - **wa680_plots.shp** (points)
- **(c:\...\TEUI_Training\Rasters)**
  - **Aspect.img** (raster)
  - **DEM.img** (raster)
  - **Slope_pct.img** (raster)
  - **Hillshade.img** (raster)

Contents

Part 1: Create a New Project

The Toolkit requires the user to specify a folder location to store the database that contains the statistics and the file location of the data used to create them. This database can use an existing project folder, but it is recommended that a new folder be created to reduce the number of extraneous files in the folder. For this tutorial, your project folder will be the **TEUI_Training** folder that you should have already uncompressed.

A. Create a new toolkit project

1. On the TEUI Toolbar, select the **Folder i**con (outlined in red below).

2. Select **Browse** in the **Select Project** dialog box (see below).
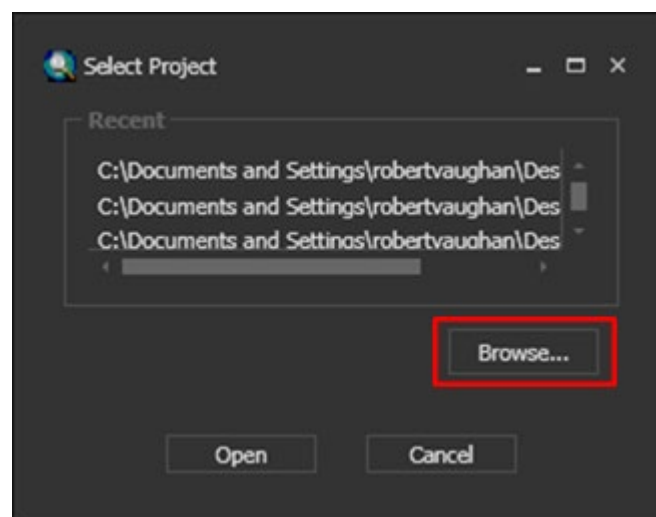
3. In the **Browse for Folder** dialog, navigate to the **TEUI_Training** folder.

4. Click **OK.**

Part 2: Manage Geospatial Data

Managing project data with the toolkit is very simple. The user simply points the program to the file that is to be used in the analysis.

> **Note:** *If you move the geospatial data used in an existing Toolkit project, you will see a yellow triangle next to the layer name in the TEUI Data dialog (see below). Click the triangle to open a file browse dialog and navigate to the layer's new location.*
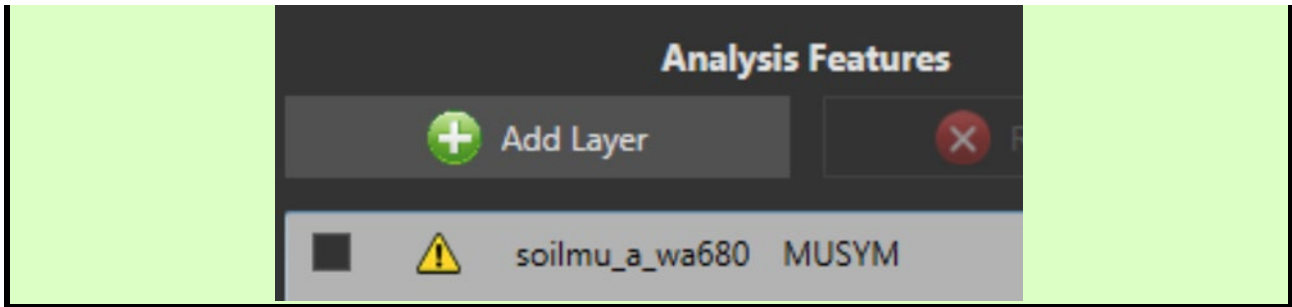
A. Add analysis features to a toolkit project

