

# Lab 5: Data Model Transformations

## 1.0 Overview

Although modern GIS systems can manage both raster and vector data, it is still often necessary to convert from one data model to the other. The primary reasons are that 1) the vector format stores discrete features more efficiently than does the raster format and 2) it is often necessary to have all data in one format for analytical operations. ArcGIS for Desktop provides simple-to-use tools for converting between data models. We will experiment with these tools in this lab. Along the way, we will also learn how to store symbology in a layer file, how to measure map distances with an ArcMap tool, and how to compute summary statistics of a numeric attribute.

The overall steps of this lab are:

- Convert vector polygon, line, and point features to a raster data model
- Create a map comparing the results of rasterized features at different raster cell sizes
- Measure the offset distance of vector point features converted to raster
- Convert raster features to vector polygon, line, and point features
- Summarize and compare area of vector polygon features to area of the same polygon features that were rasterized at different raster cell sizes, then converted back to vector
- Summarize and compare length of vector line features to length of the same line features that were rasterized at different raster cell sizes, then converted back to vector

**\*\* Before you begin this lab, review the "General" section of the [ArcGIS Tips and Tricks](#) on the course webpage. **Don't skip this step...it will ultimately save you time!****

## 2.0 Rasterization

“Rasterization” is the process of converting features in the vector data model to a raster data model. There are three possible types of conversions:

- *Polygon features to raster:* When polygons are converted, all grid cells lying within the vector polygon are assigned the attribute value of the polygon. Edges are assigned cell values by taking the vector feature that is most dominant.
- *Line features to raster:* Cells are given the value of the line that intersects each cell. Cells that are not intersected by a line are given a value of NoData. If more than one line is found in a cell, the cell is arbitrarily given the value of the first line it encounters when processing. Using a smaller cell size during conversion will alleviate this problem, but could considerably increase storage requirements.
- *Point features to raster:* When converting points, cells are given the value of the point found within each cell. Cells that do not contain a point are given the value of NoData. If more than one point lies within a cell, the cell is arbitrarily assigned the value of the first point encountered when processing. Using a smaller cell size will also alleviate this problem.

The cell size you choose when converting from vector to raster should be based on several factors, mainly the resolution of the input data, the output raster resolution needed to perform analysis, the need to maintain an efficient processing speed, and storage limitations.

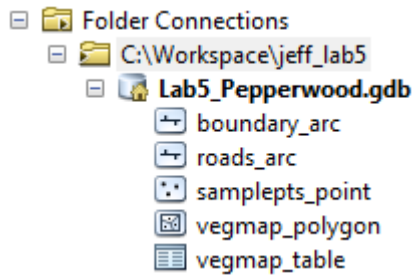
## 2.1 The Scenario

The invasive weed, Yellow Starthistle (*Centaurea solstitialis*), has aggressively spread throughout many regions of California. You have been hired by Pepperwood Preserve (just Northeast of Santa Rosa) to develop a GIS model that predicts habitats that are susceptible to the establishment of this species. To complete this task, there are several basic data layers in vector format that need to be converted to raster. The data layers to convert are 1) a polygon vegetation map classified as forest, grassland, and chaparral, 2) a line map of trails, and 3) a point map of 87 field plots characterizing the abundance of Yellow Starthistle.

## 2.2 Download and inspect the vector data

Make a Lab 5 working folder on your USB drive or local drive (C:/workspace). Download and unzip *geog387\_lab5.zip* into your Lab 5 folder. The zip file contains a file geodatabase named *Lab5\_Pepperwood.gdb*. Open a new blank ArcMap document. In the ArcMap - Getting Started dialog screen, **set the default geodatabase to your *Lab5\_Pepperwood.gdb* geodatabase**. Click OK. Alternatively, the default geodatabase can be set under File→Map Document Properties→Default Geodatabase.

In the Catalog window in ArcMap, establish a folder connection to your Lab 5 working folder. Expand the view of *Lab5\_Pepperwood.gdb* so that you can see the feature classes and the table that it contains.



Open the Item Description for the *vegmap\_table* and note the field names in the Preview tab. Also preview the *vegmap\_polygon* attribute table and note its field names (Preview tab, set to view the “Table”).

### Question 1: (2)

Which key do the two tables share that could be used in a table join? (hint: it is not OBJECTID).

## 2.3 Symbolize the vector data

We will now load the layers from our geodatabase and symbolize them with some cartographically-appropriate colors. Add the *vegmap\_polygon* feature class and the *vegmap\_table* to your map document.

Use your skills from Lab 4 to join the table to the feature class using the common key that you identified in Question 1. The fields from the *vegmap\_table* are now visible in the feature attribute table. However, the join is stored only in the ArcMap Document (.mxd file). To permanently save the joined fields to the *vegmap\_polygon* feature attribute table, export the feature class to a new feature class in your geodatabase named *vegmap\_polygon\_join*. (Use Data→Export Data; be sure that none of the polygon features are selected). The new feature class, with the *vegmap\_table* permanently joined, is added to the TOC. You can now remove *vegmap\_polygon* from the TOC.

