**Exercise 5**

**Compiling a Spatial Database & Digitizing**

**Due Dates:**

**Lab 301/302: October 15, 2018**

**Lab 303/304: October 16, 2018**

(Total of 11 points)

1. **Introduction to the Exercise**

Despite a plethora of spatial datasets freely available nowadays, there are still instances when you need to create your own data. One way of creating a new dataset is by digitizing. In the case of our class project, we will digitize or trace the boundaries of the land use polygons from a scanned land use map that we already georeferenced in Lab 4. The scanned map cannot be used directly for analysis as it is only an image made up of color pixels with no meaningful information encoded in those pixels. Once we digitize the lines we can then convert them into topologically correct polygons that can be used for further analysis.

1. **Goals**

*For the project:*

* Digitize the land use map

*Learning goals:*

* How to create a geodatabase and new feature classes
* How to create spatial datasets through basic editing process
* How to ensure topological integrity of features

1. **Introduction to the Geodatabase**

***A geodatabase*** *serves**as a physical storage container for geographic datasets, such as feature classes, tables and rasters. It is the ‘native structure for ArcGIS’ and its improved design allows for easier and more efficient storage and access to geographic data. Read more about geodatabase design and fundamentals at* <http://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/essential-readings-about-the-geodatabase.htm>

*There are three types of geodatabases –* ***file geodatabase****,* ***personal geodatabase****, and* ***ArcSDE geodatabase****.*

**Question 1:** Carefully read the section titled “File geodatabases and personal geodatabases” in the following help document and using your own words, in few sentences, compare and contrast the file geodatabase and the personal geodatabase (1pt). <http://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/types-of-geodatabases.htm>

***A feature class*** *is a vector model that represents a “homogeneous collection of common features, each having the same spatial representation, such as points, lines, or polygons, and a common set of attribute columns, for example, a line feature class for representing road centerlines”. A feature class can only be stored in a geodatabase, but it can be converted into other formats as well. It is a file format native to ArcGIS* (<http://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/feature-class-basics.htm>).

*In this lab we will create a file geodatabase that will store the datasets related to the class project. We will learn to create new feature classes, and in the next lab we will learn how to import existing data into geodatabase from various formats.*

**Task 1: Create a geodatabase and a feature class**

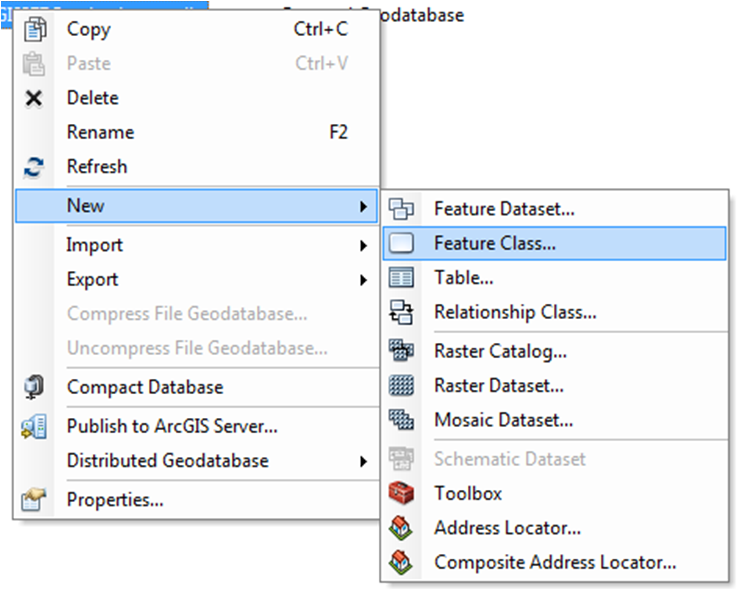
* In **ArcCatalog** browse to a folder where you would like to store the geodatabase. Right click on the name of the folder in the Catalog tree and select **New** → **File Geodatabase**. Rename it ***GEOG377Geodatabase***.
* Right click on ***GEOG377Geodatabase*** to create a new feature class within it: **New** → **Feature Class** (Figure 1).
  + Note the other types of data you can create in a geodatabase such as tables, rasters and toolboxes

Figure 1. Creating a new feature in a geodatabase.