1. On the TEUI Toolbar, select the **Manage Data** icon to open its window.
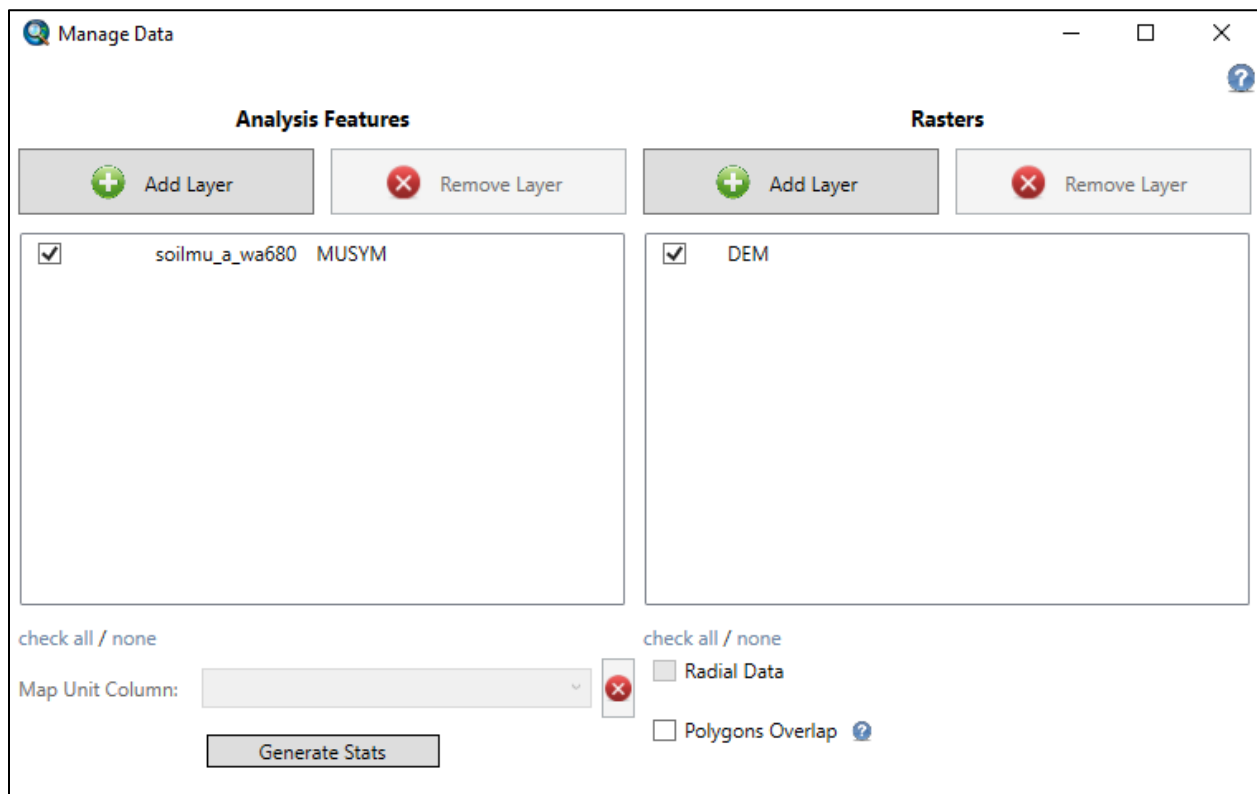
Analysis Features are zones within which you would like to generate statistics. These features can be feature classes such as polygons, points, or line features stored as a shapefile (.shp); or feature classes within a file geodatabase, or within a feature dataset within a file geodatabase. Analysis features can also be in the form of a discrete raster. In the case of the raster, the value field will be used as the identifier.

2. In the **Analysis Features** field, click on the **Add Layer** button (see below).
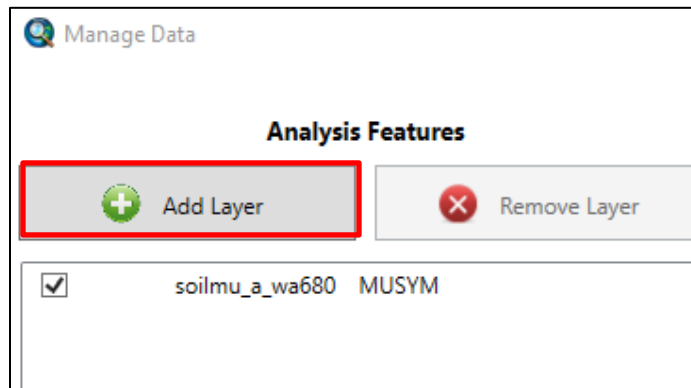
3. A data navigation window will appear. Navigate to the Analysis Features folder within the training data folder (**c:\...\TEUI_Training\Analysis_Features**).

