Demo 07: Tables and Databases

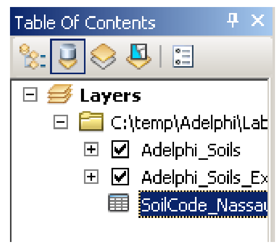
One of the great strengths of GIS is the ability to link non-spatial and spatial data for the purposes of analysis. In Demo 7, you will create relationships between a number of tabular datasets and connect them to spatial data. Specifically, you will use both your vectorized soil map from Demo 06 and a series of tables representing USGS water sampling stations in Indiana

I. First, prepare the demonstration data

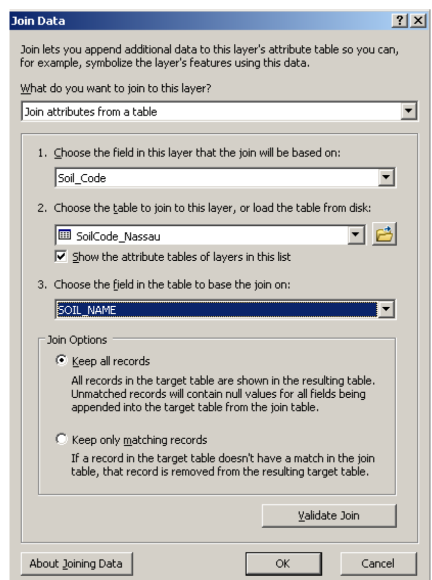
a. Download the **Demo07_Data.zip** file and extract it
b. You should have a series of DBF tables (and a shapefile representing the soils around Notre Dame campus.

II. Joining two tables (creating a *one-to-one* relationship)

a. Start a new blank ArcMap document.
b. Add the soil shapefile from last week.
   1. **Soils_NotreDame.shp**
c. Add **SoilDescriptions_StJoeseph.dbf**
d. Once you have added the table the Table of Contents should switch from "List by Drawing Order" to "List by Source."
   1. This shows all data, whether it has spatial information or not
   2. Files are grouped by the folder containing them, not by their display order.



e. Now look at the data in the new table
   1. Right-click on S**oilDescriptions_StJoeseph.dbf** and choose open.
   2. For each soil code there is also a longer description of what it is.
   3. This makes these data useful for a one-to-one relationship based on the **Soil_Code** field in the **Soils_NotreDame** shapefile
f. To set up this join
   1. Right-click the **Soils_NotreDame** layer and select Joins and Relates→Join… [Note: you can also do this by clicking Add… in the Joins section of the Joins & Relates tab in the **Soils_NotreDame** Layer Properties window.]
   2. Where prompted to indicate what you want to join to this layer, make sure Join attributes from a table is selected;
   3. **Soil_Code** (the field with the identifier for each polygon) is the field in the layer that the join will be based on;
   4. **SoilDescriptions_StJoeseph.DBF** is the table to join to this layer;
   5. **MUSYM** is the field in **SoilDescriptions_StJoeseph.DBF** that the join will be based on. Click OK.
   6. If you are asked to create an index, click Yes.

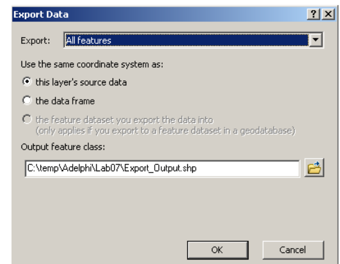


g. Right-click the **Soils_NotreDame** layer and select Open Attribute Table. Scroll to the right, and note that all of the fields from **SoilDescriptions_StJoeseph.DBF** have been appended to **Soils_NotreDame.shp**.
   1. For now, this is a temporary relationship, there is no new data stored in the attribute table **Soils_NotreDame**
   2. There is duplicate (Soil_Code and MUSYM) and unnecessary (OID) data.
   3. You can hide this by right-clicking on the field name (eg Soil_Name) and choosing "**Turn Field Off**"
   4. When you do this for your soils, you may have blank records.
      i. These are rows in your soils shapefile where there is no corresponding value in the