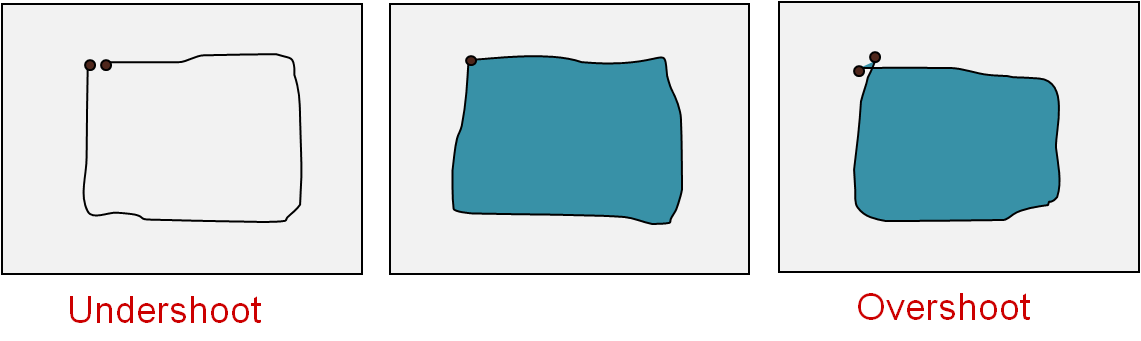
* Name the feature class *landuselines* and change the type of features stored to “**Line Features**”. Click Next. In the following window, browse to choose the coordinate system **WGS 1984 UTM Zone 18N**. In which folder would this coordinate system be, geographic or projected? Why? Click Next. Accept the default value for **XY Tolerance**. To learn more about what this number means click on **About spatial reference properties**. Click Next again to skip the screen about **Specify the database storage configuration**.
* In the last window you can add fields that will show up in the attribute table of your feature class. We don’t need to add any new fields at this time but notice that there are already two fields listed here, **OBJECTID** and **SHAPE**. Those two fields are maintained by the software and you will not be able to change them anyway. OBJECTID assigns a unique number for each feature while SHAPE specifies what type of feature it is – point, line, or polygon.
* Click **Finish** to create the new feature class named *landuselines* in ***GEOG377Geodatabase***.

*Note that you can also create a geodatabase and feature classes through the Catalog Window within ArcMap.*

1. **Digitizing**
   1. **Overview**

* ***The first step*** is to digitize or trace the boundaries between the land use types on the map using arcs (lines). This is the ***spaghetti data model*** which means it does not record the topology. When digitizing you should consider a few things. First you need to determine the scale you will be working in. For example, if you digitize at 1:20,000 scale the lines will be less accurate as if you were to digitize at 1:5,000 scale. In addition, when creating arcs you will have to decide on the location of the nodes and shape points (**vertices**). Try to create long continuous lines instead of dozens of small short segments to limit the opportunity for digitization errors. Last but not least, keep in mind that we only want to trace each line once so that we can then create polygons out of them.
  + You will be using the **Editor Toolbar** (Figure 2) to perform the digitizing and editing tasks. This toolbar allows you to start and stop an edit session, use a variety of tools and commands for creating and editing features, as well as to save your edits.
  + Some of the most common digitizing mistakes are undershoots and overshoots (Figure 3). To minimize such errors, we can use **Snapping** in ArcGIS which ensures that features connect to each other by making the feature act like a magnet. Snapping is based on the **snapping tolerance** which is a distance within which a pointer or a feature snaps to another location. You can adjust the snapping distance depending on your needs, however be careful not to set the distance either too small or too large. You can read more about snapping at <http://desktop.arcgis.com/en/arcmap/latest/manage-data/editing-fundamentals/about-snapping.htm>

Figure 2. Editor Toolbar.



