Union: A polygon-on-polygon overlay method that preserves all features from the input layers.

Update: A GIS operation that replaces the input layer with the update layer and its features.

REVIEW QUESTIONS

1. Define a buffer zone.
2. Describe *three* variations in buffering.
3. Provide an application example of buffering from your discipline.
4. Describe a point-in-polygon overlay operation.
5. A line-in-polygon operation produces a line layer, which typically has more records (features) than the input line layer. Why?
6. Provide an example of a polygon-on-polygon overlay operation from your discipline.
7. Describe a scenario in which Intersect is preferred over Union for an overlay operation.
8. Suppose the input layer shows a county and the overlay layer shows a national forest. Part of the county overlaps the national forest. We can express the output of an Intersect operation as [county] AND [national forest]. How can you express the outputs of a Union operation and an Identity operation?
9. Define slivers from an overlay operation.
10. What is a minimum mapping unit? And, how can a minimum mapping unit be used to deal with the sliver problem?
11. Although many slivers from an overlay operation represent inaccuracies in the digitized boundaries, they can also represent the inaccuracies of attribute data (i.e., identification errors). Provide an example for the latter case.
12. Both nearest neighbor analysis and Moran's I can apply to point features. How do they differ in terms of input data?
13. Explain spatial autocorrelation in your own words.
14. Both Moran's I and the G-statistic have the global (general) and local versions. How do these two versions differ in terms of pattern analysis?
15. The local G-statistic can be used as a tool for hot spot analysis. Why?
16. What does a Dissolve operation accomplish?
17. Suppose you have downloaded a vegetation map from the Internet. But the map is much larger than your study area. Describe the steps you will follow to get the vegetation map for your study area.
18. Suppose you need a map showing toxic waste sites in your county. You have downloaded a shapefile from the Environmental Protection Agency (EPA) website that shows toxic waste sites in every county of your state. What kind of operation will you use on the EPA map so that you can get only the county you need?

APPLICATIONS: VECTOR DATA ANALYSIS

This applications section has four tasks. Task 1 covers the basic tools of vector data analysis including Buffer, Overlay, and Select. Because ArcGIS does not automatically update the area and perimeter values of an overlay output in shapefile format, Task 1 also uses the Calculate Geometry tool to update the area and perimeter values. Task 2 covers overlay operations with multicomponent polygons. Task 3 introduces two different options for measuring distances between point and line features. Task 4 deals with spatial autocorrelation.

Task 1: Perform Buffering and Overlay

What you need: shapefiles of *landuse, soils,* and *sewers.*

Task 1 simulates GIS analysis for a real-world project. The task is to find a suitable site for a new university aquaculture lab by using the following selection criteria:

- Preferred land use is brushland (i.e., LUCODE = 300 in *landuse.shp*).
- Choose soil types suitable for development (i.e., SUIT >= 2 in *soils.shp*).
- Site must be within 300 meters of sewer lines.

1. Start ArcCatalog, and connect to the Chapter 11 database. Launch ArcMap. Add *sewers.shp, soils.shp,* and *landuse.shp* to Layers, and rename Layers Task 1. All three shapefiles are measured in meters.

2. First buffer *sewers.* Click the Show/Hide ArcToolbox Window button to open the ArcToolbox window. Select Environments from the context menu of ArcToolbox, and set the Chapter 11 database to be the current workspace. Double-click the Buffer tool in the Analysis Tools/Proximity toolset. In the Buffer dialog, select *sewers* for the input features, enter *sewerbuf.shp* for the output feature class, enter 300 (meters) for the distance, select ALL for the dissolve type, and click OK. Open the attribute table of *sewerbuf.* The table has only one record for the dissolved buffer zone.

Q1. What is the definition of Side Type in the Buffer dialog?

3. Next overlay *soils, landuse,* and *sewerbuf.* Double-click the Intersect tool in the Analysis Tools/Overlay toolset. Select *soils, landuse,* and *sewerbuf* for the input features. (If you are using the ArcInfo version of ArcGIS, you can input all three layers; otherwise, overlay two layers at a time.) Enter *final.shp* for the output feature class. Click OK to run the operation.

