



Geographical Information Systems Institute
Center for Geographic Analysis, Harvard University

LAB EXERCISE 6: Vector data manipulation

1. Clipping and Selecting


If you were only interested in looking at part of a larger dataset you could select or clip from your data for the entire state.

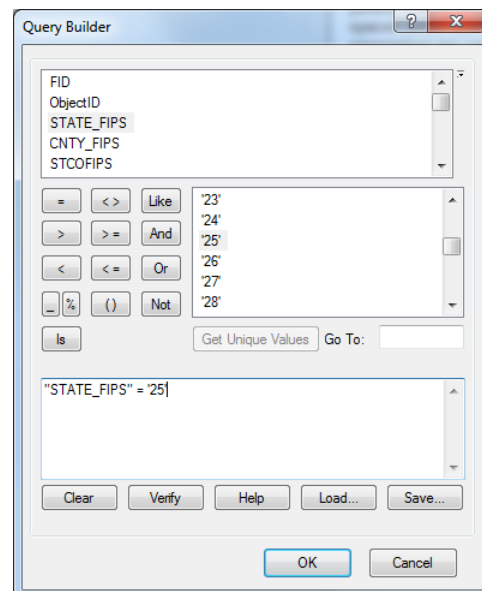
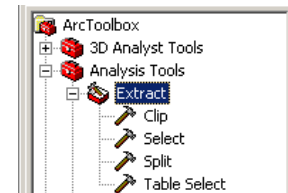
1. Copy USA_Data\Massgis to your desktop
2.  Add 2010 tract data: census_tract_2010_esri.shp that you used for the previous lab (which is in the folder USA_Data\USA_Census)

3. Activate **ArcToolbox**  and then by expanding “**Analysis Tools**” then “**Extract**” and “**Select**” you can get the **Select** wizard. Then follow the steps below.

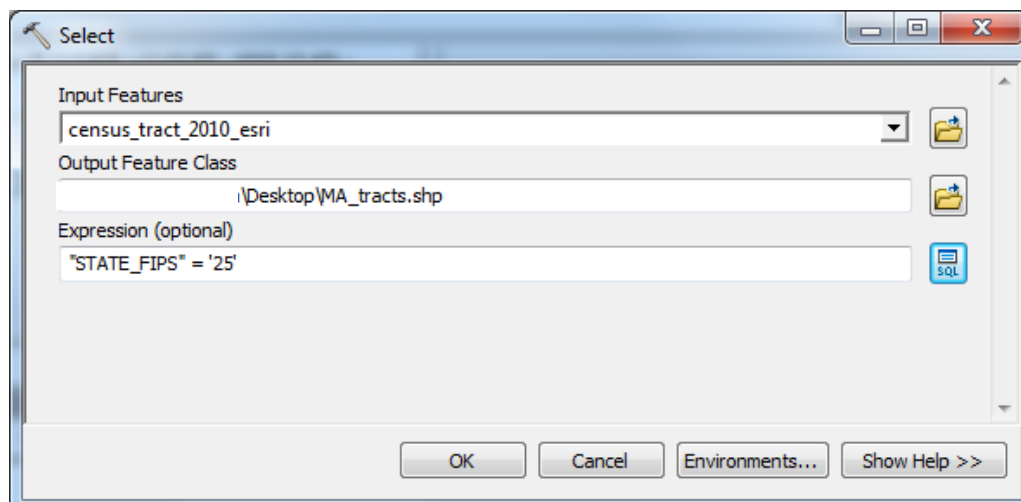
4. The input features are the data that you want select from census_tract_2010_esri.shp – the shapefile for 2010 tracts for the US; the output should be to a location where you have permission to save – the desktop

5. the **Expression** can be built by clicking on the

SQL icon  that will bring up the **Query Builder** window where you can specify that you want to select. You could use “STATE_FIPS” = ‘25’ to get Massachusetts tracts. (Double-click on the variable names, functions and values to make them appear in the Query builder window). Once you have the Select window fully specified as shown below click OK.