**Question 2:** Explain undershoot and overshoot. How does snapping prevent these errors? What would happen if the snapping distance was set to be too small or too big when editing? (1pt)

Figure 3. Graphical illustration of undershoot and overshoot.

*For more detailed information about editing see the following help documents:*

* + ***What is editing?***<http://desktop.arcgis.com/en/arcmap/latest/manage-data/editing/what-is-editing-.htm>
  + ***Editing tutorial*** *– a full tutorial with data and directions about basics of editing* <http://desktop.arcgis.com/en/arcmap/latest/manage-data/editing-fundamentals/introduction-to-the-editing-tutorial.htm>
  + ***Quick tour of editing*** <http://desktop.arcgis.com/en/arcmap/latest/manage-data/editing/a-quick-tour-of-editing.htm>
* ***The second step*** is to assemble the lines to create area (polygon) features. This step creates a topological data layer from a spaghetti data layer. The polygons will ‘know’ their neighbors.
  + You will be using **Construct Polygons** tool on the **Advanced Editing Toolbar** (Figure 4). You have to be in an edit session in order to use tools from the Advanced Editing toolbar. For the Construct Polygons tool to be active you first need to select features that you want to convert into polygons. Before running the tool you can specify the **cluster tolerance** which is the distance when two points are considered to be coincident.

Advanced Editing toolbar

Figure 4. Advanced Editing Toolbar.

* ***The third step*** is to examine the constructed polygons for any missing or unintentional polygons. No matter how careful we are digitizing errors can happen. For example, you might have missed a line when digitizing so now you have too few polygons. Or you might have digitized a line twice in which case you now have too many polygons. Due to a larger cluster tolerance one polygon might have collapsed into two during Construct Polygons procedure so now you have too many polygons. In this step we will find these errors and fix the lines so that we can reconstruct the polygons.
* ***The last step*** is to repeat the Construct Polygons process to recreate the polygons now based on the corrected lines. You might need to go through the constructing process a few times before you find and fix all of the digitizing errors.

*Keep this big picture overview in mind while you progress through the lab.*

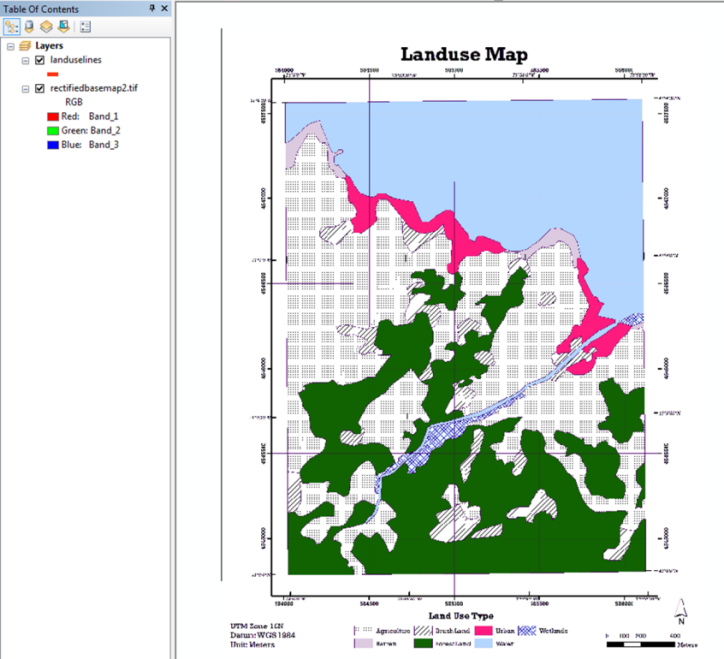


Figure 5. Landuselines and georeferenced map before digitizing.