SoilDescriptions_StJoeseph.DBF file

- ii. You will need to fix them in the shapefile's attribute able and redo the join
  - a. The two most common errors are
    1. A space at the beginning of the SoilCode (e.g. “ Ug” vs “Ug”)
    2. Mistaking a lowercase L for an uppercase I (e.g. “PlB”: “P-ell-B” vs “P-eye-B”)
- iii. To fix the errors:
  - a. Remove the join:
    1. Right click on the soils shapefile and select Joins and Relates→ Remove Join → SoilDescriptions_StJoeseph
  - b. Start editing the shapefile and fix the errors
  - c. Redo the join as above in II.f

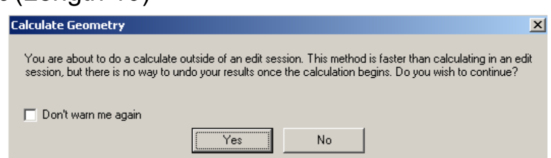
5. To permanently store the new data to the attribute table of **Soils_NotreDame** you will need to export a new copy of it.
   - i. Right-click on **Soils_NotreDame** in the Table of Contents (either view) and choose **Data → Export Data**
   - ii. For this time make use that **All Features** is listed under Export
     - a. Using Selected Features is a quick way to make a new file.
   - iii. Use **The layer's source data** for the coordinate system
   - iv. Store the new file in your Lab07 folder and call it **Soils_NotreDame_WithDesc.shp**



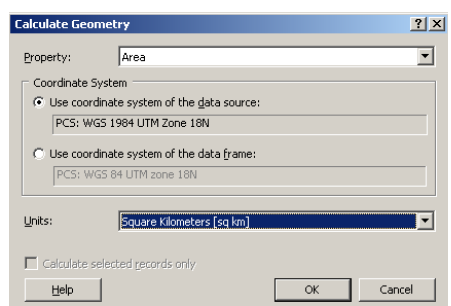
6. In the attribute table for **Soils_NotreDame_WithDesc.shp** you can delete the duplicate **Soil_Name** field by right-clicking on its name and choosing **Delete Field**

III. Calculating fields

- a. In ArcGIS we have two main options for calculating the values of new fields
  1. The **Field Calculator** is used for calculating a new value based on other fields
  2. **Calculate Geometry** lets you calculate x and y locations, area or perimeter.
- b. First we will calculate a field holding the area of each polygon and then we classify theses areas into a text field.
- c. For both of these we must first create a new field to hold the new data.
  1. To add a new field, you need to be in the Attribute Table for that layer
  2. Then click on the Table Options button ( ) and choose **Add Field.** This will bring up a new pop-up window
  3. Add a field called **Area** of type **Double** (leave scale and precision at 0)
  4. Add a second new field called **SizeClass** of type **Text** (Length 10)
- d. To calculate the Area field
  1. Right-click on the name **Area** and choose **Calculate Geometry**
  2. You will get a warning about calculating a field outside of an editing session. This basically means that you can not undo it. If you feel more comfortable you can say No and then start editing.

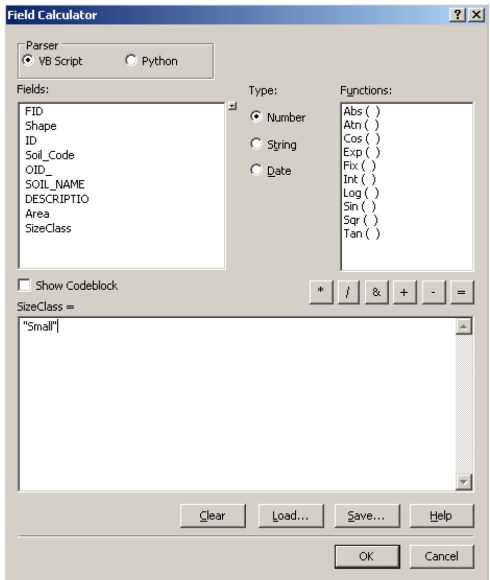


  3. When your proceed you will get another popup where you select the specifics of the calculation
  4. Choose **Area** for the **Property**
  5. The **Coordinate Systems** should be the same, but if not you can choose the correct one.
  6. I would change the **Units** from **Square Meters** to **Square Kilometers**
     - i. This is just because there are 1,000,000 square meters in a square kilometer, so the numbers get huge.
  7. You should now have a field that displays the area for each feature



