Demo 11: Basic Spatial Analysis

This demo will introduce you to several methods for spatial analysis. Most of the ones here are based on distance, but we will work with creating models from your preexisting data as well. Part of the assignment for this week is to find an optimal location for defense within your study area.

I. Prepare the Demo data

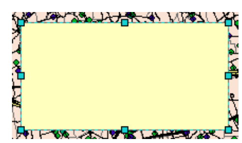
a. There is a **Demo11_Data.zip** file. It contains the example data I am using in the demo
   1. These is also a repeat of some data from Demo 03.
b. When you do the assignment, you will be using a lot of data you created or processed in past weeks.
   1. From Demo 06:
      i. Your Soils shapefile, preferably with the descriptions appended to the attribute table (Demo 07)
      ii. Your Soils_Extent shapefile
   2. From Demo 10:
      i. The DEM, Slope and Aspect rasters for your study area. The TIN will not be used.
      ii. The water shapefile you used to make the TIN

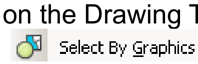
II. Interactive selections

i. We have done most of this in earlier demos, so this will be a review

b. Select by graphic
   1. You can use select by graphic as a quick way to select features in a specified area
   2. These graphics are not shapefiles, they are simply a drawing in your viewer.
   3. They do have some spatial information, but it is not stored anywhere permanently.
   4. To **Select by Graphic**
      i. Draw a graphic in the viewer using the Drawing Toolbar

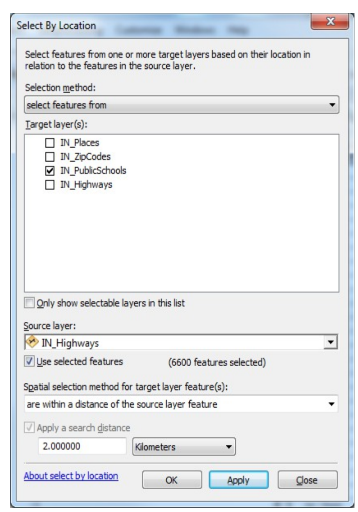
      ii. If the Drawing Toolbar is not visible, you can activate it by right-clicking on a blank spot at the top of the screen and choosing "Draw" or by going to **Customize → Toolbars → Draw**
      iii. In the **Feature Type Selection** () choose either **Rectangle** or **Polygon**
      iv. Draw your selection in the viewer.
      v. Here I will draw a rectangle through the middle of Michiana
      vi. Make sure that your graphic is selected: ie outlined in cyan and with the resize squares



         a. It not, click it with the black arrow() on the Drawing Toolbar
      vii. Now, in the top menu choose **Selection** → 

      viii. The limitation of **Select by Graphic** should now be clear: It will select features in all layers.
         a. If you want to control this, you can set the Selectable Layers (like Demo 5, section IV.c)
         b. First, change the Table of Contents from "List by Drawing Order" to "List by Selection" 
         c. Then click on the () button next to each unwanted layer to disable selection
      ix. Experiment with this. Try selecting features in just one layer from the graphic

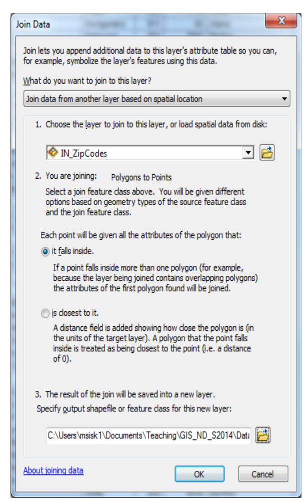
c. Select by Locations
   1. We have already done this (Week 03), but as a quick review, **Select by Location** allows you to quickly query vector data based on their spatial location.
   2. The Select by Location tool is accessed via Selection → Select by Location



- i. You have the following options
  - a. **Selection Method**: Here you can
    1. Create a new selection replacing any previous selection
    2. Add the newly selected features to a previous selection
    3. Subtract the newly selected features from a previous selection
    4. Select features from within the currently selected features and not from the whole data set.
  - **b. Target Layer(s)**: Here you set one or more layers to be selected
  - c. **Source Layer**: Here you set which layer you a selecting relative to
  - d. **Spatial Selection Method**: Here you set whether you are interested in:
    1. Intersections
    2. Within
    3. Contain
    4. Within a distance
    5. There are others here as well. A full graphical description of each can be found here, or by searching the ArcGIS help for "Select by Location: Graphic Examples"
  - e. **Search Distance**: This is optional for most (except Within a Distance), but it sets a buffer around the Source Features