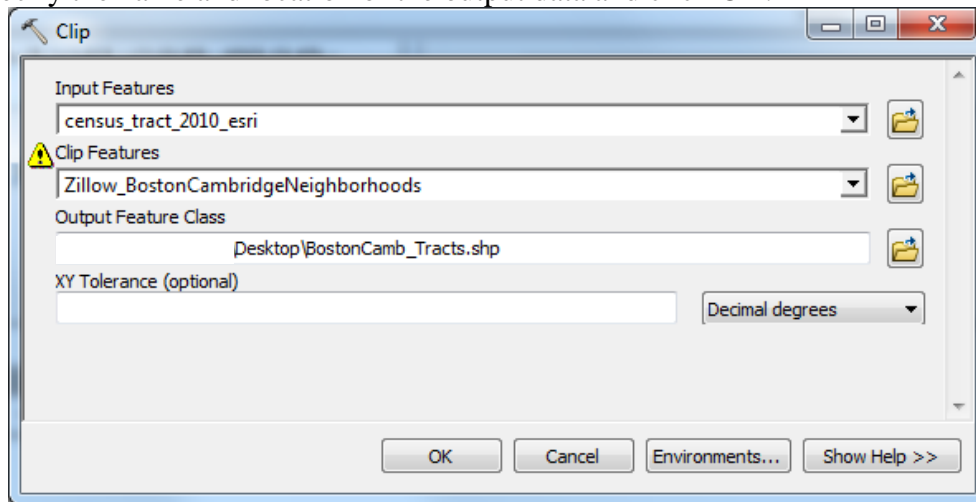


to




You can do a selection using another tool called **clip**. Clip is useful if you do not have a specific attribute like STATE_FIPS. For example if you want only the Boston and Cambridge tracts but did not have an attribute showing the cities themselves you would do the following:

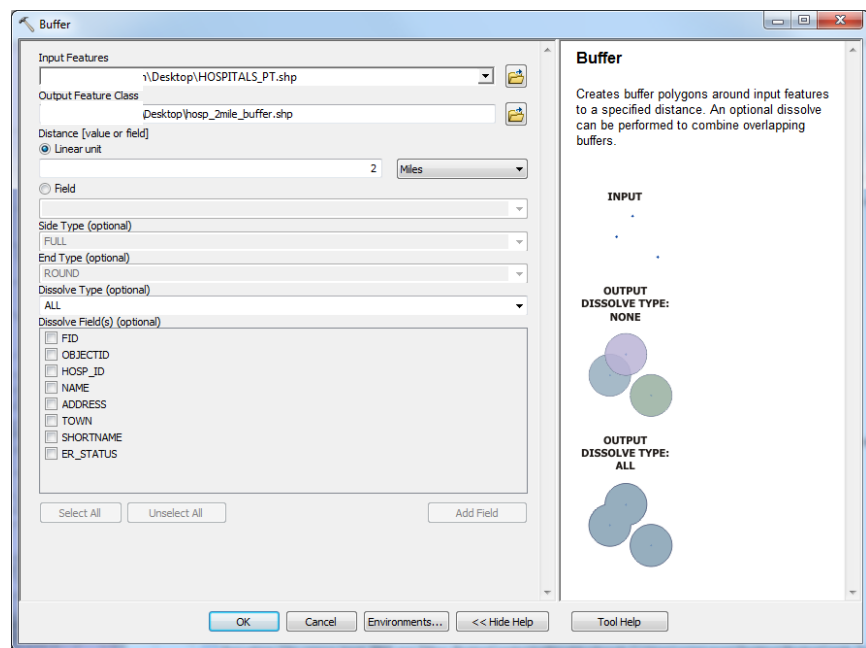
6. Either use the main menu **Geoprocessing** and then **Clip** or in ArcToolbox, expand “**Analysis Tools**,” then “**Extract**” and lastly “**Clip**” to get the **Clip** wizard.
7. The **Input Features** are the 2010 tracts shapefile, **Clip Features** are Zillow_BostonCambridgeNeighborhoods in the USA_Data folder. Note the warning (there is a datum conflict between the layers). What does this mean? (The data use different datum. You can project the layers to match their datum – if you use the shapefile census_tract_2010_esri_NAD83 as the input you will not get a warning).
8. Specify the name and location of the output data and click OK.