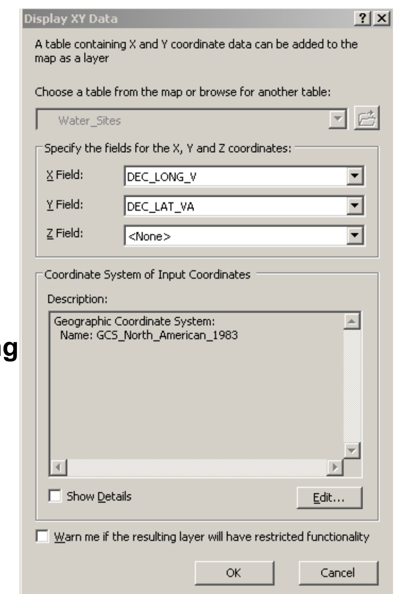
- e. As an example of calculating features based on other aspects, we will now do a quick and dirty classification of these areas into Small, Medium, and Large polygons.

1. If you have selected features, all of the calculate tools will default to only calculating the new field for these features
2. Thus, here we will select small features and then calculate their value, then select medium features and calculate their value.
3. You should have already created the **SizeClass** field, now we should decide what values correspond to what size classes
   i. You should look over your data a bit to decide
   ii. For this example:
      a. Small: less than 0.05 km^2 (258 features)
      b. Medium: between 0.05 and 1 km^2 (147 features)
      c. Large: greater than 1 km^2 (12 features)
4. To perform the classification, we will first select the those polygons with an area less than 0.05 square kilometers
   i. Access the **Select by Attribute** tool either in the Table Options menu ( ) or in the main program S**election → Select by Attribute**
   ii. In the box either type **"Area" <0.05** or use the buttons to match it.
   iii. Click **Apply**
   iv. Now you should have several selections in the attribute table.
   v. Right-click on the **SizeClass** filed and choose **Field Calculator**
   vi. From this new popup you can create a new value using any of the other fields in addition to VisualBasic or Python Code
   vii. Here simply type **"Small"**
      a. The quotes are important, otherwise it thinks it is a variable
   viii. When you choose OK, all of the selected values should be updated in the attribute table
   ix. Now we will repeat this for medium and large
   x. For medium,
      a. Select by Attributes with the code **"Area" > 0.05 AND "Area" < 1**
      b. Repeat as above, but type **"Medium"** in the **Field Calculator** window
   xi. For large,
      a. Select by Attributes with the code **"Area" > 1**
      b. Repeat as above, but type **"Medium"** in the **Field Calculator** window
   xii. You should not be left with any unclassified polygons.



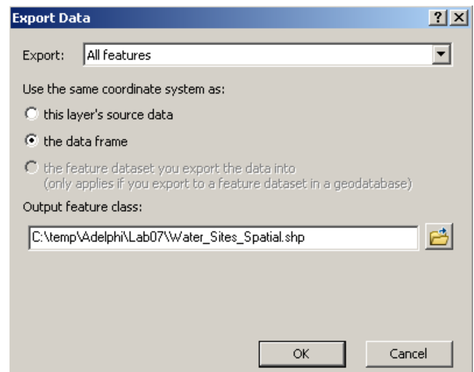
IV. Creating a point shapefile from XY data

a. We will now work with a series of USGS water sampling sites for all of Indiana.
b. This data is in two tables,
   1. **Water_Sites.dbf:** Contains spatial and descriptive information for each Long Island sampling site
   2. **Water_Measurements.dbf:** Contains the type and value of each measurement.
c. Add both of these tables to your ArcMap document.
d. Open **Water_Sites.dbf** by right-clicking on the name in the Table of Contents and choosing **Open**
e. You should see the name (or number) of each sampling site with the coordinates in decimal degrees (**Lat** and **LONG**) and in UTM (**Northing**



and **Easting**).

f. To create a shapefile from these coordinates:
   1. Right-click on the **Water_Sites.dbf** in the Table of Contents and choose "**Display XY Data**"
   2. In this menu choose:
      i. X Field: **Easting**
      ii. Y Field: **Northing**
      iii. Z Field: Leave as **<None>**
      iv. The Coordinate System should be correct (WGS_1984_UTM_Zone_16N)
         a. If not, Click Edit to set the Coordinate System
      v. Choose OK when you have finished.
   3. This will bring up a temporary layer in the Table of Contents called **Water_Sites Events**
      i. Although this layer is temporary, you can treat it as you would a normal shapefile
      ii. Be careful about editing this temporary layer, it will change the file as well.
   4. You should now make this temporary events layer permanent the same way above in step **II.g.4**
      i. Right-click on **Water_Sites Events** in the Table of Contents and choose Data → Export Data
      ii. Verify that All Features is shown at the top
      iii. Call the file **Water_Sites_Spatial.shp** and put it in your Lab06 folder.
      iv. In this case under "**Use the same coordinate system as:**" choose **the layer's source data**
      v. Click OK and say yes when it asks about adding the new layer.
   5. You can now remove both **Water_Sites Events** and the original **Water_Sites** table