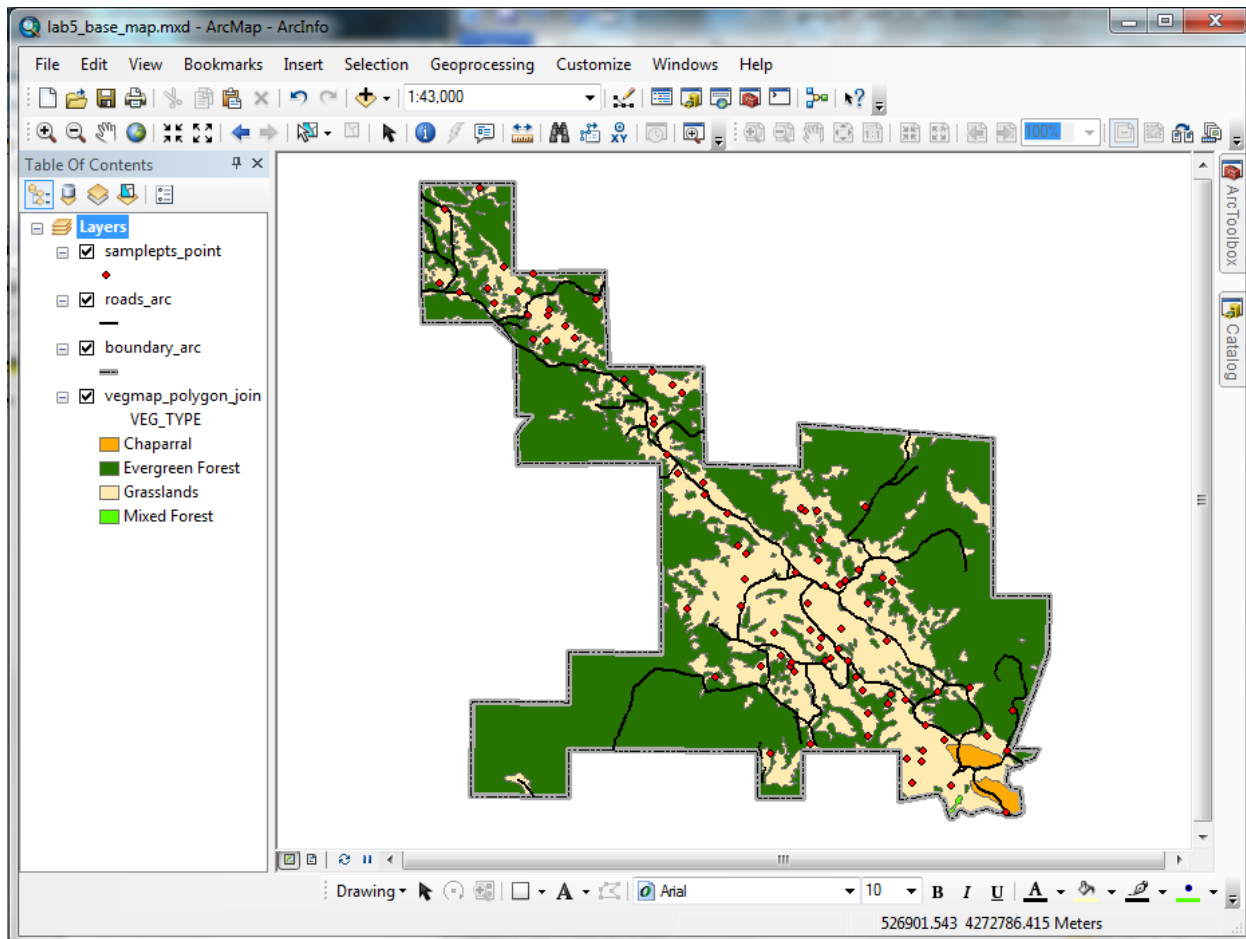
Open the feature class symbology tab for *vegmap\_polygon\_join* and chose the Categories, Unique Values symbology option. For value field, choose the VEG\_TYPE field. Click the Add All Values button. Deselect the <all other values> option. Considering the lecture on cartographic design, choose appropriate colors for each class by double-clicking on the color patch for each class, and finally click OK. Change the TOC view from List by Source to List by Drawing Order.

Load the *boundary\_arc* feature class, which delineates the boundary of Pepperwood Preserve. Single-click the line symbol below *boundary\_arc* layer in the table of contents (TOC). The symbol selector box opens. Choose the Boundary, State symbol in the Symbol Selector window.

Next load the *roads\_arc* feature class. Open the Symbol Selector and choose the Major Roads symbol.