Q2. How is the XY Tolerance defined in the Intersect dialog?

Q3. How many records does *final* have?

4. The final step is to select from *final* those polygons that meet the first two criteria. Double-click the Select tool in the Analysis Tools/Extract toolset. Select *final* for the input features, name the output feature class *sites.shp,* and click the SQL button for Expression. In the Query Builder dialog, enter the following expression in the expression box: "SUIT" >= 2 AND "LUCODE" = 300. Click OK to dismiss the dialogs.

Q4. How many parcels are included in *sites?*

5. Open the attribute table of *sites.* Notice that the table contains two sets of area and perimeter. Moreover, each field contains duplicate values. This is because ArcGIS Desktop does not automatically update the area and perimeter values of the output shapefile. An easy option to get the updated values is to convert *sites.shp* to a geodatabase feature class. The feature class will have the updated values in the fields shape_area and shape_length. For this task, you will use a simple tool to perform the update. Close the attribute table of *sites.*

6. Double-click the Add Field tool in the Data Management Tools/Fields toolset. Select *sites* for the input table, enter Shape_Area for the field name, select Double for the field type, enter 11 for the field precision, enter 3 for the field scale, and click OK. Use the same tool and the same field definition to add Shape_Leng as a new field to *sites.*

7. Right-click Shape_Area in the attribute table of *sites* and select Calculate Geometry. Click Yes to do a calculate outside of an edit session. In the Calculate Geometry dialog, select Area for the property and square meters for units. Click OK. Shape_Area is now populated with correct area values.

8. Right-click Shape_Leng and select Calculate Geometry. Click Yes in the next dialog. In the Calculate Geometry dialog, select Perimeter for the property and Meters for units. Click

OK. Shape_Leng is now populated with correct perimeter values.

Q5. What is the sum of Shape_Area values in *sites.shp*?

Task 2: Overlay Multicomponent Polygons

What you need: *boise_fire, fire1986,* and *fire1992,* three feature classes in the *regions* feature dataset of *boise_fire.mdb. boise_fire* records forest fires in the Boise National Forest from 1908 to 1996, *fire1986* fires in 1986, and *fire1992* fires in 1992.

Task 2 lets you use multipart polygon features (Chapter 3) in overlay operations. Both *fire1986* and *fire1992* are polygon layers derived from *boise_fire.* An overlay of multipart polygons results in an output with fewer features (records), thus simplifying the data management task.

1. Insert a new data frame in ArcMap, and rename it Task 2. Add the feature dataset *regions* to Task 2. Open the attribute table of *boise_fire.* Historical fires are recorded by year in YEAR1 to YEAR6 and by name in NAME1 to NAME6. The multiple fields for year and name are necessary because a polygon can have multiple fires in the past. Open the attribute table of *fire1986.* It has only one record, although the layer actually contains seven simple polygons. The same is true with *fire1992.*
2. First union *fire1986* and *fire1992* in an overlay operation. Double-click the Union tool in the Analysis Tools/Overlay toolset. Select *fire1986* and *fire1992* for the input features, and enter *fire_union* for the output feature class in the *regions* feature dataset. Click OK to run the operation. Open the attribute table of *fire_union.*

Q6. Explain what each record in *fire_union* represents.

3. Next intersect *fire1986* and *fire1992.* Double-click the Intersect tool in the Analysis Tools/Overlay toolset. Select *fire1986* and *fire1992* for the input features, and enter *fire_intersect* for the output feature class. Click OK to run the operation.

Q7. Explain what the single record in *fire_intersect* represents.

Task 3: Measure Distances between Points and Lines

What you need: *deer.shp* and *edge.shp.*

Task 3 asks you to measure each deer location in *deer.shp* to its closest old-growth/clear-cut edge in *edge.shp.* There are two options to complete the task. The first option is to use the join data by location method, which is used for this task. The other option is to use the Near tool in the Analysis Tools/Proximity toolset (available to the ArcInfo version), which is explained at the end of the task.