2. Buffering

A buffer is a region of a specified distance that surrounds certain features, and can be uniform or variable in width.

9. Add  the HOSPITALS_PT layer from the USA_data\Massgis folder
10. From the main menu **Geoprocessing**, select **Buffer** or from ArcToolbox, the Buffer wizard (expand **Analysis Tools**, **Proximity**, and finally **Buffer**).



11. In the first dialogue box, select **Hospitals** as the layer to buffer, specify the **Buffer Distance** as **2 miles**, Specify the name and location in your folder to save the new Buffer Layer, **Dissolve type** is “**All**.” Then Click OK.

Note that the dissolve specifies whether you want to remove **overlaps** between buffers which why you selected “all.” Try this again with NONE to see what the difference is.

You could now estimate the population living in these buffers around hospitals. To do this you need a way to “overlay” this spatial data with a data set that has population (like tracts, block groups or blocks from the Census for example).

3 Intersecting features

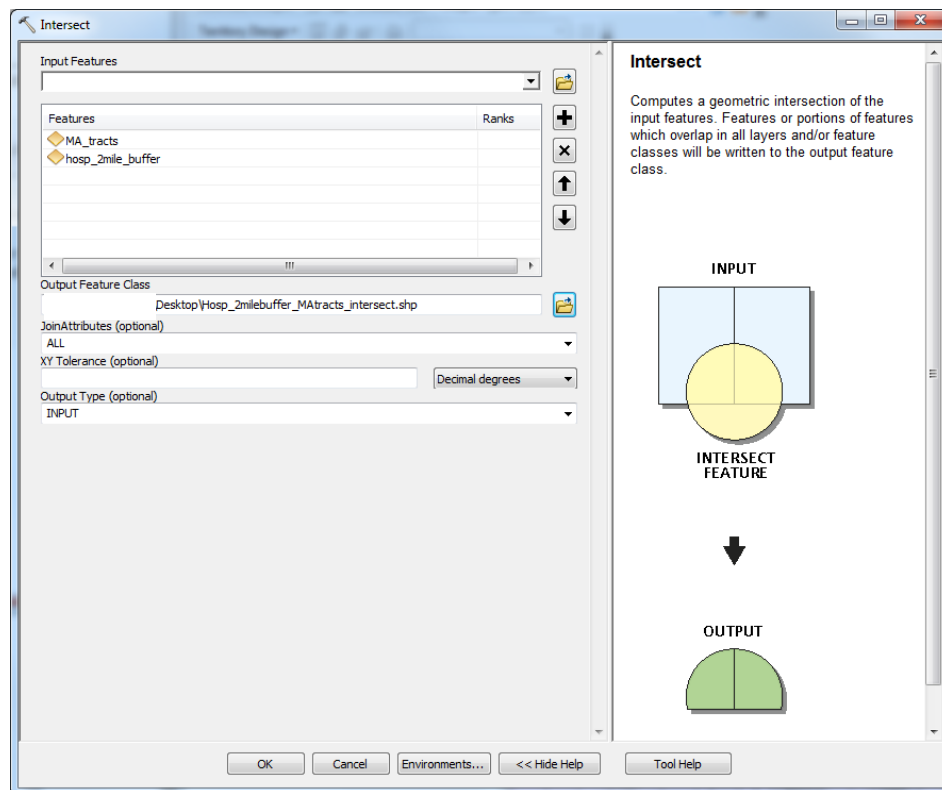
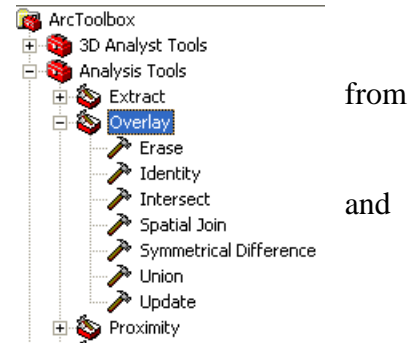
Recall that you have tracts data that you created in Section 1 using select for a State.