**Digitizing the land use map**

*Now that we have an empty feature class, we will trace the boundaries off the rectified (georeferenced) image and save into this feature class.*

**Task 2: Digitize by arcs**

* Start ArcMap (or open the saved Lab4 map document if you have one), add the *landuselines* feature class you created in Task 1 and the *rectifiedbasemap.tif* you completed in Lab 4 (Figure 5).Save the map document to your flash drive by clicking **File** → **Save As** (better use relative path). Name the map document appropriately.
* Notice that even though the layer *landuselines* has a symbol you cannot see it. That is because this feature class is still empty, meaning there are no features stored within it. To check that let’s right-click on the layer name in the TOC and choose **Open Attribute Table**. In an attribute table columns are called **fields** and they represent a specific attribute. Rows are called **records** and they correspond to the actual features. Notice that there are no records in this table, hence no features to be displayed. Close the table.
* Double-click on the line symbol for your landuselines feature to change the lines to a bright color for easy viewing so that the lines stand out once you start digitizing.
* Open the **Editor Toolbar** by choosing **Customize** → **Toolbars** → **Editor** or clicking **Editor Toolbar**.
* Currently all tools on the Editor Toolbar are grayed out which means that you cannot use them. In order to create new features into *landuselines* you need to be in the **Edit Session**. Click on **Editor** → **Start Editing**.
  + You can edit multiple layers at once, as long as they are saved in the same location such as in one geodatabase. In our case we only have *landuselines* to edit but you might find yourself in a situation where there are multiple layers from a variety of sources in your map document. In that case, before the Edit Session would start the software would prompt you to choose which layer or which location you would like to edit.
* Once in an edit session you should see **Create Features** window (Figure 6) on the right side of the screen. If you cannot see it, open it by clicking **Create Features** button on the Editor Toolbar.

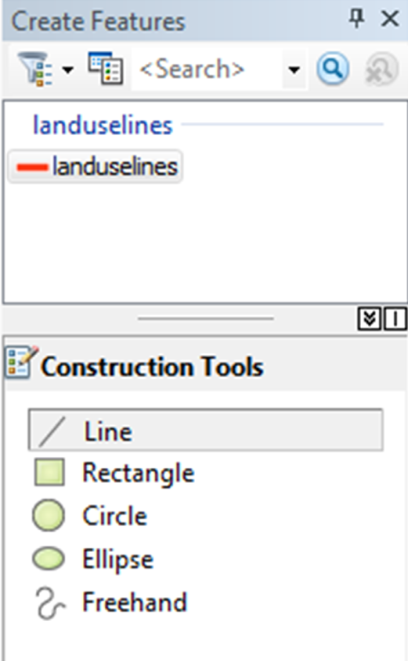


Figure 6. Create Features window.

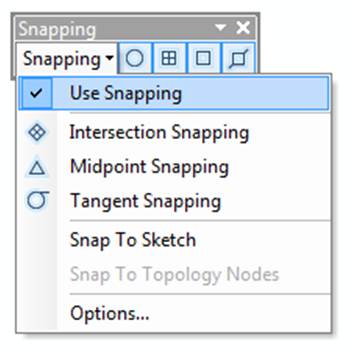
* Select *landuselines* as the feature you want to create and then select **Line** as the **Construction Tool** (Figure 6).
* Zoom to the top left corner and start tracing the edge of the land use map by left clicking in each corner. Remember that you only need to trace each boundary once; do NOT double trace the lines.
* You are creating an **arc** composed of **vertices**. Every time you click you add a new vertex.
  + Adding more vertices at a larger scale (when zoomed in closer) will create a more accurate representation of the scanned map, however it will also take more time and effort. It is important to find your balance between time and accuracy.
  + The coordinates of each of the vertices that create the arc are stored; therefore the size of your feature class increases with more added vertices. In our case, the size will not be an issue, but if you were to digitize a very large image then you might need to consider the size of the feature class as well.

Figure 7. Snapping toolbar.