3. Experiment with selecting by location in the Indiana data. In particular remember that you can stack selections.
   - i. For example, if I was interested in finding all of the Public Schools that were within 2 kilometers of a major road I would
     - a. Select major roads in the IN_Highways.shp using **Select By Attributes**
       1. Major roads are CODE I or U:
       2. So the code should look like **"CODE" = 'I' OR "CODE" = 'U'**
     - b. Then, I would use Select by Location to only choose the schools that were within 2 km of these selected roads
       1. **Selection Method**: Select features from
       2. **Target Layer:** IN_PublicSchools.shp
       3. **Source Layer:** IN_Highways.shp (make sure Use Selected Features is checked) ##
       4. **Spatial Selection Method**: Target features(s) are within a distance of the Source layer feature
       5. **Search Distance**: 2 Kilometers.
   - ii. You will be expected to perform two other selections like this during the assignment.

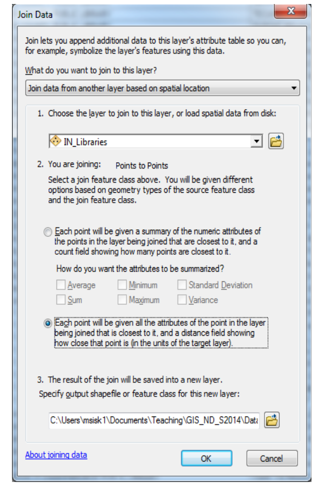
4. Spatial joins



   - i. We can also permanently add the data from one layer to the attribute layer of another based on its spatial location.
   - ii. This is like the Joins based on a field we did in Week 06
   - iii. As an example, we will add information on Zip Codes to the shapefile of cities and towns (**IN_Places.shp**)
   - iv. To perform a spatial join:
     - a. Right click on the layer name (of the table to which you want to add data) and choose **Joins and Relates → Join...**
       1. Here, right click on **IN_Places.shp**
     - b. In the top box change it from "**Join Attributes from a Table"** to "**Join Data from Another based on Spatial Location"**
       1. This sets it as a spatial join, which is different than what we did before.
     - c. In Section 1. you choose the layer from which data will be added.
       1. For this example, choose **IN_ZipCodes.shp**
     - d. Depending on the type of data in the two files, the options in Section 2 will be different.
       1. Here we have "Polygons to Points" so we have the options to just use the polygons that

points are within or the closest (with a distance measure)

2. These options are one of the few things in ArcGIS that are very well explained, so as you experiment you should be able to figure out the options, but if not, try the help file or email me



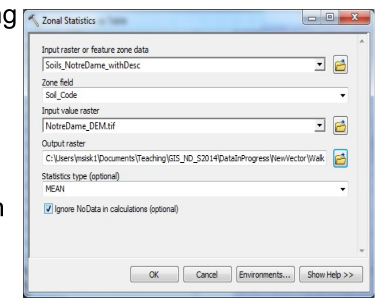
e. Finally, Section 3 is where you export the new shapefile.
   1. Save it permanently. Here I call it **IN_Places_With_ZipCodes.shp**

v. We can also find out what is the closest library to every public school.
   a. Make sure you clear the selected features
   b. Right-click on **IN_PublicSchools** and choose **Joins and Relates → Join...**
   c. Section 1: Choose IN_Libraries
   d. Section 2: Choose the second option (each point will get the value of the closest library and its distance in map units)
   e. Section 3: Call the output **Schools_With_Libraries**

vi. Again, experimenting with this technique is part of the assignment.

III. Zonal Statistics

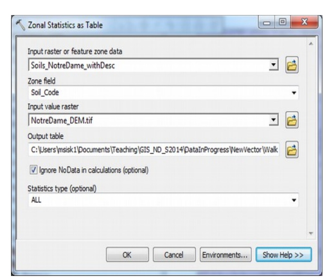
a. Zonal statistics are an easy way to summarize the values of a raster over particular areas
b. Here, we will compute the average elevation and slope for each of your soil classes.
   1. This could be useful for characterizing each soil type or for looking at erosion risks.
c. First, we will make a new raster layer representing the mean of each class using the Zonal Statistics tool