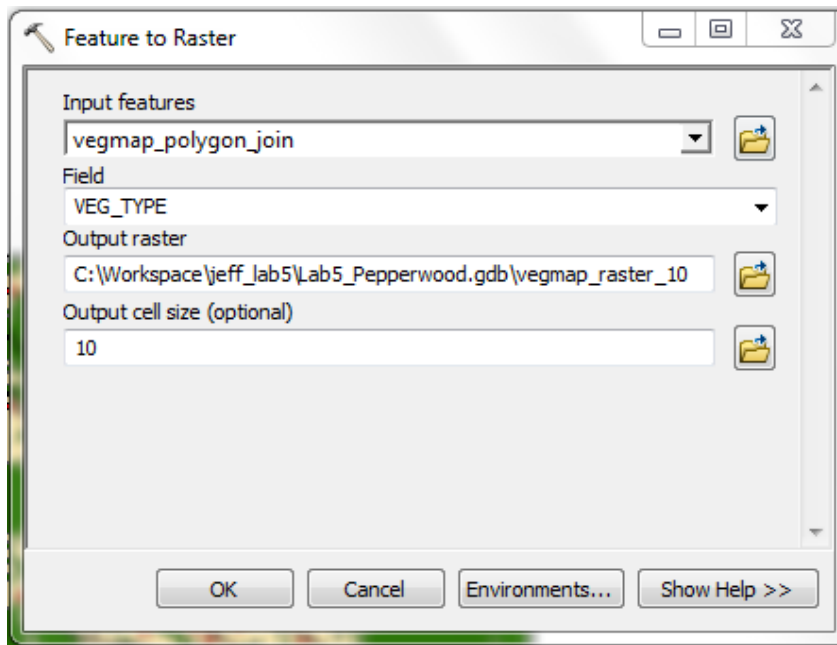
Load the *samplepts\_point* feature class. Just change the color to red.

Save your map document as *lab5\_base\_map.mxd*. Your layers should look similar to the following screen shot:



## 2.4 Polygon features to raster

We will now convert the feature classes to raster layers. In ArcToolbox, go to Conversion Tools, then To Raster, and select the Feature to Raster tool. For Input Features, select your *vegmap\_polygon\_join* feature class. For the field, we are going to use the **VEG\_TYPE** field from the feature attribute table. Direct the output raster to your *Lab5\_Pepperwood* geodatabase, and name it *vegmap\_raster\_10*. Use an output cell size of 10. The units of the cell size are in linear units of the feature class coordinate system, in this case meters (m). See the screenshot below for the setup of this tool. Click OK to run the Feature to Raster conversion tool.



The output raster *vegmap\_raster\_10* is added to the TOC.

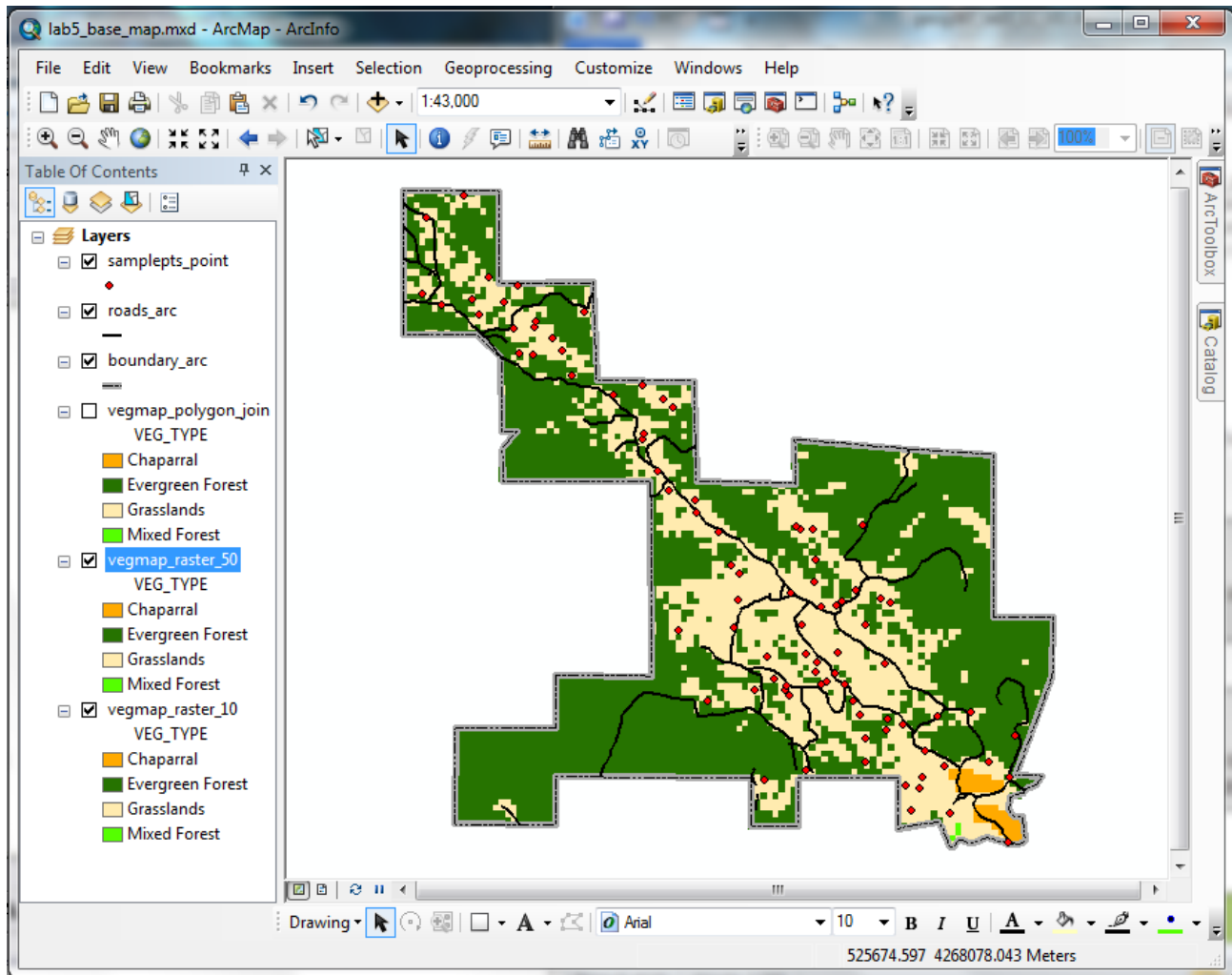
### Question 2: (4)

Go to the Layer Properties of your output *vegmap\_raster\_10* raster, then view the Source tab to answer the following:

- How many columns and rows make up the raster?
- What is the pixel type?
- What kind of raster compression method is used?
- What is the projected coordinate system?