4. Hold down the shift key and select all three layers, then select **Add**. You can add as many analysis zone data layers as you wish (see below).
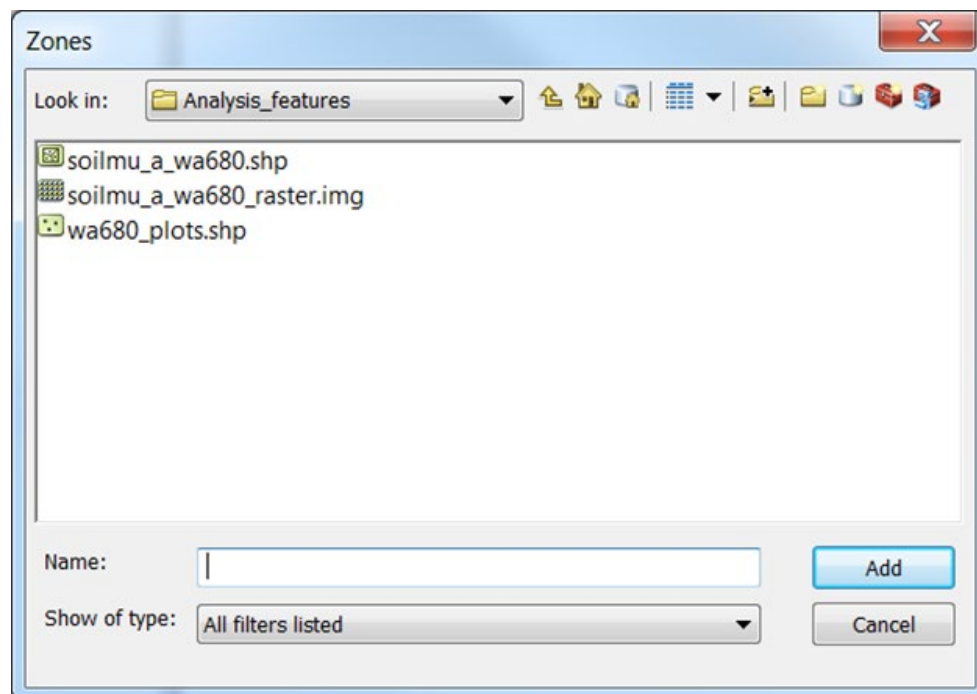
> **Note**: *All data layers to be analyzed (analysis features and raster data) must have the same geographic coordinate system and projection. You will receive a warning if they are different.*

B. Remove analysis features from a toolkit project

1. To remove a layer from the **Analysis Features** field, click on the layer you wish to remove in the grey layer list, and click **Remove Layer**.
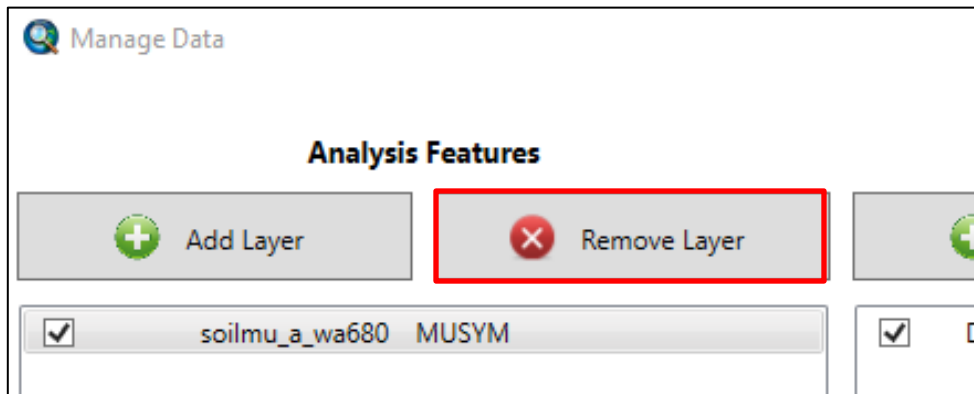

2. Click **Yes** when prompted to confirm that this is correct.

C. Choose a map unit

In some instances, such as TEUI mapping or soil survey, you may wish to identify individual polygon features as belonging to a specific **map unit**. This is simply a repeating identifying symbol or value which will be used to aggregate the statistics of the polygons belonging to that group. You **do not** need to select a map unit column in order to calculate statistics. The symbol must be present either as an attribute column in the feature class or as an attribute in the VAT table of a discrete raster.

1. Select the **soilmu_a_wa680** layer by clicking on it in the grey layer list, and click the **Map Unit Column** drop down menu.