   1. Click on the toolbox icon () or go to **Geoprocessing → ArcToolbox**
   2. Within the ArcToolbox window choose: **Spatial Analyst Tools → Zonal → Zonal Statistics**
   3. In the new window you will set:
      i. The **Zone Data**: This can be either a polygon shapefile or a raster with a discrete number of classes.
         a. Here it is your soils shapefile (**Soils_NotreDame_WithDesc.shp**)
      ii. The **Zone Field:** This is the field with information about the zones.
         a. Here we use the **Soil_Code** because we want a summary of each soil type.
      iii. The **Input Value Raster**: This is the raster data to be summarized.
         a. Here it is **NotreDame_DEM.tif**
      iv. For the output put it someplace permanent. I called mine **SoilZones_ElevMean.tif**
      v. Choose the statistic to be calculated. Here is should be MEAN.
   4. Choose OK and you will get a new layer with a different color representing the mean of each soil type.

d. I find these images kind of useless for two reasons.
   1. There is only one statistic (min, max, mean, std) in each layer
   2. They repeat spatial data already available in the inputs

e. Thus, I find it much more useful to calculate **Zonal Statistics Tables,** which contain as many statistics as you want and can be joined to spatial data.
   1. Within the ArcToolbox window choose: **Spatial Analyst Tools → Zonal → Zonal Statistics as Table**
   2. In the **Zonal Statistics as Table** window the options will be largely similar



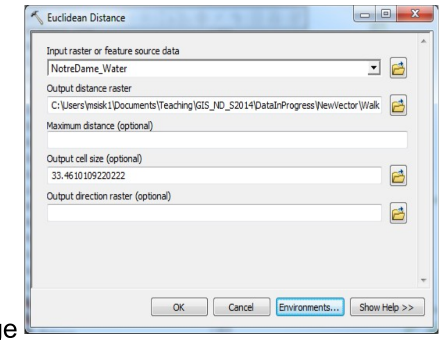
      i. The **Zone Data** is **Soils_NotreDame_WithDesc.shp**
      ii. The **Zone Field** is Soil_Code
      iii. The **Input Raster** is **NotreDame_DEM.tif**
      iv. Put the **Output Table** in your permanent Lab09 folder. I called mine **SoilZones_Elev_Table.dbf**
      v. Leave **Statistics** as ALL.
   3. Click OK and look at the new table. For each of the soil codes you have many different statistics on their elevation values

f. Now repeat this process for the Slope.
g. You now have two data tables with descriptive statistics for the Elevation and Slope of each soil_code.
    1. These data could be combined into one table in Excel or OpenOffice
    2. If, when creating the table, you choose to use a unique ID field (instead of the soil code) you could join the summary data back onto the soils shapefile attribute table.

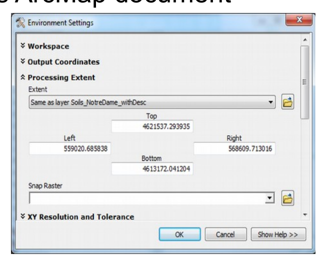
IV. Distance Measurements

a. Euclidean (straight-line) distance

1. The simplest distance measurement is a simple straight line, or Euclidean distance.
2. Here, we will calculate the distance to water within the study area
3. To calculate a Euclidean distance: Within the ArcToolbox window choose: **Spatial Analyst Tools → Distance → Euclidean Distance**



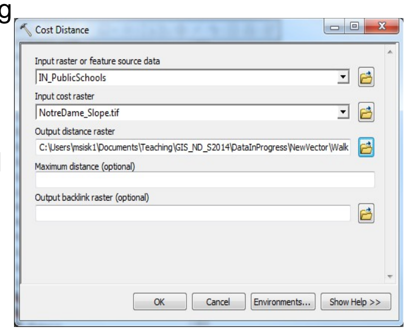
4. In the Euclidean Distance Window
    i. The **Input Raster or Feature Source Data** is the shapefile NotreDame_Water.
        a. It it actually rare and difficult to do this to a raster, so it will nearly always be a shapefile
        b. Remember, that if any features are selected only they will be used.
    ii. I called my output: **EucDistToWater.tif**
    iii. Before proceeding, there is one complex setting we need to change
        a. This, and most of the spatial analyst tools, default to using a strange extent.
        b. For this to work correctly, we have to change one of the **Environment Variables**,
        c. There are two options
            1. We can change the Environment Variable for all analysis in this ArcMap document
                i. To access the variables this way, exit the Euclidean Distance window and choose **Geoprocessing → Environments** at the top of the screen.