Go to the Symbology tab for the *vegmap\_raster\_10* and select Unique Values, and **VEG\_TYPE** for the value field (**note: the default field is Value**). Use the same colors for the different classes that you used for *vegmap\_polygon*. Right-click on the *vegmap\_raster\_10* layer in the TOC and select "Save As Layer File...". **Be sure to save this file to your Lab 5 working folder**. This will save your symbology for the raster layer in a small file called a "layer file" with a .lyr file extension. We can then symbolize similarly-coded rasters by using this layer file. By the way, layer files work with feature classes as well.

Repeat the Feature to Raster tool steps above to convert *vegmap\_polygon\_join* to a raster with a 50-m cell size. Call the raster *vegmap\_raster\_50*. Go to the symbology tab for this raster and click on the "Import..." button. Browse to the layer file (.lyr extension) that you created above. Click OK. The symbolization for *vegmap\_raster\_10* and *vegmap\_raster\_50* should now be exactly the same.



Use the zoom tools and turn on and off the various vegmap polygon and raster layers and to answer the following questions:

**Question 3:** (2)

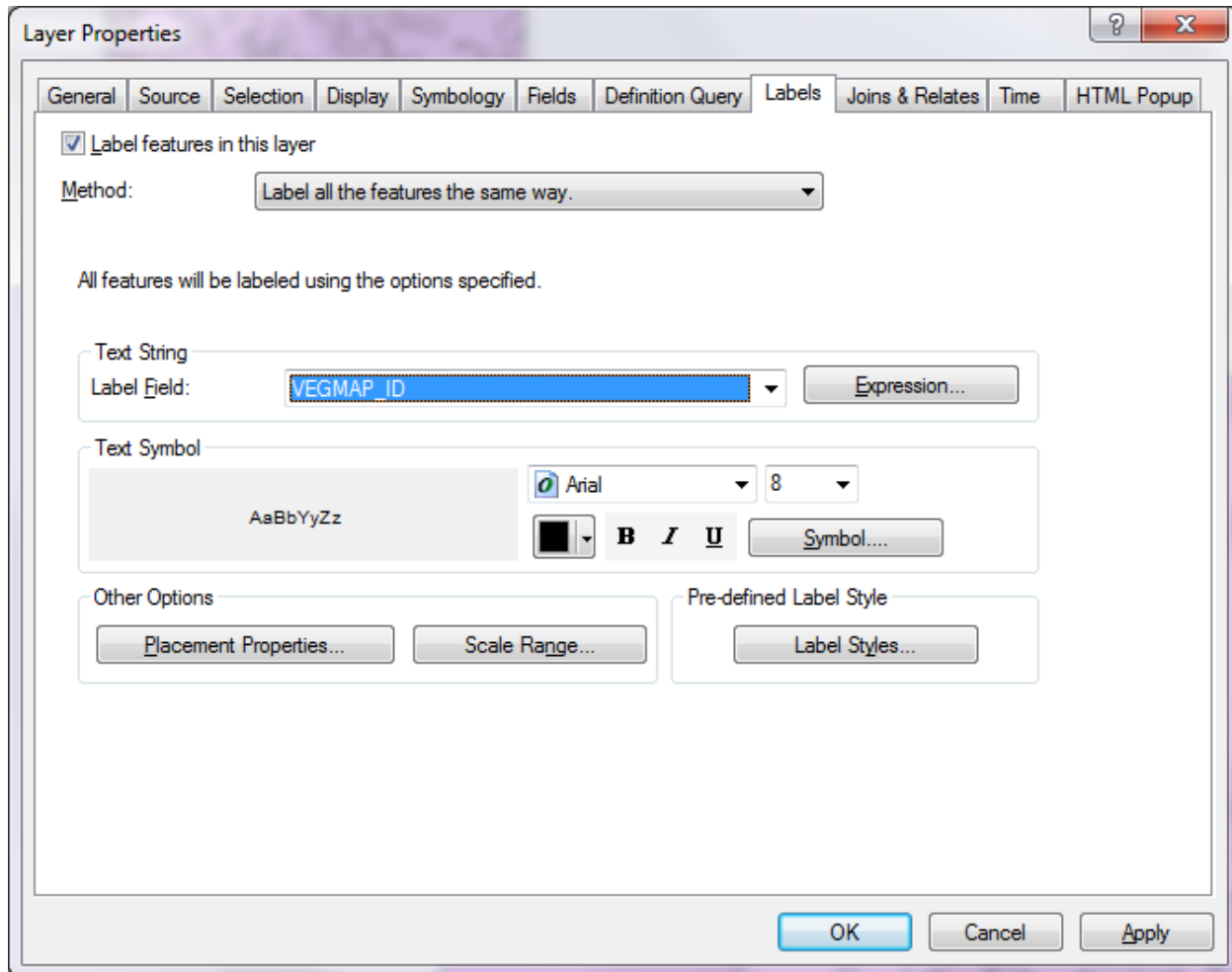
Which of the conversions (10 m or 50 m) do you think more accurately represents the original vector polygon map? Why?