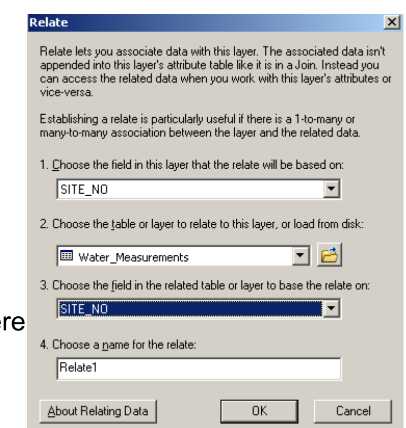


V. Relating tables (creating a *one-to-many* relationship)

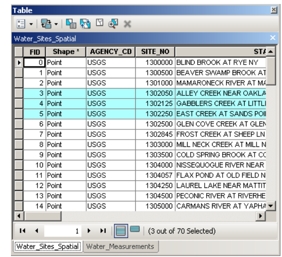
a. Now we want to create a *one-to-many* relationship between the features in the **Water_Sites_Spatial.shp** and the measurements in **Water_Measurements.dbf**

b. If you open Water_Measurements.dbf you will see five fields
   1. **OID:** Is the ordering field created in the dbf format. You should not change these values
   2. **ID**: A more complicates Station ID
   3. **STATION:** The site number. This will be the field related to the features in **Water_Sites_Spatial**
   4. **PARAM:** What measurement is being recorded
   5. **YEAR :** The year being measured
   6. **Value :** The mean value of Param_ during YEAR_NU.
   7. **Number**: The total number of measurements contributing to Value

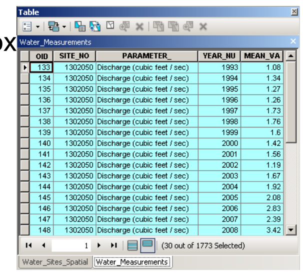
c. To relate these two tables so we can query values based on spatial location
   1. Right-click on **Water_Sites_Spatial** in the Table of Contents and choose **Joins and Relates → Relate...**
      i. You can actually do this from either **Water_Sites_Spatial** or **Water_Measurements**
   2. In the Relate window:
      i. The first box is the similar field in the first table. Here is it **STATION**
      ii. The second box is where you decide what table to relate it to. Here use **Water_Measurements**
      iii. The third box is the similar field in the other table. Here is is also **STATION**
      iv. In the last box you can give the relate a name if you want. This is only really useful if you have many related tables.
   3. Finish and choose **OK**
   4. To explore the new relate first open the attribute table for **Water_Sites_Spatial**



5. Select one or more features in this attribute table.
6. To select the the data for the same sites in **Water_Measurements** click on the related tables button () at the top of the attribute table. You then choose the related table in a dropdown. Here choose "**Relate1: Water_Measurements**"
7. This will open the table for **Water_Measurements** with the features from these sites collected



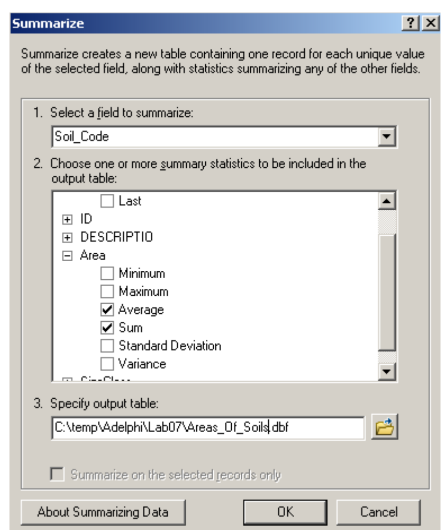
8. Experiment with selecting the features that are within or near your study area.
9. The relate goes in both directions; you can select a Param from the Water_Measurements table and see the related sites
   i. Use Select By Attribute to select all of the Salinity Measurements
      a. Access the **Select by Attribute** tool either in the Table Options menu () or in the main program S**election → Select by Attribute**
      b. Either type **"Param" = 'OXYGEN, DISSOLVED'** in the bottom box or build the equation by double-clicking on **"Param"** then the **=** button then **Get Unique Values** and then '**OXYGEN, DISSOLVED'**
      c. Click **APPLY**