12. From the main menu **Geoprocessing** select **Intersect** or ArcToolbox, expand **Analysis Tools**, **Overlay** and then **Intersect** to get an Intersect wizard.

13. Select the two layers to be intersected (Hospitals buffer the tracts data), indicate where you want to save the results, keep the default settings for the rest of the inputs. Finally click OK (see figure below).

The newly “intersected” layer is added to the data frame automatically.

14. Right click on this new layer in the TOC and zoom to layer. (Also uncheck the other layers in the TOC).

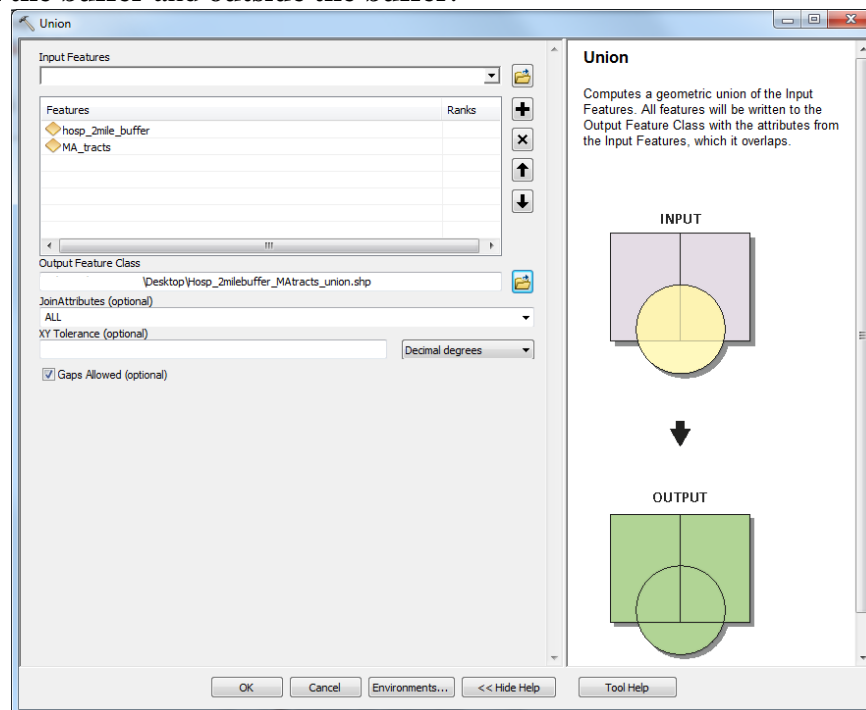


In this new data set you have the outline of the buffer as well as the data from the tracts that fall within it. You could now sum the number of households, population data, etc to find out how many people or households live within 2 miles of a hospital in Massachusetts. Note that this is an estimate. The buffer only has a part of some of the tracts. You have to figure out a way to adjust the data to account for this. One way to start correcting this would be to use the area values of the tracts within the buffers. (So you can use density as a measure for population living in that area).

4 Union of Layers

A union is “a **topological overlay** of two polygonal spatial datasets that preserves features that fall within the spatial extent of either input dataset.” In this example, you will use Union to get a new layer that includes the **all** the tracts in MA and the buffer area that you created earlier in the exercise.

15. From the main menu **Geoprocessing** select **Union** or from ArcToolbox, expand **Analysis Tools, Overlay** and **Union** to get the wizard.
16. For Input features add the two datasets (tracts and the buffer) you added in the last section, specify output and leave the other options to the default. Click OK.
17. Open the attribute table for this new layer which is automatically added to your TOC. Notice that the attribute table for this new shapefile you created has more records than the one you created in the intersection in the previous because it includes **both data for locations in the buffer and outside the buffer**.