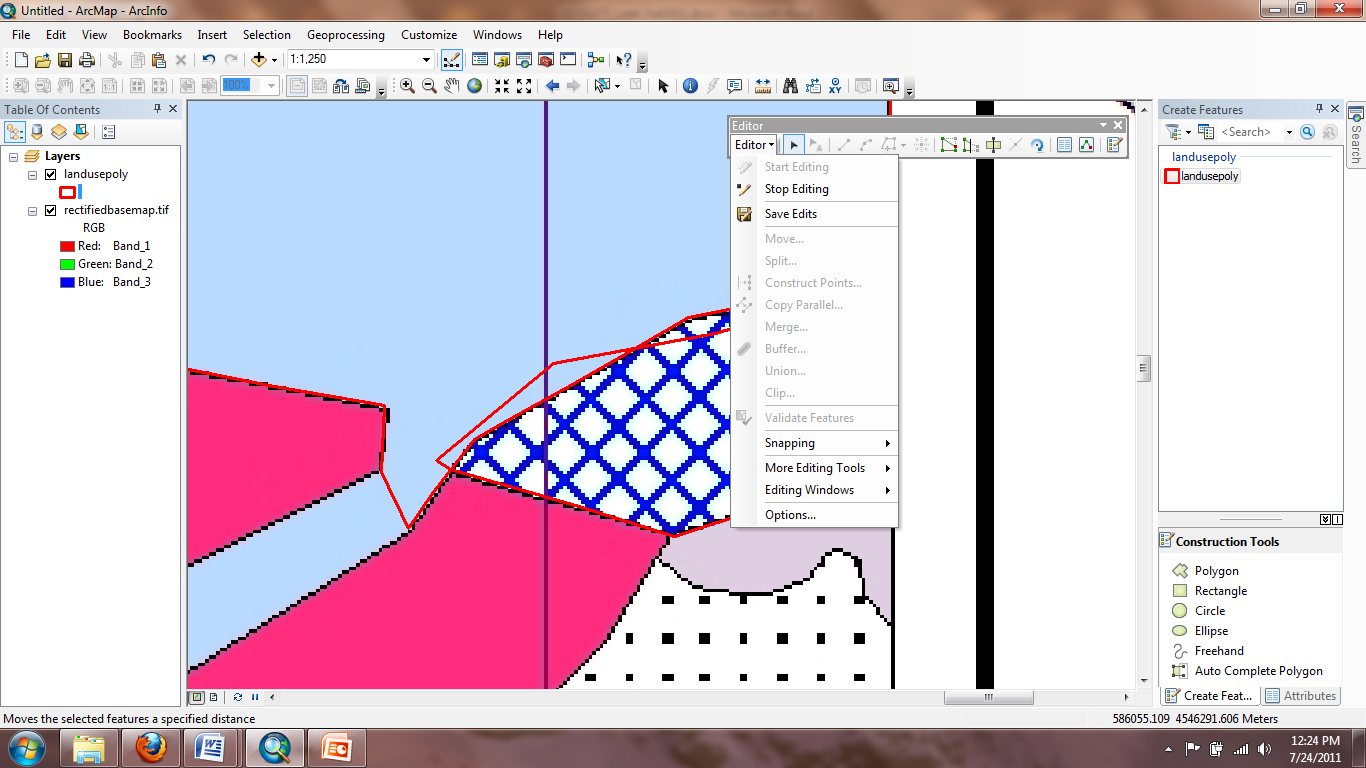
* Once you come back to the starting point of the arc outlining the map you can either double click to finish the feature, or right-click to bring up a context menu and choose **Finish Sketch**
  + Notice what other options are also available in that **right-click context menu**. This menu changes depending on where you right click. For example, if your cursor is over a vertex then the context menu will be specific to editing vertices, but if you right-click anywhere else, the options in the menu will change accordingly.
* Start tracing the boundaries between the land use types using the same method as described above.
* By now you might have noticed that when your cursor is close enough to an already digitized arc, the line acts like a magnet and the cursor jumps to it. This is **Snapping** and it will help you digitize and edit more accurately.
  + In ArcGIS 10 Snapping is enabled by default and its various settings are controlled through **Snapping Toolbar** (**Customize** → **Toolbars** → **Snapping** or **Editor** → **Snapping** → **Snapping Toolbar**) (Figure 7). Currently you can snap to points, endpoints, vertices and edges of other features. You can turn off Snapping by unchecking **Use Snapping** or you can temporarily suspend snapping by holding the Space bar when editing.
* Here are some **useful keyboard shortcuts** you can use to zoom and pan around the map. While holding the shortcut key you temporarily switch your current tool to the shortcut tool, but once you let go of the key, you are back to your original tool. Using shortcuts is a very efficient way to edit, so try these out and use them in this lab for practice.
  + **Z** – Zoom In, **X** – Zoom Out, **C** – Pan, **Ctrl Z** – Undo Last, **Ctrl Y** – Redo Last, **V** – Show Vertices