2. Select the **MUSYM** (map unit symbol) attribute name for your map unit symbol in the drop down menu. This selection applies only to the layer selected.
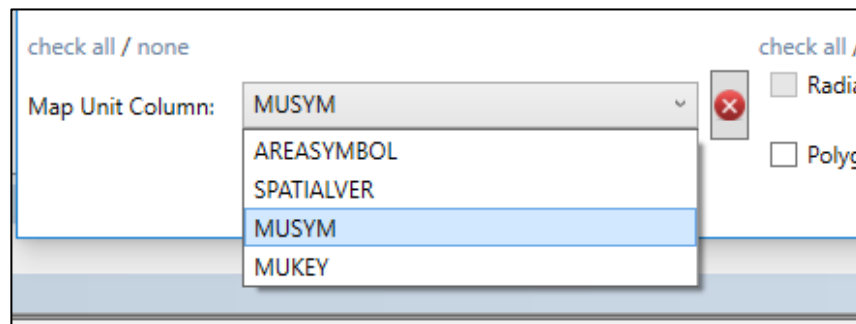

3. Select **wa680_plots** layer and then set its **Map Unit Column** to **Id**.

Hint: To **remove an attribute name** *from the map unit symbol selector, simply select the layer you wish to change, and hit the* **red circle X** *next to the Map Unit Column dropdown menu.*

D. Add raster data to the toolkit project

In the TEUI Data dialog window, the **Rasters** field is for raster data that is to be used to generate the descriptive statistics. An example would be a digital elevation model (DEM), percent slope, aspect, or land use. Raster data can be either continuous or discrete and can have multiple bands, as with Landsat data.

1. To **add raster data** to your Toolkit project, select the **Add Layer** button under the Rasters heading.
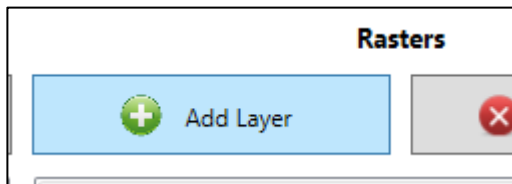
2. In the Rasters navigation window, navigate in the training data folder to (**c:\...\TEUI_Training\Rasters**).

3. While holding down the **Ctrl** key, select all the data layers and click **Add**.
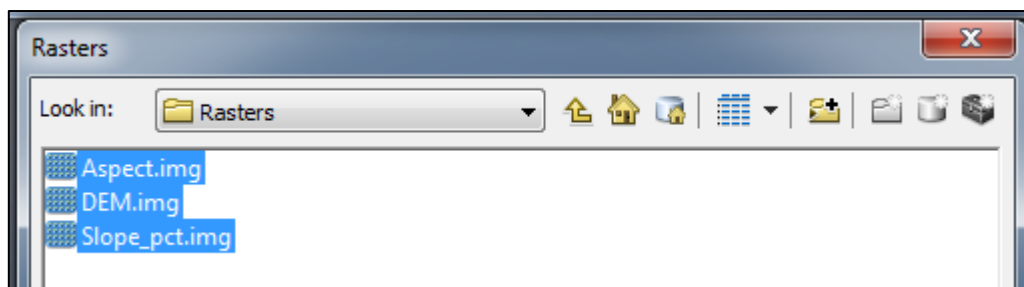

Part 3: Calculate Statistics

1. To calculate statistics, you must first select the analysis feature layers and raster data layers you wish to run statistics on by checking the box next to each layer. You must select at least one Analysis Feature layer and at least one Raster layer.

   i. For this exercise, select the **soilmu_a_wa680** and **wa680_plots** Analysis Feature layers and the **DEM**, **Slope_pct**, and **Aspect** Raster layers. You can check or uncheck every layer at once by clicking the **check all or none** buttons below the layer list.

*If your polygons overlap then make sure to click the **Polygons Overlap** box at the bottom of the window. This method will significantly increase processing time but will ensure accurate statistics for overlapping polygons.*