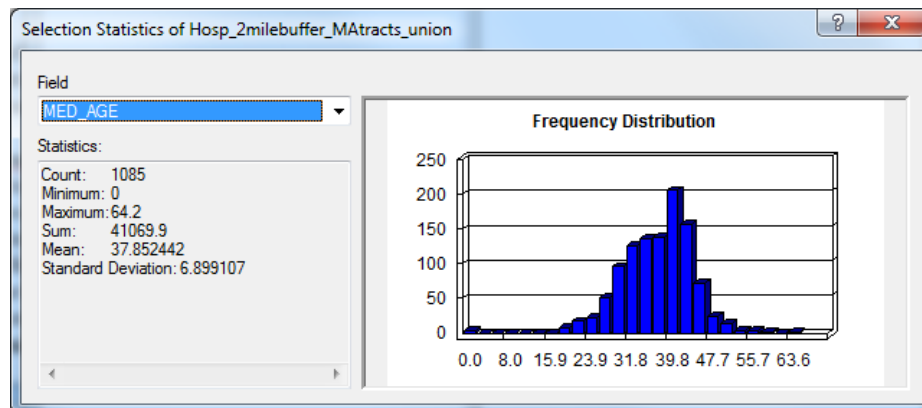
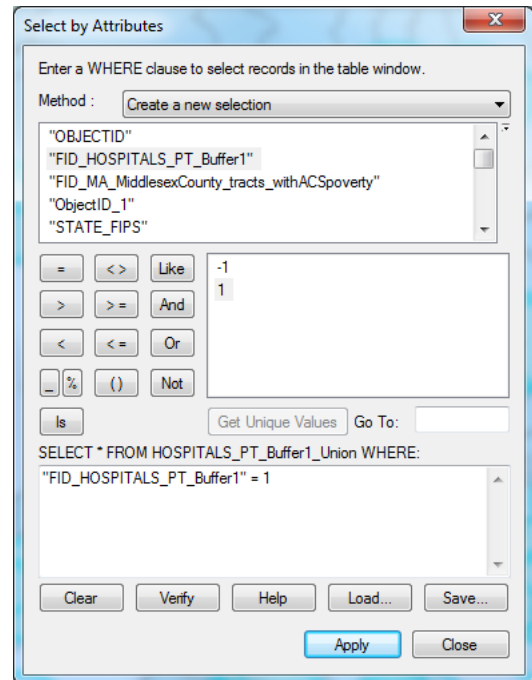


Notice that, there is a new attribute in this shapefile that was not there in the intersected shapefile that indicates whether or not a polygon is in the buffer or not.

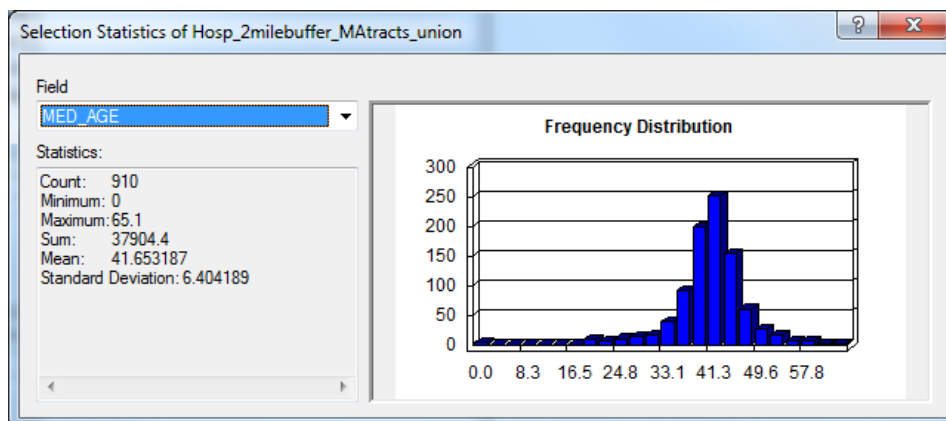
18. Select by attribute as shown in the screenshot (this may not be the name for your dataset – look for an attribute with values -1 or 0 or -1 and 1).

You can use this selection to find out if there is a difference in the types of locations that are close to Hospitals versus those that are farther from hospitals. For example you could look at the Statistics for MED_AGE for locations close to hospitals versus farther from hospitals. (See the statistics graphs below)

- 19.** On the attribute column for MED_AGE right click and select Statistics for the selection then switch the selection and calculate statistics once again.



Statistics for median age for locations within 2 mile hospital buffer



Statistics for median age for locations not in 2 mile hospital buffer