   ii. Move this selection back to Water_Sites_Spatial by clicking () at the top of the attribute table and choosing "**Relate1: Water_Sites_Spatial**"
   iii. The attribute table for **Water_Sites_Spatial** indicates that 1055 out of 1922 records have been selected.
   iv. You can locate the selected collection unit spatially by minimizing the tables. The selected features will appear on the map as a cyan circle. If it is not visible, click the Full Extent button to zoom all the way out.
10. Experiment with querying back and forth on this relate.

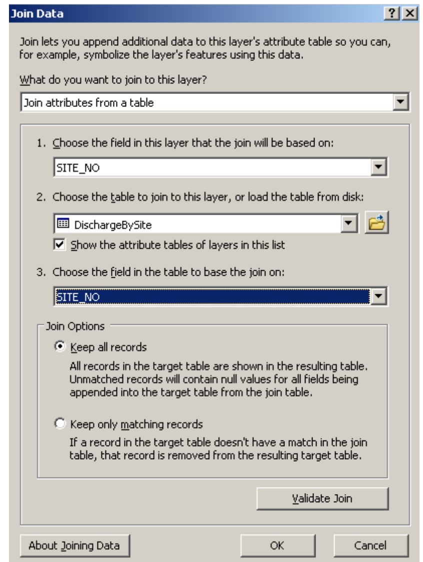
VI. Summarizing fields

a. Here we will do two examples of using field summaries to display data in novel ways. The first reduces spatial data to a non-spatial summary table. The second summarizes many cases for joining to spatial data.
b. First, we will create a summary table for **Soils_NotreDame_WithDesc** that tells the average and total area for each soil type
   1. This example assumes that the **Area** field was calculated in section **III.d.** If you have not done this, complete that part first.
   2. Now open the attribute table for **Soils_NotreDame_WithDesc**
   3. At the top of the **Soil_Code** (or Soil_Name if you kept it) field right-click on the name and choose **Summarize**
      i. Verify that **Soil_Code** is selected at the top
      ii. In the big box, expand **Area** and check **Average** and **Sum**
         a. You may choose others if you wish
         b. The **Count** is included by default
      iii. Specify an output location in your Lab07 folder. Call the file **Areas_Of_Soils.dbf**
      iv. Say **Yes** when it prompts you to add it.
   4. Open the Attribute table for this new Areas_Of_Soils table.
      i. For each soil code there is the number of instances in the original shapefile, the total area and the average area.



c. Second, we will summarize the **Water_Measurements** values for each site and join them to the **Water_Sites_Spatial**
   1. Open the **Water_Measurements** table

2. Check that you do not have any features selected.
   i. If you do, use the **Clear Selection** button ( )
   ii. If the table looks blank then you may need to switch the buttons at the bottom of the table from **Show Selected Features**( ) to **Show All Features** ( )
3. Now you need to select one particular **Param** value to summarize.
   i. Use **Select By Attributes** to select only those entries with **Discharge (cubic feet / sec)**
   ii. The code should look like **"PARAMETER_" = "OXYGEN, DISSOLVED"**
4. With this selection in place, summarize the **STATION** field
   i. Right-click on the name of the field and choose **Summarize**
   ii. In the Summarize window
      a. Under **Value** check **Average**
      b. Under **Number** check **Sum**
      c. Output the new file to your Lab07 folder and call it **DissOxyBySite.dbf**
      d. **IF YOU GET AN ERROR WHEN YOU PRESS SAVE**
         1. You probably need to change the Save as Type from “File and Personal Geodatabase Tables” to **dBase tables**
   iii. Say yes when it asked to add the new table
   iv. This will give you a summary table that contains the average dissolved oxygen for each site across the entire time period as well as the total number of samples that contributed to that measurement.
      a. Yes, I know this is not statistically valid, if anyone wants to know how to weight it based total numbers, just let me know



5. Now, join these data to the **Water_Sites_Spatial** file using the **STATION** field in both of them.
6. You should see that there are some values listed as <NULL>. These are sampling stations where the DO was not recorded.
7. As before, you can make this file permanent.