**Question 4:** (2)

Are any polygons in the original vector layer eliminated due to the rasterization using 50 m resolution? How about using 10 m resolution? Why is there a difference in results?

Use File→Save As to save a copy of your map document to your Lab 5 folder. This time name the map document *lab5\_map1.mxd*. Switch the map document to layout view.

For the following map layout, you will need to add another copy of the *vegmap\_polygon\_join* layer to your TOC. Go to the symbology tab and this time just select Single Symbol symbology. Click on the colored Symbol patch and select "Hollow" in the Symbol Selector. Make sure that this hollow symbolized *vegmap\_polygon\_join* is on top of the raster layers. Select the Labels tab in the *vegmap\_polygon\_join* layer properties. Check the "Label features in this layer" box and change the label field to VEGMAP\_ID. Go ahead and turn off the drawing of all the other layers in the TOC except the hollow *vegmap\_polygon* and the two rasters.



**Map 1 (8)**

Produce a map with a layout containing two data frames--one for each raster resolution. For the 50-m resolution raster, zoom in on at least **five polygons** that have been eliminated due to rasterization. For the 10m-resolution raster, zoom to the same extent and show how these polygons were preserved. To display the correspondence between the vector and raster formats, overlay the hollow *vegmap\_polygon* on top of the raster in each map. There should be labels for each polygon displayed on your map. Remember that presentation quality also counts. Use guides and rulers to help design your map. Be sure to include a meaningful title, legend, scale bar, north arrow, credits (your name, date), etc. Label each data frame with the raster resolution. Export the map to a PDF file with 150 dpi.

Additional help: A way to make this task easier is to set up one data frame with the 50-m raster showing and with the 10-m raster turned off. Overlay the hollow vector polygons and zoom into an area with at least 5 patches that have been eliminated due to the rasterization process. Size this data frame into one half of your map layout. Next, copy and paste the data frame so that you have a duplicate copy. Move this new data frame into position in the second half of your layout. In the second data frame, turn off the 50-m raster and turn on the 10-m raster. You should now have two data frames with the same map extent (e.g., zoomed in on 5 eliminated polygons), but one with the 50-m raster underneath and the other with the 10-m raster underneath.

**Be sure to save your changes to your *lab5\_map1.mxd* map document.**



## 2.5 Line features to raster

We will now investigate the effects of converting linear vector features to raster. We will no longer work with the *lab5\_map1.mxd* map document. Open your *lab5\_base\_map.mxd* map document instead.

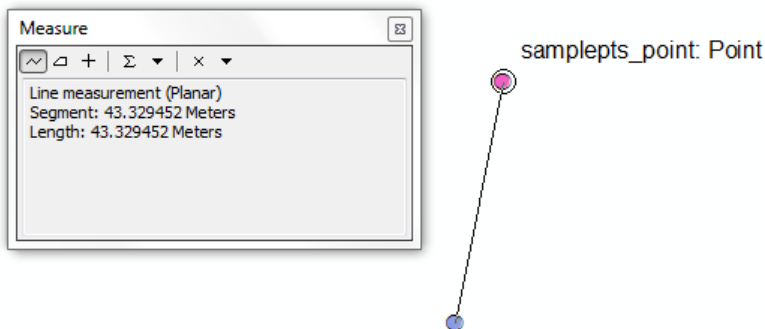
Turn off the display of all layers in the TOC except for the *roads\_arc* feature class. Use the Feature to Raster tool to convert the *roads\_arc* feature class to a 10-m raster using *ROADS\_ID* as the cell value. Choose an output raster name that allows you to remember what you did, such as *roads\_raster\_10*, and be sure to save the output to your Lab5 geodatabase. Run the tool again to create a 50-m raster, and use the same name format (i.e. *roads\_raster\_50*). If you like, you can change the symbology to display each raster in a single color by selecting Classified in the Symbology tab and changing the number of classes to 1. Arrange the layers in the TOC so that *roads\_arc* is above *roads\_raster\_10*, which is above *roads\_raster\_50*. Zoom in on areas to visually compare the original vector lines to the new raster road features. We will return to these raster road layers in the Vectorization section.

## 2.6 Point features to raster

We will now investigate the effects of converting vector point features to raster. Turn off the display of all layers in the TOC and turn on the *samplepts\_point* feature class. Use the Feature to Raster tool to convert the *samplepts\_point* to a raster layer with 10-m cell size and save to your geodatabase. Use SAMPLEPTS\_ID as the value field. Repeat the steps to create a raster with 50-m cell size. Be sure to give your output rasters descriptive names so that you can tell what you did later.


For point data, there can be a considerable amount of precision error when converting vector data layers to raster data. In this exercise, you will estimate how much the rasterization process offset the location of sample points. To do this, familiarize yourself with the Measure tool  located on the Tools toolbar. With the "Measure Line"  option selected, the Measure tool allows you measure the distance between one point and another by clicking on the first point and dragging the cursor the second point. The distance (length) measurement appears in the Measure tool window.