1. Insert a new data frame in ArcMap, and rename it Task 3. Add *deer.shp* and *edge.shp* to Task 3.
2. Right-click *deer,* point to Joins and Relates, and select Join. Click the first dropdown arrow in the Join Data dialog, and select to join data from another layer based on spatial location. Make sure that *edge* is the layer to join to *deer.* Click the radio button stating that each point will be given all the attributes of the line that is closest to it, and a distance field showing how close that line is. Specify *deer_edge.shp* for the output shapefile. Click OK to run the operation.
3. Right-click *deer_edge* and open its attribute table. The field to the far right of the table is Distance, which lists for each deer location the distance to its closest edge.

Q8. How many deer locations are within 50 meters of their closest edge?

4. The Near tool uses a dialog to get the input features (*deer*), the near features (*edge*), and the optional search radius. The Near tool does not have the option of creating a new output data set. The result of the analysis is stored in the attribute table of *deer,* which may also include location

(x- and y-coordinates of the near feature) and angle (the angle between the input and near features), in addition to distance.

Task 4: Compute General and Local G-Statistics

What you need: *adabg00.shp,* a shapefile containing block groups from Census 2000 for Ada County, Idaho.

In Task 4, you will first determine if a spatial clustering of Latino population exists in Ada County. Then you will test to see if any local "hot spots" of Latino population exist in the county.

1. Insert a new data frame in ArcMap. Rename the new data frame Task 4, and add *adabg00.shp* to Task 4.
2. Right-click *adabg00,* and select Properties. On the Symbology tab, choose Quantities/Graduated colors to display the field values of Latino. Zoom in to the top center of the map, where Boise is located, and examine the spatial distribution of Latino population. The large block group to the southwest has a high percentage of Latino population (11%) but the block group's population is just slightly over 4600. The visual dominance of large area units is a shortcoming of the choropleth map.

Q9. What is the range of % Latino in Ada County?

3. Open ArcToolbox. You will first compute the general G-statistic. Double-click the High/Low Clustering (Getis-Ord General G) tool in the Spatial Statistics Tools/Analyzing Patterns toolset. Select *adabg00* for the input feature class, select Latino for the input field, and check the box to display output graphically. Take the default for the other fields. Click OK to execute the command.
4. The graphic output shows the general G-statistic and its associated *Z* score. It also provides an interpretation of the result (i.e., clustering of high values) and the likelihood that the result is due to random chance (i.e., less than 1%). Close the graphic output.
5. Next you will run the local G-statistic. Double-click the Hot Spot Analysis (Getis-Ord Gi*) tool in the Spatial Statistics Tools/Mapping Clusters toolset. Select *adabg00* for the input feature class, select Latino for the input field, and specify a distance band of 5000 (meters). Click OK to execute the command.
6. Open the attribute table of *adabg00-Hotspots.* The table is the same as *adabg00* but has the added fields of GiZScore and GiPValue, storing the *Z* Score and the probability value, respectively, of the local G-statistic for each block group.

Q10. What is the value range of GiZScore?

7. On the map you can "see a "hot spot"" in Boise and another in the large block group to the southwest.

Challenge Task

What you need: *lochsa.mdb,* a personal geodatabase containing two feature classes for the Lochsa area of the Clearwater National Forest in Idaho.

lochsa_elk in the geodatabase has a field called USE that shows elk habitat use in summer or winter. *lt_prod* has a field called Prod that shows five timber productivity classes derived from the land type data, with 1 being most productive and 5 being least productive. Some polygons in *lt_prod* have the Prod value of −99 indicating absence of data. Also, *lochsa_elk* covers a larger area than *lt_prod* due to the difference in data availability.

This challenge task asks you to prove the statement that "the winter habitat area tends to have a higher area percentage of the productivity classes 1 and 2 than the summer habitat area." Specifically, you are asked to calculate (1) the area percentage of the winter habitat area with the Prod value of 1 or 2, and (2) the area percentage of the summer habitat area with the Prod value of 1 or 2. The statement would be true if the percentage from (1) is higher than the percentage from (2).

Q1. What is the area percentage of the summer habitat area with the Prod value of 1 or 2?

Q2. What is the area percentage of the winter habitat area with the Prod value of 1 or 2?