**Question 4:** When digitizing, explain what are the two considerations in choosing the number and position of vertices. (1 pt)

**Question 3**: What are the two ways you can delete an unwanted vertex while editing? (1pt)



Figure 7. Digitized landuselines.

* Digitize all lines on the map that serve as boundaries between different land use types. To ensure that your edits will be saved in case the software unexpectedly closes, click **Editor** → every 5-10 minutes or so. Saving the map document is NOT the same as Save Edits.
* If you have already created a new arc but would like to edit a vertex or two you can use the **Edit Vertices** tool or double click the feature with **Edit Tool** so that you can see the vertices. Click on a vertex and hold to drag the vertex to a different location (Figure 8). To make the changes apply to the feature, deselect it by right-clicking anywhere on the map and choosing **Finish Sketch**.

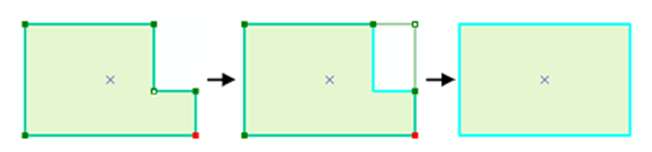
****

Figure 8. Editing vertices.

* To learn more about editing vertices and segments see the following help document <https://desktop.arcgis.com/en/arcmap/latest/manage-data/editing-existing-features/editing-vertices-and-segments.htm>
* Double check your work – make sure that all boundaries between land use types are digitized and that no line is digitized twice (Figure 9). Click **Editor** → **Stop Editing** to end the edit session (save edits if prompted).

1. **Building Polygons and Creating Topology**

*How the data is stored determines how the data can be used. Currently our land use map is stored as a line feature class; however, we are interested in the land use patches (polygons) instead of the boundaries (lines). Therefore, we need to convert the “spaghetti” lines into polygons, and we will use the* ***Construct Polygons*** *tool to do this.*

*Most digitizing errors (overshoots and undershoots) can be fixed through* ***Construct Polygons****. At the same time, topology will be constructed with that function. (Topology refers to relative spatial relationships between features, but the “spaghetti” model doesn’t contain topological information. Hereafter, each land use polygon should “know” its neighbors.)*

*Here is an outline for this process:*

1. *Create an empty polygon feature class and start an editing session*
2. *Select all lines and click Construct Polygons (on Topology Toolbar) to populate polygon feature class*
   1. *Adjust cluster tolerance if needed*
3. *Check the attribute table of the constructed polygons* 
   1. *Should have 46 polygons, check for sliver polygons*
4. *Edit lines if necessary*
   1. *Find digitizing errors and correct them*
   2. *Turn off the polygons when editing lines*
5. *Select all polygons and delete them* 
   1. *The polygon feature class has to be empty before you can use Construct Polygons again*
   2. *Turn off lines so you don’t select and delete them by accident*
6. *Go back to step 2 and repeat the process*

**Task 3: Constructing Polygons**

* First you need to create an empty polygon feature class to contain the land use polygons which will be constructed from the “spaghetti” lines. Create a new feature class in the GEOG377Geodatabase and call it *landusepoly*. Make sure to set the type of feature to **Polygon Features** and assign the same coordinate system as *landuselines*. Add *landusepoly* to your current map document.