If a layer is radial data (such as aspect) click that layer to highlight it then click the Radial Data box at the bottom of the window.
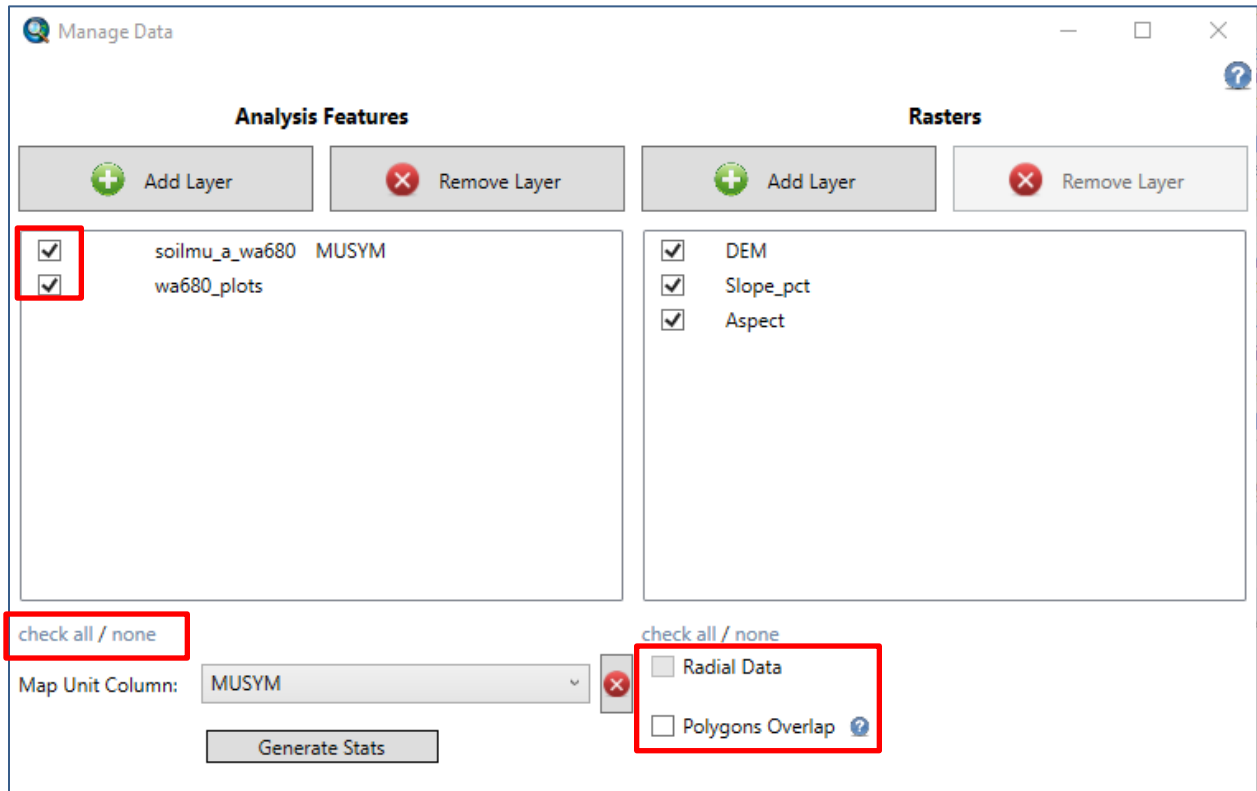
2. Click on the **Generate Stats** button. Select **Yes** when prompted by the time warning that appears. This process may take a significant amount of time depending on the amount of data selected.

3. A dialog will appear to inform you of the duration, the features and rasters being used, and the progress (see below). The tool can process more than 1 million cells a second (see next page)!
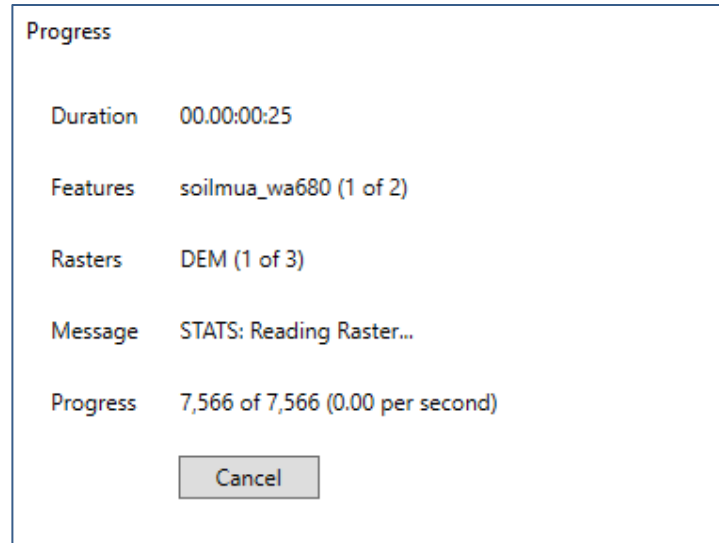
4. Click **OK** when the statistics run is done.