You can snap to the vector points by turning on the Snapping tool bar (from Customize→Toolbars) and clicking on the Point Snapping (the circle icon on the toolbar). See example below of measuring the distance between 2 vector points. I first clicked the light blue point (snapped), then positioned the tool near the purple point to snap to it. A circle is displayed around the point and a text label appears that identifies the data source of the point and the type of snapping. A linear distance for the line is shown for "Line measurement" in the Measure tool window. For the first measurement the segment length and total line length are the same. If you were to continue to click more points, then there would be a length displayed for the last line segment digitized, and a total length for the entire line composed of multiple segments. Double-click with the measure tool to clear the measurement.



Change the symbology of both raster point layers to Unique Values (any color scheme is fine). Arrange the layers in the TOC so that *samplepts\_point* is above *samplepts\_raster\_10*, which is above *samplepts\_raster\_50*.

**Question 5:** (8)

Using the measure tool and the table below, measure and record the displacement/error due to rasterizing of the *samplepts\_point* feature class for the selected sampling points. You can use a combination of select by attributes, labeling features in the *samplepts\_point* properties, and the identify tool  to find each vector point's location. Zoom in so that a single 50-m raster point feature nearly fills the data view and measure the distance from the vector point to the center of the raster cell. You will NOT be able to snap to the raster's cell center because snapping only works on vector features. However, you can at least snap to the vector point and then estimate where the cell center is located.

We have provided a blank table in a Word document for your answers to this question. Please use this table for your answers. Also, your answers do not have to be exact!

SAMPLEPTS_ID	Displacement/Error (m)	
	10 m Raster	50 m Raster
21		
22		
24		
30		
33		
40		
43		
46		

**3.0 Vectorization**

“Vectorization” is the process of converting raster feature data to a vector format. This process is usually done on raster data that encode discrete categories, such as land-cover classes. Continuous raster data (e.g. DEMs) are rarely converted to vector because raster is a superior data model for representing spatial variability of continuous phenomena. An exception to this example would be the conversion of a DEM to a TIN or vector elevation contour lines.

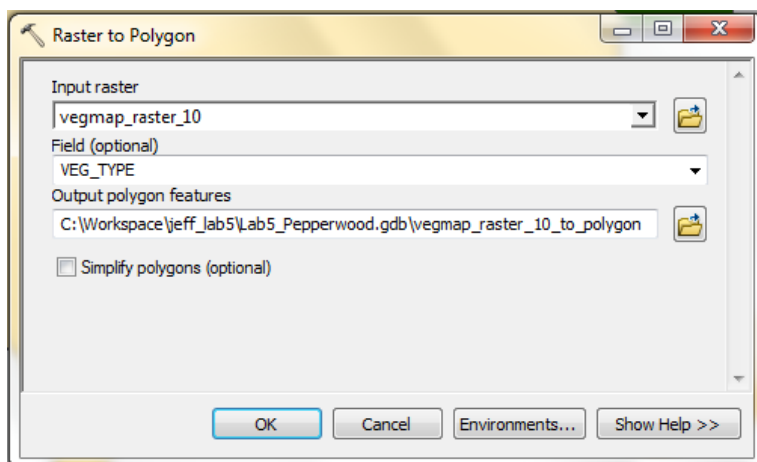
As with rasterization, there are three possible types of conversions.

- *Grouped raster features to vector polygons:* Vector polygons are built from groups of contiguous raster cells having the same cell values. Polygon outlines are created from cell borders of contiguous cells with the same value. Cells that have NoData values in the input raster will not become features in the output polygon feature.
- *Linear raster features to vector lines:* When you convert a raster representing linear features to vector lines, a vector line is created from each cell in the input raster, passing through the center of each cell. Cells that are NoData in the input raster will not become features in the output vector line feature.
- *Raster point features to vector points:* When you convert a raster representing point features to vector point, a point will be created in the output vector point layer for each cell with a value in the input raster. Each point will be positioned at the center of the cell that it represents. NoData cells will not be transformed into vector points.

To examine the effect of vectorization of rasters, we will use our output rasters from Section 2.