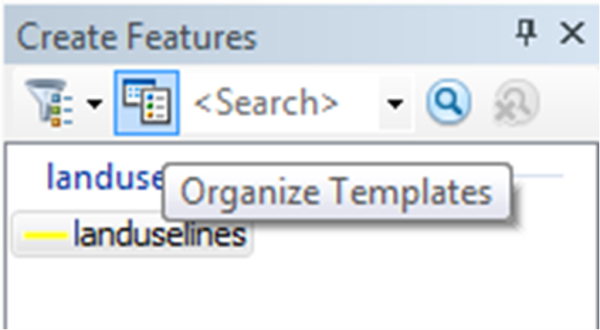


Figure 9. Organize Templates button.

* Zoom to the full extent of *landuselines* layer.
* Open the **Advanced Editing Toolbar** by choosing **Customize** → **Toolbars** → **Advanced Editing**
* Start an edit session again. Can you see both *landuselines* and *landusepoly* layers listed in the **Create Features** window? If you cannot, click on **Organize Template** button and choose **New Template** button to open the **Create New Templates Wizard** window (Figure 10). Click on *landusepoly* and create **New Template**. Close the window. Now you should be able to see both layers in the Create Features window.
* Use the **Edit tool**  from the Editor Toolbar to select all *landuselines* by drawing a box around all the features. Features will be highlighted with a cyan color (light blue).
* Click on the **Construct Polygons** button on the ***Advanced Editing*** toolbar. Choose *landusepoly* as the layer in which the new polygons will be created. Keep the default **Cluster Tolerance** and click OK (Figure 11).

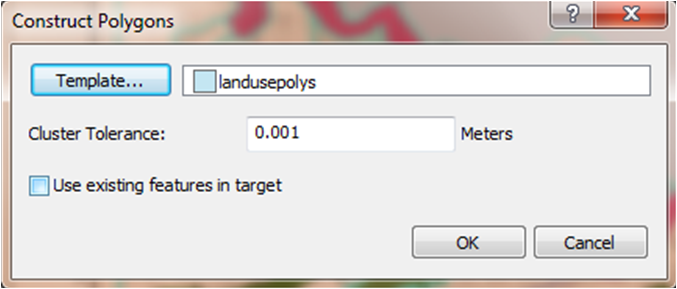


Figure 10. Construct Polygons window.

* + **Cluster Tolerance** is a distance within which two points are considered coincident. You can adjust the cluster tolerance later when you are reconstructing the polygons again.
* The new polygon features were created in the *landusepoly* from *landuselines*. You should see the polygons displayed in the data frame.

*However, no matter how careful you are in digitizing, errors are inevitable. Some errors may have been fixed through constructing features, but some may not have. However, we can always go back and fix errors and update the topology. This does not mean that you do not need to be careful when digitizing. Your care in digitizing will save you a lot of time later on.*

**Task 4: Correcting Digitizing Errors**

* Look at the attribute table of *landusepoly*.You should have 46 rows in the table representing 46 unique polygons.
  + Each feature class has an attribute table, with each **record** (row) corresponding to a feature on the map. For example, in a polygon feature class, each row corresponds to a polygon on the map and for a line feature class, each row corresponds to a polyline.
* While in the Attribute Table we can check for any unintentional polygons (**slivers**), which are very small features created through digitizing errors. Right-click the **Shape\_Area** column (the software automatically updates the area of the polygons in this field) and **Sort Ascending**. There should not be any polygons with an area of less than 2,000 square meters.
  + If you find any, select it in the table by clicking on the gray square on the left side of the table. The polygon that corresponds to this record will also be selected on the map. Zoom to it, turn off the polygon layer and make the necessary edits to the *landuselines*.
* If you don’t have 46 polygons, you made a mistake in the digitization and you need to fix it now. Compare the two layers to find major differences. Look for missing polygons that have resulted from undershooting or omitting a line when digitizing, and too many/inaccurate polygons that may have resulted from collapsing (Figure 12). **Edit the line features as necessary**. When editing lines turn off the polygon layer so that you don’t select it and start editing it by accident.

