Assignment 8. Dissolve, Union, Intersect Overlay, Buffer, & Compute Centroids

Preliminary

- Review Gorr/Kurland, Tutorial 8 ('Spatial Data Processing') and Tutorial 9 ('Spatial Analysis').
- In ArcCatalog, preview the previously joined tables mdtgr2 (table, geography) and examine the metadata (especially 'Spatial Reference Information'). Then preview Phurevac (hurricane evacuation zones), making sure that its coordinate system matches that of mdtgr2.
- In ArcMap, set to store relative path names, set the appropriate coordinate system, and check the display units.

I

'Add Data': Phurevac, mdtgr2, and mdpopulation.
- List Phurevac first in the table of contents, followed by mdtgr2. Rename Phurevac 'Evacuation zones'; mdtgr2 'Miami-Dade County'; and mdpopulation 'Population'.
- Create a new group layer. Display 'Evacuation zones' first, then 'Miami-Dade County' and 'Population'.
- In 'Evacuation zones', symbolize the value field HURR_CATEG, unchecking 'All Other Values' (so that this category does not display in the table of contents).
- Adjust the colors of the symbols to reflect the gradation from Category 1 (greatest urgency: need to evacuate for weakest hurricane) to Category 5 (need to evacuate only for strongest hurricane). Rename HURR_CATEG to HURRICANE CATEGORY, and rename the hurricane category labels in the table of contents: Category 1, Category 3, Category 5. Adjust the colors on the Miami-Dade County map.

- There is no category for 'No evacuation'. While this can be easily rectified in terms of mere map display, the problem requires greater work to enable spatial analysis. How can you add a 'No evacuation' category that allows layers to be combined for spatial analysis?

II

Dissolve and union:
- Dissolve aggregates features (i.e. reduces features from more to fewer fields by dissolving the boundaries between adjacent polygons that share the same value for the specified attribute).
  - A dissolve's results depend on the particular field(s) selected for the procedure.
- Union combines all of the features of two layers.

- You can add a 'No evacuation' category to the layer discussed above and permit layers to be combined for spatial analysis by using union to combine all of the features of two layers, 'Evacuation Zones' and 'Miami-Dade County'.
  - The new layer will be named 'Miami-Dade, incl. Evacuation Zones'.
- This procedure will be easier if we first use *dissolve* to reduce the number of fields in the layer 'Evacuation zones' to only those that we need.
- We will use *dissolve* to extract from 'Evacuation zones’ the field HURR_CATEG, which will become a new layer to be named ‘Hurricane categories’. We will then union it with Miami-Dade County to make ‘Miami-Dade, incl. Evacuation Zones’.

  - *Dissolve* reduces a larger number of fields to a smaller number. Put differently, it aggregates features (e.g., census tracts).
    - ArcToolBox>right-click in white space>set Environments. Then Data Management Tools>Generalization>Dissolve.
    - Input feature class: Evacuation zones. Output feature class: C:\TEMP\...\hurrcats. In Dissolve Fields click FID HURR_CATEG. OK.
    - How can you save the new layer?
    - Rename the layer ‘Hurricane categories’. Inspect its attribute table.
    - Remove the layer ‘Evacuation zones’ (Phurevac) from the map.
  - Why is *dissolved* located in ArcToolBox under ‘Generalization’?

  - *Union* combines all of the features of two layers.
    - ArcToolBox>Analysis Tools>Overlay>Union>Input features: Miami-Dade County and Hurricane Categories>Output feature: specify path and name the file mdhc_union.
    - Rename the new layer ‘Miami-Dade, incl. Evacuation Zones’.
    - Save the new layer.
    - Uncheck the other layers.
    - Inspect the attribute table. Under symbology, uncheck ‘All other values’ so that it does not display. Symbolize the evacuation and non-evacuation zones, showing by color and/or size and shape the gradation from most to least dangerous zones. Edit the layer’s labels to say ‘Evacuation Zones’, ‘No evacuation’, ‘Category 1’, etc.
    - Remove the layers ‘Hurricane categories’ (hurrcats) and ‘Miami-Dade County’ (mdtgr2) from the map.

### III

- **Mapping and counting population across the zones:**
  - Using the layers ‘Miami-Dade, incl. Evacuation Zones’ and ‘Population’, what is the size of the residential population in each evacuation zone and in the ‘no evacuation’ zone?
  - Symbolize POP2000 (without normalizing it), clicking ‘Show class range using feature values’.
Symbolize population as follows (see Ormsby et al., chap. 6):

- Via the initial default symbolization, experimenting with the sizes and shapes of the symbols (including rank order).
- Via dot density, experimenting with ways of displaying dot density (including turning off the district boundaries ['Background’ button] and changing the order and transparency of layers).
  - Use the Identify tool, Selection tool, or other selection procedures to estimate, by examining the map and the attribute table for 'Miami-Dade, incl. Evacuation Zones’, how much the mapped symbols correspond to the population’s actual spatial distribution (see Ormsby et al. on dot placement, page 156).
  - Next, symbolize POP2000 and normalize it (try both ‘area’ and ‘percent of total’), clicking ‘Show class range using feature values’. Why doesn’t normalizing by ‘area’ make an adequate map?