            2. Or, we can change it just for this analysis by clicking the **Environments** button at the bottom of the Euclidean Distance window.
            3. In either case, you then click on Processing Extent in the new window and change the **Extent** from "**Default"** to "**Same as Layer Soils_NotreDame_withDesc"**
                i. Yours may be different.
    iv. We will not use the other options for now.
5. Click OK
    i. If it comes up as an odd size, you did not set the Environment Settings correctly
6. This new image is a raster for which the only pixel value is distance from one of the polygons in NotreDame_Water.shp

b. Cost-weighted distance

1. Euclidean Distances are useful, but in many cases difficulty in crossing a landscape is not as simple as pure distance
2. In these cases we can weight the distance by a factor representing ease of travel
3. The creation of this factor can be extremely complex, including vegetation types, slope, soil types, types of roads and a multitude of other factors.
4. Here, we are simply going to weight the distance by the slope and determine how hard it would be to walk to the various public schools in the study area.
5. In the ArcToolbox window choose: **Spatial Analyst Tools → Distance → Cost Distance**

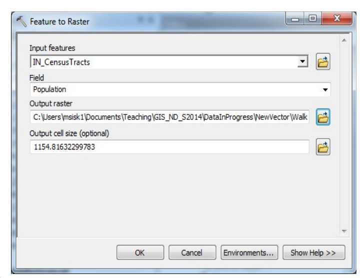


i. The **Input Raster or Feature Source Data** is **IN_PublicSchools.shp**
ii. The **Input Cost Raster** is the raster representing how easy it is to cross the terrain. Here it is **NotreDame_Slope.tif**.
iii. Save the output to your Week08 folder and call it **CostDistToSchools.tif**
iv. **BEFORE CLICKING OK.** Make sure that you have set the Extent in the Environment Variables (In Section IV.a.4.iii above) to the extent of the soils. Failure to do this will mean it runs the analysis for all of Indiana and creates a huge file.

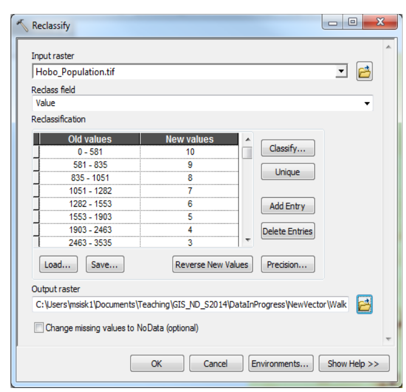
6. Once you have set the Environment Variables, click **OK**
7. The output is much more complex, but gives a better indication of how hard it would be to reach the closest school from each cell of the raster.

V. Raster Math

a. Until now we have been creating individual rasters representing a single factor. In the following section we will introduce using these tools to combine layers and model optimal locations.
   1. We will continue this process next week by modeling erosion and hydrology
b. The construction of these sorts of models is complex, and the best way to do them is to start with a good question and build all of the data layers. Then we can combine all the layers into a single model.
c. Here, we will help a young hobo find the best locations to live in Indiana.
d. First we need to know what factors are important to the model.
   1. Our newly graduated hobo has a series of needs that we want to think about and convert into raster layers that we can combine
      i. The most attractive thing about being a hobo is solitude, so we will avoid high population areas.
      ii. Being a hobo does not pay well, so being near public libraries is important.
      iii. Railroads are the lifeblood of the hobo, so we must be near locations of railroads
e. We are now ready to create layers representing these three components (populations, libraries and railroads)
   1. We already have all of the data we need to model this, but we will need to transform it all into rasters that are at the same scale
   2. **Population**:
      i. Using the **IN_CensusTracts.shp** layer we want to create a raster layer based on the population and then reclassify it into good and bad areas.
      ii. To convert a shapefile to a raster:



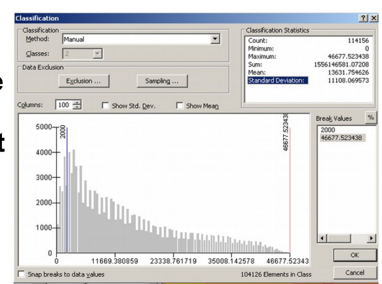
         a. In the ArcToolbox choose **Conversion Tools → To Raster → Feature to Raster**
            1. The input features is **IN_CensusTracts.shp**
            2. The field is **Population**
            3. Call it **Hobo_Population.tif** and save it to your drive.
         b. The new raster has values ranging from 0 – 25,000 based on the population of each area.
         c. It will be more useful for our model if we scale this from 1 – 10 based on how LOW the population is (this means the higher values would be more preferable for our hobo).
      iii. We do this with the **Reclassify Tool:**



         a. ArcToolbox → Spatial Analyst Tools → Reclass → Reclassify
         b. Here we can reassign the cell values based on ranges. This is like a classified symbology.
         c. The Input Raster is **Hobo_Population.tif** and the field should be Value
         d. The initial Old and New Values are a little odd. The easiest way to update these is by clicking the **Classify** button.
            1. In the new window select either **Natural Breaks** (**Jenks**) or **Equal Interval** and set the number of

classes to 10.
        2. Choose OK.
    e. Now remember, that we want a good locations (few people) to have the highest values. Right now, the higher the population, the higher the new values.
    f. Click the **Reverse New Values** button and this will change
    g. Save your output and call it something like **Hobo_PopulationRECLASS.tif**
    h. Click OK

3. **Libraries**
    i. Now we want to create a layer that ranks the distance of each cell by how our hobo is from the nearest library.
    ii. We will not use Cost-Weighted here (Although we could) because we do not have slope information for all of Indiana.
    iii. Instead, run a Euclidean Distance (Section IV.a) on the libraries data
        a. The Input is **IN_Libraries**.**shp**
        b. Save the output as **Hobo_Libraries.tif**
        c. You don't need to reset the Environment Settings (unless you made a global change) because we want the full extent of the file.
    iv. Now we want to use the **Reclassify Tool** to rank each cell based on how close it is.
        a. This will be exactly the same process as for the population
        b. Call the output **Hobo_Libraries_RECLASS.tif**
        c. Make sure that you **reverse the values** so that 10 is closest and 1 is furthest
4. Railroads.
    i. Finally, our hobo will not, under any circumstances live more than 5 kilometers away from a railroad. To do so would be make him a hermit, not a hobo. And those guys hate eachother.
        a. Thus, this is binary character: distance more than 5 km is impossible, distance less than 5 km is unimportant
    ii. To calculate this, we first want to create a Euclidean Distance layer to the University data
    iii. Again, follow Section IV.a
        a. The Input is **IN_Railroads**.**shp**
        b. Save the output as **Hobo_Railroads .tif**
        c. You don't need to reset the Environment Settings (unless you made a global change) because we want the full extent of the file.



    iv. Now we want to reclassify again (**ArcToolbox** → **Spatial Analyst Tools** → **Reclass** → **Reclassify**) but this time we only want two classes, less than 5 km away, and more than 5 km.
        a. The Input is **Hobo_Railroads.tif**
        b. The Output is **Hobo_Railroads_RECLASS.tif**
        c. Click the **Classify** button and change the number of classes to 2. Now you can edit the value by clicking on the top number under **Break Value**s
        d. Change it to 5000
            1. 5000 meters = 5 km
        e. Click OK to exit the Classification window.
        f. Now Change the New Values from 1 and 2 to 1 and 0
            1. This makes it a binary system.
        g. Click OK and you should be left with a two color image

Old values	New values
0 - 5000	1
5000 - 90066.671875	0
NoData	NoData

f. Creating the final model of a hobo's ideal habitats.
    1. Now we are ready to combine these three new layers into one spatial model.
    2. We do this with the **Raster Calculator**: **ArcToolbox** → **Spatial Analyst Tools** → **Map Algebra** → **Raster Calculator**
    3. This works like a regular calculator, following the same order with multiplication or addition
    4. Here, we want to take the average of the Population and Libraries layers and then multiply by the railroads.

5. The equation should look like:
    i. **"Hobo_Railroad_Reclass.tif" * (("Hobo_Libraries_Reclass.tif" + "Hobo_Population_Reclass.tif") / 2)**
    ii. The initial multiplication by the binary Hobo_Railroad_Reclass.tif meakes it so that any pixel within 2 km of a university will get a value of 0 and any other a value of 1
    iii. Be careful with parentheses
    iv. You can do the equation by hand, or by clicking on the icons and layer names.
6. Save this output in a permanent location, called **Hobosity_Indiana.tif**

g. You now have a single raster layer that represents how suitable each area is for a hobo. This example is a bit silly, but imagine that you would make these sorts of models using real data as well.