Part 4: Query Statistics

You can quickly query zonal statistics using the TeuiLite and Query Stats tools.
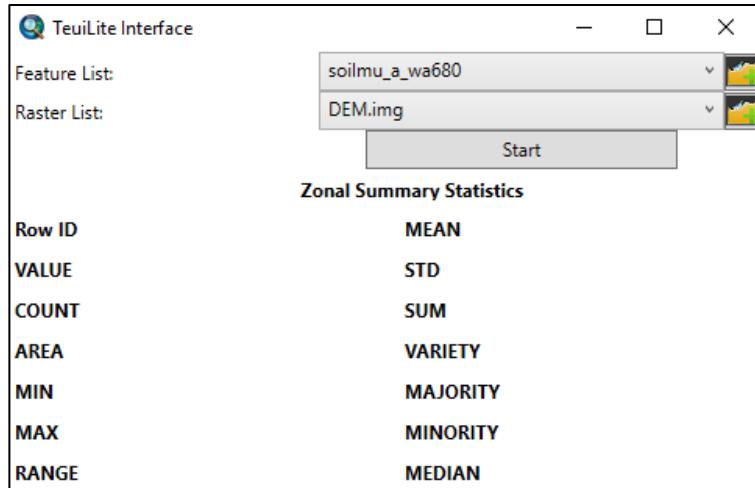
A. TeuiLite

TeuiLite displays the zonal statistics of a polygon which you must select in ArcMap.

*The **Spatial Analyst Extension** must be activated for this tool to work.*
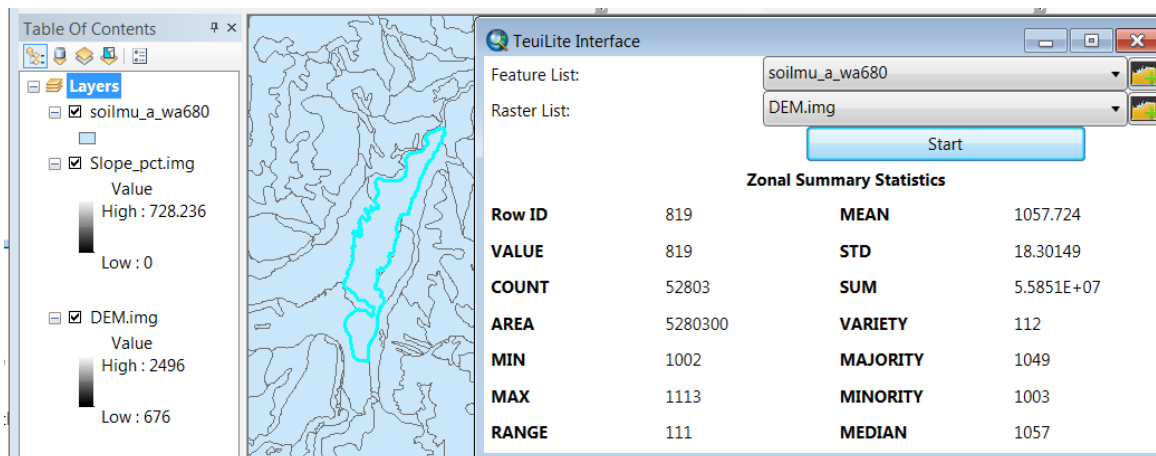
1. On the toolbar, click the **TeuiLite** button.

2. In the **Feature List** menu**,** select the **soilmu_a_wa680 layer** (this tool must use a polygon feature layer).

3. Select a **Raster** layer to query.

4. Click **Start**.

5. In the ArcMap viewer, make sure the **soilmu_a_wa680** layer is visible.

6. Select a polygon and you should then see its zonal stats displayed in the TeuiLite Interface.

B. Query Statistics

The Query Statistics tool provides advanced filtering options for map unit and zonal statistics.

1. On the toolbar, click the **Query Statistics** button.

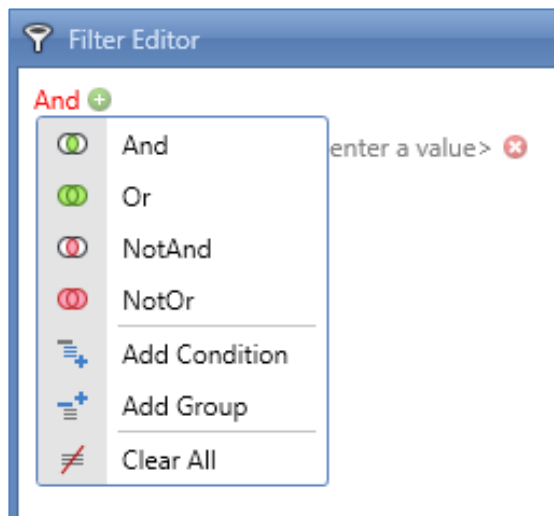2. In the Query Statistics window select the **Zone Layer**, **Data Layer**, **Band**, and either **Map Units** or **Zones**.