Normalized or not, what has been misleading about the symbolization of POP2000 in the various forms? What could be done to improve the map?

- **Using Select by Attribute, count the residential population in each evacuation and non-evacuation zone:**
  - Find the map’s total population: open the attribute table for ‘Population’>right-click column header for POP2000>statistics.
  - Find the population in each evacuation and non-evacuation zone. Use Select by Attribute (zone) and then Select by Location (‘Population’ that ‘is contained by’ the selected zone), clearing the selected features each time.
  - Are the results different if, in Select by Location, you specify ‘Population’ that ‘intersects’ the selected zone? If so, what’s the lesson learned?
  - How can you save each location-mapped layer, including in a geodata base?
  - What’s the population in evacuation zones 1, 3, and 5 and in the ‘no evacuation’ zone? Do these add up to the map’s total population?

- There’s a much quicker, easier way to count the population in each zone, but to do it we first have to **spatially join** 'Miami-Dade, incl. Evacuation Zones’ with ‘Population’. We will do so next.

IV

- **Spatial join: joining two layers based on location:**
  - *Spatial join* is used to link a spatial feature (such as the set of evacuation and non-evacuation zones) with non-spatial data.
  - E.g., to map data for poverty levels, join the poverty data to spatial data. Begin the join by right-clicking the spatial data.
Right-click ‘Miami-Dade, incl. Evacuation Zones’ and do a spatial join with ‘Population’. The spatial join will allow you both to compute the statistics that you did above and to map them.

Do the spatial join:
- Click on ‘Miami-Dade, incl. Evacuation Zones’ in the table of contents>Joins and Relates>‘Join’ with the layer ‘Population’.
- In response to ‘What do you want to join to this layer?’ enter ‘Join data from another layer based on spatial location’.
- Choose to give each population data point all the attributes of the polygon that it falls inside.
- Specify the file path and name.
- Inspect the new layer’s attribute table.
- Rename the new layer ‘Miami-Dade, incl. Evacuation Zones and Population’.
- Remove ‘Miami-Dade, incl. Evacuation Zones’ from the map.

Here’s how to do the quick and easy computation—of a quantitative or count field summarized by a categorical field—as mentioned before. It will produce a statistical output table.
- Open the new layer’s attribute table.
- Right-click the column header of the categorical field of interest, ‘HURR_CATEG’.
- Click to sum ‘the attributes to be summarized’ (i.e. sum ‘Population’ by ‘HURR_CATEG’).
- Specify the path and name the output table ‘evacpop’.
- Click ‘Source’ so that the output table displays.
- Open and inspect the table. How much population falls within each evacuation and non-evacuation zone?
- Create a graph that displays population size by zone (layer=evacpop; field=POP2000).

You can do such computations based on any spatial join. At any one time you can summarize more than one quantitative or count field by a categorical field.

Another exercise with spatial join:
- Uncheck all layers except ‘Miami-Dade, incl Evacuation Zones and Population’.
- ‘Add data’: Lmobhomes. Rename the layer ‘Mobile home parks’.
  Inspect its attribute table.
- Symbolize Lmobhomes.
- Use Select by Attributes and Select by Location to find the number of mobile homes in each evacuation and non-evacuation zone. Clear selection.
- Next, right-click ‘Miami-Dade, incl. Evacuation Zones’ and do a spatial join with ‘Mobile home parks’, giving each mobile-home data point all the attributes of the polygon that it falls inside.
- Make an output table that displays the number of mobile homes by each zone.
- Click ‘Source’ so that the output table displays.
o Make a graph showing the number of mobile home parks by evacuation and non-evacuation zone.
o Remove ‘Mobile home parks’ from the map.

VI

- **Intersect overlay:**
o What is the total length of highways?
o Using 'Miami-Dade, incl. Evacuation Zones and Population', what total length of highways falls within the evacuation zones as a whole?
o Select by location. Try various selection criteria (e.g., intersect; completely within; contained by) and compare the mapped and statistical results produced each way. Which way is most precise?
o Next, disaggregate the total length of highways into amounts for each evacuation and non-evacuation zone. Do so using **intersect overlay**.
o Finally, use a spatial join to link evacuation zones to ‘Highways’. Examine the length of highways for the evacuation zones in general and for each evacuation zone. Do the results differ from those based on **intersect overlay**.
o Do not remove Ahighway from the map.

o When should you use **spatial join** and when should you use **intersect overlay**?
  - Intersect overlay may not work for quantitative-discrete data (e.g., population size) or count data (e.g., number of HIV/AIDS cases). It generally requires quantitative-continuous data (e.g., length of roads).
  - Intersect overlay does not change the characteristics of the data layers.
  - Thus, intersect overlay does not disaggregate the output data into, say, the specific evacuation categories. Nonetheless, this disaggregation can be obtained within the attribute table via ‘Summarize’.
o Why are **intersect** and **union** located in ArcToolBox under ‘Overlay’?