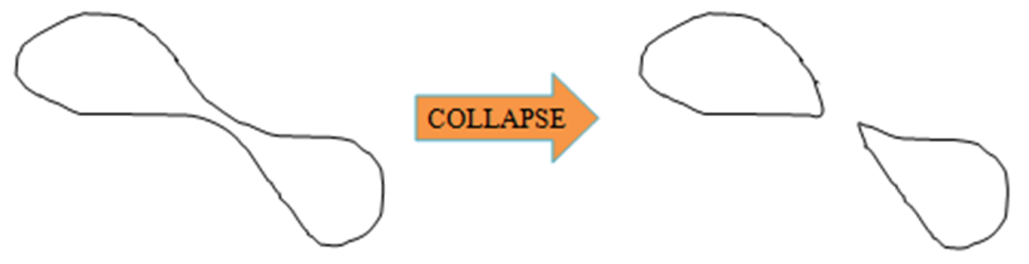


Figure 11. Collapsing polygon.

* Once you are satisfied with your edits repeat the process of constructing polygons from lines. First we should delete all of the existing polygons as we want to create brand new ones based on the edited lines.Turn off *landuselines* in TOC (to ensure that you don’t select the lines and delete them by accident), and with **Edit Tool** select all polygons. Hit Delete.

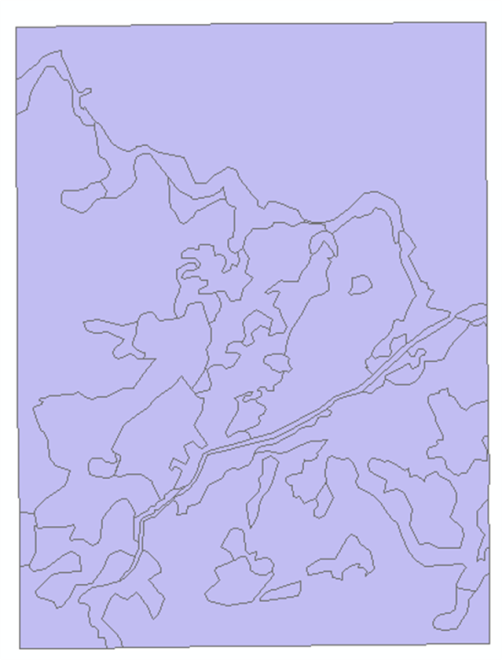


Figure 12. Constructed landuse polygons.

* Turn the landuselines back on and select them all. Click **Construct Polygons** button again. You can adjust cluster tolerance if desired.
* Look at the attribute table to make sure there are 46 polygons and visually compare the line and polygon layers again. If you don’t have 46 polygons, you will need to find the errors and edit repeat the process again. If everything looks good, you have succesfully created a topological data layer of land use polygons (Figure 13).
* Save your edits and stop editing.

**Lab 5 Assignment:**

**Question 7:** Taking information from the real world, converting it to a paper map, georeferencing and then digitizing that map is one way to input data into a Geographic Information System. Throughout this process, errors will occur. Name AND describe three of potential errors in this process (1.5pt). How do you think they will affect the final site selection? (0.5 pt)

**Question 6:** Explain the functionality of **Construct Polygons** tool by considering the differences between the two layers – *landuselines* and *landusepoly*. Discuss the type of features (points, lines, polygons), data models (spaghetti or topological) as well other characteristics of each layer. (2pt)

**Question 5:** Explain exactly how overshoots and undershoots can be fixed through Construct Polygons tool. Explain the cluster tolerance AND how should you set it? What happens if you set your cluster tolerance too small or too large? (2pt) Hint: the question is not asking about how you can manually fix these errors but rather how they are automatically fixed when you construct polygons from lines.

* + Typed answers to the questions.
* 7 questions total for 10 points (+1 for attendance)

Be sure to save your work! We will be using *landusepoly* in the remaining labs assignments. You will need *landusepoly* completed before proceeding with Lab 6.