3. Click the **Filters** button in the upper right corner of the window to open the **Filter Editor** window.

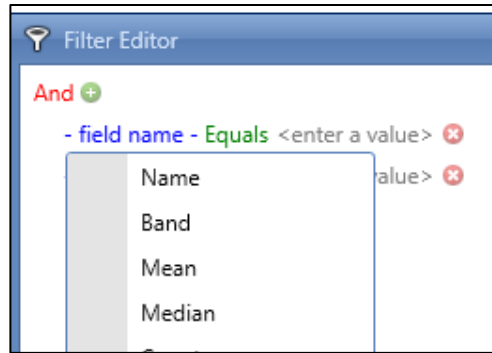   i. Click And to open a menu of filter commands.

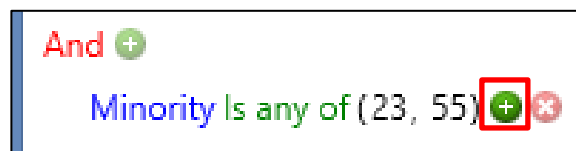   ii. Clicking the green plus button next to And will add a row for filter commands.

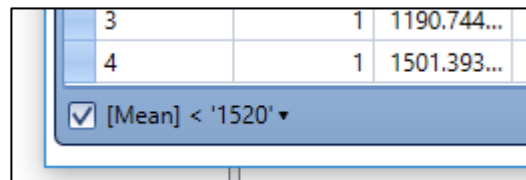4. In the filter command line click the green and blue words to adjust the command.

5. To enter multiple values in a command you must click the green + button to add more fields.


6. Enter value(s) then click **OK**.

7. The bottom left of the Query Statistics window will display the current filter.

- Clicking checkbox on the left will turn the filter on/off.
- The dropdown menu on the right contains all the filter commands you have created.

8. On the bottom right of the Query Statistics window you can click the **editing pencil button** to edit the filter or the **X button** to erase the filter.
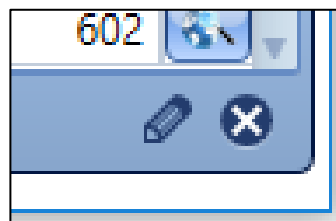
9. Click the globe button on the right side of a row to zoom to that map unit/zone in ArcMap (see below).
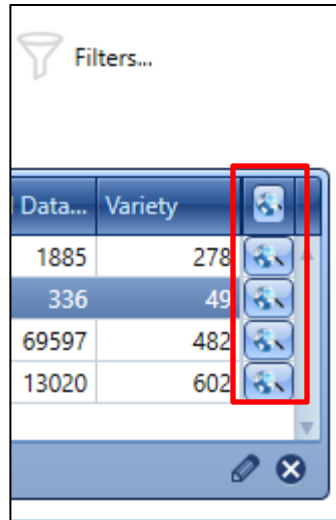
Part 5: Use the Symbology Check Tool

The Symbology Check tool is designed to check the map unit symbology of individual features (zones) against the map unit symbology of individual points which reside within those features. Therefore, if the map unit symbol (or any ID #) of the point data does not match the symbol of the polygon feature within which it falls, then it will be flagged. Points that do not fall within any feature will also be flagged. The user can then zoom to each flagged point and make necessary changes.