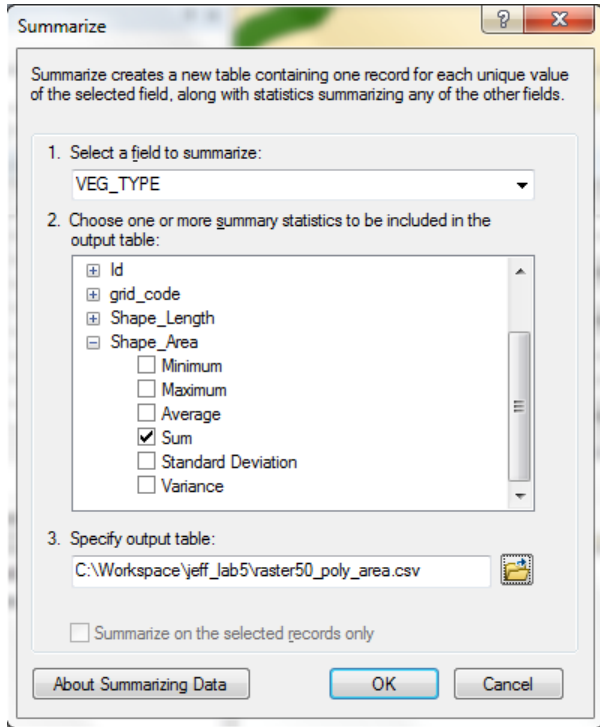
### 3.1 Raster to polygon features

Turn off the display of all layers in the TOC...you can do this quickly with a right-click on the data frame, then "Turn All Layers Off". In ArcToolbox, go to the Conversion Tools, then From Raster, select the Raster to Polygon tool. For the input raster, select your *vegmap\_raster\_10* raster layer. Use **VEG\_TYPE** as the field. *Be sure to save your output polygon feature class to the Lab5\_Pepperwood geodatabase, and give it a logical name, like "vegmap\_raster\_10\_to\_polygon".* There is an option in the tool to "Simplify polygons". This will smooth the boundaries of the output polygons, reducing their "stair-stepped" appearance resulting from the raster cells. *Uncheck the "Simplify polygons" box--we will not use this feature for now.* Click OK to run the tool. Repeat these steps for the *vegmap\_raster\_50* raster layer.

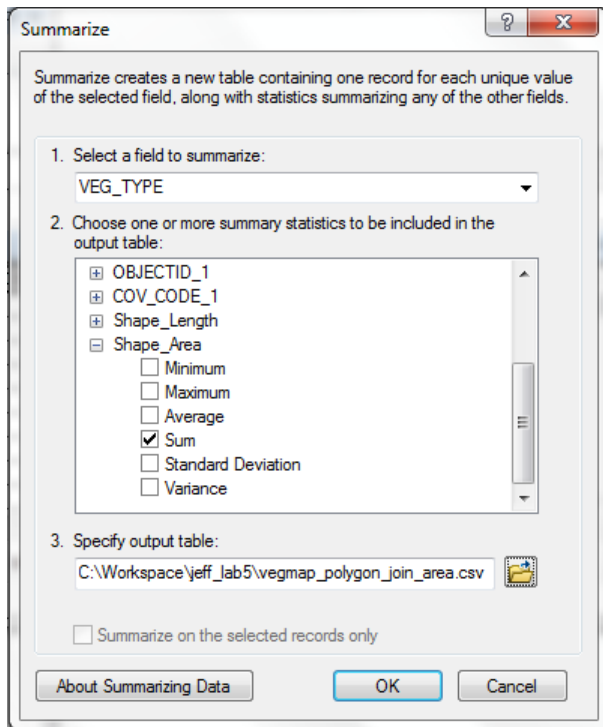


Compare the area of each land-cover type as represented in the original *vegmap\_polygon\_join* feature class and the polygon feature classes vectorized from the 10-m and 50-m rasters by completing the table below for Question 6. To summarize the areas of the four cover types in each layer, open the attribute table for each polygon feature class, then select the VEG\_TYPE field and right-click and select Summarize. Use VEG\_TYPE as the field to summarize, and then check the Sum of Shape\_Area for the summary statistic.

Save your output tables as Text Files with a .csv extension outside of the geodatabase so that you could open them in Excel (like we did in Lab 2). Give the output tables names that identify the source layer (e.g. *raster10\_poly\_area.csv*). For the vectorized feature classes, your Summarize window will look like this:



For the original vector layer, use the *vegmap\_polygon* feature class with the joined table. Your Summarize window will look like this:



Click OK, and Yes to add the table to the ArcMap document. The TOC view will change to List by Source, and the tables will be displayed in the TOC.

You can combine your area summaries from the exported CSV files into one Excel worksheet and paste your results into the table for Question 6. Alternatively, you can open each of the tables in ArcMap and copy and paste the area results into the table for Question 6. When viewing the tables in ArcMap, you can round the area values to the nearest integer by right-clicking at the top of the Sum\_Shape\_Area field, selecting Properties, clicking on the button next to Numeric, and entering 0 for the number of decimal places. **Use the template table for Question 6 in the provided Word document. Also, be sure to enter the areas in the correct source polygon layer column in the table!**

**Question 6:** (4)  
 Copy the area sum for each vegetation type from the exported CSV files from each feature class and enter the values into the answer summary table for Question 6. Your values for area should all be fairly similar regardless of the feature class. Large differences mean that there was a problem.


	<b>Vegmap_polygon</b>	<b>Vectorized Raster 10m</b>	<b>Vectorized Raster 50m</b>
<b>Chaparral</b>			
<b>Evergreen Forest</b>			
<b>Grasslands</b>			
<b>Mixed Forest</b>			

For Question 7, refer to the data in the table for Question 6. Students may have answers that vary, so interpret the data that you have in YOUR table and you will get full credit for Question 7.