VII

- **Computing and mapping centroids:**
o Converting polygon features (e.g., property parcels) into point features may make a map’s features more accurate. It may do so by keeping features from being erroneously multi-counted because their locations sometimes spill cross spatial boundaries (see [http://www.gsd.harvard.edu/gis/manual/vector/](http://www.gsd.harvard.edu/gis/manual/vector/)).
o **Use centroids only if the features are roughly the same size and shape** (see Mitchell, *The ESRI Guide to GIS Analysis, vol. 2, Spatial Measurements & Statistics*, 184-85).

- Here’s how to compute and map centroids, using the layer ‘Miami-Dade, incl. Evacuation Zones’.
To obtain the population points we have to compute centroids (see the web site’s document ‘Computing Centroids’).

Open the attribute table of ‘Miami-Dade, incl. Evacuation Zones and Population’ and create an ‘X’ field and a ‘Y’ field (field type= double; precision=10; scale=4).

- Compute the X (longitude, ‘northing’) and Y (latitude, ‘easting’) centroids for pop2000 (see the class document on ‘Computing Centroids’). Do so as follows:
  - If the attribute table does not already have a column for the ‘X’ and a column for the ‘Y’ coordinates, add these fields (type=double; leave the other categories blank).
    - Recall: X=longitude (‘northing’), Y=latitude (‘easting’).
  - The ‘X’ and ‘Y’ columns are initially populated with 0’s.
  - Do the following for the ‘X’ column:
    - Right-click column.
    - Click Calculate Values (outside of an edit session: click ‘Yes’ to the prompt).
    - Click Advanced.
    - In top text box type the following (or load the ‘CentroidsX_calc’ formula from the class web site; click ‘Load’ in the dialogue box):
      ```vba
      Dim dblX As Double
      Dim pArea As IArea
      Set pArea = [Shape]
      dblX = pArea.Centroid.X
      ```
    - In the bottom text box type the following: dblx
    - Click OK.
  - Repeat for the ‘Y’ column, changing the X’s to Y’s (or load the ‘CentroidsY_calc’ formula).
  - The calculations have assigned centroid values to the ‘X’ and ‘Y’ columns.
    - Are the computed centroids correctly? Do some research to find out what the geographic coordinates should be for Miami-Dade County.

- **From within the attribute table**, click Options>Export (i.e. not from a layer in the table of contents but from within the attribute table itself). Then list the path for the centroids output table and name it ‘centroids’.

- Click the ‘Source’ tab at the bottom of the table of contents so that the centroids layer displays in the table of contents.
- Click Tools>Add XY Data and import the coordinate system from ‘Miami-Dade, incl. Evacuation Zones’ so that the centroids display on the map.
- For the moment, experiment with symbolization:
Select POP2000: Resymbolize the layer to display effectively (e.g., perhaps symbol=Circle1, size=2.0, choose a color).
Try other symbols (e.g., symbol=Circle1, black, graduated symbol 1 to 5).
Note: the most appropriate symbolization varies according to the extent of the display (i.e. full extent versus zooming into various scales).

VIII

**Buffer analysis:**
- Create a buffer that contains hurricane category 3.
  - Select by Attribute: hurricane category 3.
  - Click the buffer tool.
  - Buffer: Miami-Dade, incl. Evacuation Zones and Population>Next>At a specified distance: 5 miles>Based on a distance from an attribute: HURR_CATEG>Distance units: kilometers>Next>Barrier output type: dissolve barriers between ‘Yes’. Create buffers so they are ‘only outside the polygon(s)’. Name path and file. Finish.
  - Rename the buffer layer ‘Buffer zone: hurricane category 3’. Adjust the buffer color if desired, and display the layer at 30%.
- Create another buffer for hurricane category 3, but this time click ‘At a specified distance: 5 miles’ but do not click ‘Based on an attribute’. What’s the difference in the result? When should you use one or the other option?
  - Using either one of the buffer layers, map the population within the buffer: Select by location (‘Population’) that is contained by the buffer. How many residents does the buffer contain?
  - If you needed to combine the population and the buffer zone as a separate data file, how would you do so?
  - Next, select the length of highways within the buffer. What is the length of the highways within it?
  - If you needed to combine the highway layer with the buffer layer, how would you do so? How would you save the combined layers?

IX

**Reduce the number of displayed fields without permanently changing the data set:**
- If an attribute table has more fields than you want to display but you don’t want to eliminate any fields from the attribute table, how can you reduce the number of displayed fields?
  - Right-click ‘HURRICANE CATEGORY’>Properties>Fields.
  - Uncheck the Visible box for each field you don’t want to display.
  - OK.
**Change the names of fields in a table:**
- If you want to change the names of fields in a table: Right-click 'HURRICANE CATEGORY'>Properties>Fields.
- Click on each field whose name you want to change and type an *alias* in place of the original name.
- OK.

**Review:**
- Why are procedures such as clip, union, intersect overlay, spatial join, dissolve, buffer, and select by location or attribute called 'spatial analysis'?
- To summarize what you’ve done, describe what each of these procedures does and when you would use them.