A. Add data layers

1. To use the Symbology Check tool you must add (to the Analysis Features field in the TEUI Data dialog) a polygon or raster feature analysis layer that has a map unit symbol identified. You should already have the **soilmu_a_wa680** layer loaded with the Map Unit Column set to **MUSYM** and the **wa680_plots** layer loaded with its Map Unit Column set **to Id**.
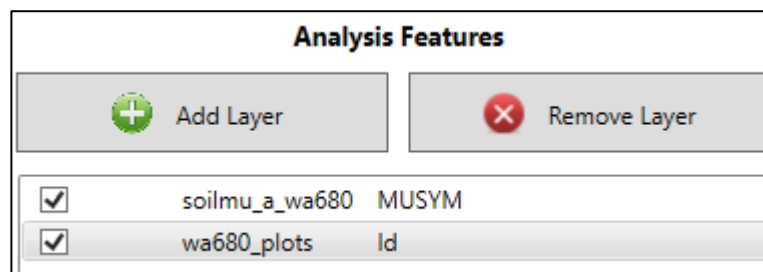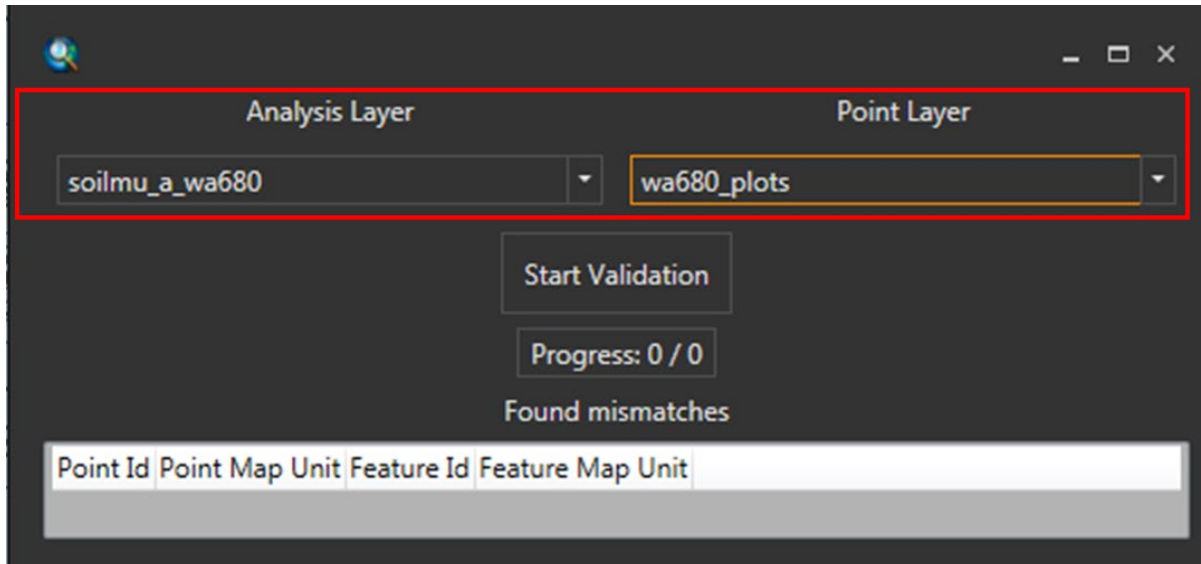

2. Close the TEUI Data dialog.
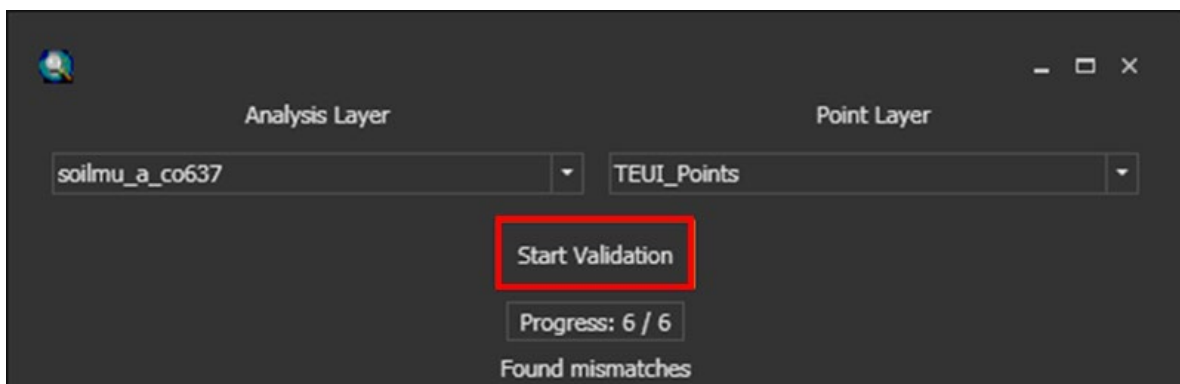
B. Check map unit symbology

1. Click the **green pin icon** on the toolbar to open the Validate Points window (see below).

2. Set the Analysis Layer to **soilmu_a_wa680.**

3. Set the Point Layer to **wa680_plots** (see below).

4. Click on the **Start Validation** button (see below).

5. The notification 'Found mismatches' should appear in the window. Double click in one of the rows and ArcMap will automatically zoom to the corresponding point. In the highlighted row in the following graphic, point **ID 5** has mismatching point and feature map unit symbols. The point map unit symbol is **99** while the feature map unit symbol is **80** (see below). The point map unit symbol should be changed from **99** to **80** so it matches the feature map unit symbol.

6. To correct the mismatches, start an editing session in ArcMap for the **wa680_plots** layer and adjust the **Id** values in the attribute table so that they match the map unit values of the features they reside in.

You have successfully learned the basic process of the managing your geospatial data, running statistics, and checking map unit symbology using the Toolkit!

To learn how to visualize the statistics in both graph and table form, continue on to the next exercise: **Graphs and Tables**.