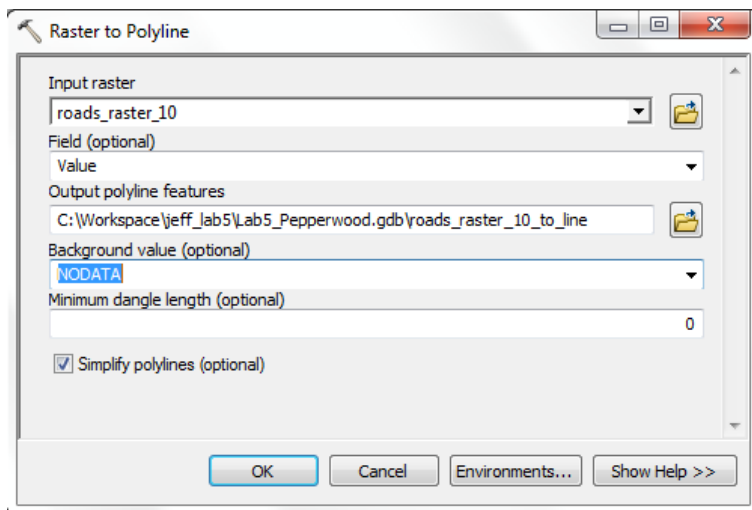
**Question 7:** (4)

- a. In the polygons vectorized from the 10-m raster, which class(es) increased in area relative to the original vegmap\_polygon feature class?
- b. In the polygons vectorized from the 10-m raster, which class(es) decreased in area relative to the original vegmap\_polygon feature class?
- c. In the polygons vectorized from the 50-m raster, which class(es) increased in area relative to the original vegmap\_polygon feature class?
- d. In the polygons vectorized from the 50-m raster, which class(es) decreased in area relative to the original vegmap\_polygon feature class?

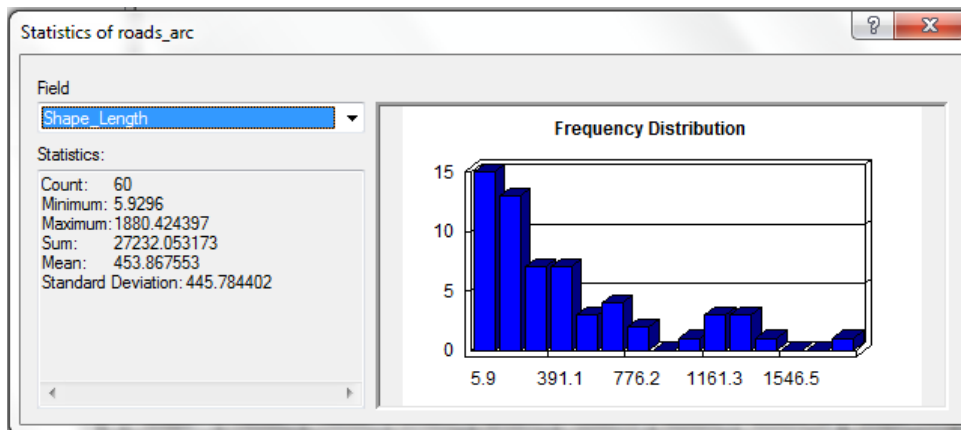
### 3.2 Raster to line features

Turn off the display of all layers in the TOC, and switch back to List by Drawing Order . We are going to vectorize the raster roads from Section 2.5 and compare the output to the original *roads\_arc* feature class.

In the Conversion toolbox in ArcToolbox, open the Raster to Polyline tool. For the Input Raster, select your rasterized roads with a 10-m cell size. For the field, leave the default at Value. For the output polyline features, create a name in your geodatabase (e.g., *roads\_raster\_10\_to\_line*). Change the background value to NODATA. This tells the tool to treat all No Data values as blank space. Leave the remaining parameters in their default settings, including a check next to Simplify polylines. Click OK. Repeat the process to vectorize the roads raster with 50-m cell size.



Summarize the total length of the entire road network for the original *roads\_arc* feature class and the vectorized versions. You can retrieve the sum of the line lengths by going to the attribute table of each feature class, then select the *Shape\_Length* field and right-click, then select Statistics (see screenshot below). For the chosen field, this tool provides general statistics (e.g., mean, min., max, sum) from the records in the feature class (e.g., the lengths of each road line segment).



**Question 8:** (3)  
 Select and copy the Sum statistic for the Shape\_Length field for each layer and enter the result into the table for Question 8. **Use the template table provided in the Word document.**

	<b>roads_arc</b>	<b>Vectorized Raster 10m</b>	<b>Vectorized Raster 50m</b>
<b>Total Road Length (m)</b>			

**Question 9:** (3)  
 In terms of length and visual properties, which vectorized raster layer is more similar to the original *roads\_arc* feature class? Why?

## 4.0 Conclusions

In this lab you began to experiment with data model transformations and their effects on the original data. In particular, rasterization cell size has profound effects on the generalization of vector features. You also found that once a vector layer has been rasterized, one cannot expect to retrieve the same shape and length of features in the original data source by re-vectorizing the raster (e.g., vectorization).

## 5.0 To turn in

- Your answers to the 9 questions in a Word document (32)
- Map 1 (pdf format) (8)

Submit electronic files via email to your instructor, with the subject "G387, Lab 5, [your last name]".

**Remember to put your last name before each file name (e.g., clark\_lab5.docx, clark\_lab5\_map1.pdf, etc.). We will deduct 0.5 points for each file not properly named.**

Credits: Text portions of this lab were developed by Dr. Ross Meentemeyer, University of North Carolina at Charlotte. Condensed version by Matthew Clark, Geography and Global Studies Department, Sonoma State University. Edited for ArcGIS 10 by Elizabeth Lotz.