Max Calzada Muñoz

Prof. Jason Brown

ENVS-654 Geographic Information Systems

17 December 2022

Final Exam

1. Why is logical consistency important when overlaying data sets with spatial joins? 5pts

Logical consistency when important when overlaying data sets with spatial joins because the logical consistency ensures that no area is being double counted or that no area is left not counted (so that slivers are avoided). This can be accomplished by making sure that the different layers are all in the same coordinate system and capture data at the same level of resolution.

When logical consistency is missing in overlaying data sets with spatial joins, missing values will be produced. Statistics and arithmetic operations will not be calculable, or they will be erroneous.

2. Describe the difference between an attribute table and a standalone table. 5pts

An attribute table is a table containing rows of information for features. A standalone table is a table of information not linked to spatial data features

In other words, tables associated with spatial data sets are called attribute tables and contain records, one for each feature in the data set. Standalone tables are not associated with map features.

3. Explain the purpose of a map topology. 5pts

Topology refers to the spatial relationships between features in terms of adjacency, connectivity, intersection, or overlap

A topology can also define rules for how features can be related to each other, determine whether the rules have been broken, and assist the user in finding and fixing these errors.

4. List the main types of topological errors. 5pts

A dangle is when a line that fails to connect to another line.

A coincident boundary is when the shared edge contains exactly the same x-y pairs for two polygons which should not overlap

5. Which one(s) of the following tools or functions create a new data set in a geodatabase? 'Select Layer By Location', 'Make Feature Layer', 'Buffer', 'Clip', 'Definition Query', 'Export'. 5pts

> **'Make Feature Layer', 'Clip', and 'Export' create a new data set in a geodatabase.**

> **'Select Layer By Location', 'Buffer', 'Clip', and 'Definition Query', do NOT create a new data set in a geodatabase.**

6. Explain the difference between cardinality and join type. Describe why a one-to-many cardinality is often handled using a one-to-one join. 10pts

> **Cardinality is defined by the book as, "the numeric relationship between matching records in two tables: one-to-one, one-to-many, many-to-one, or many-to-many" and a join is defined by the book as "the temporary combination of data from two tables based on a common attribute field or location". In plain language, a join is how one relates different data sets and cardinality is how one relates the observations in the different data sets.**

> **A situation where one-to-many cardinality would be handled using a one-to-one join is if the user had a target table that listed Michigan's counties, but the target table did not list the population in each county. The user would then have to find a join table that lists each town in Michigan as well as its population and county it belongs to. The user would then sum all of the populations of the towns (like Waterford, Clarkston, Auburn Hills, etc.) in Oakland County to find the population for Oakland County. The join would allow the user to find the populations for Wayne County, Macomb County, and all the other counties in Michigan.**

7. If a point feature type is joined to a polygon layer, with the points as the target layer, what will the feature type of the output layer be?

> **The output layer will be a point feature layer since the point feature layer has the lowest dimensionality of geometry.**

> **The target feature class determines the type of features in the output feature class.**

8. What is an operator? Describe and give examples of each of the following: arithmetic operators, logical operators, spatial operators, and Boolean operators. 10pts

> **An operator is a kind of query that allows user to extract a set of features from a map or objects from a table.**

> **An arithmetic operators is when data from one feature layer is affected by data from another feature layer, using an arithmetic expression. For example, if you wanted to find how much the population of Malibu, California increased between**

2020 and 2010, the user could subtract the 2010 population from the 2020 population.

Boolean operators (AND, OR, NOT, XOR) evaluate a pair of true or false propositions and return a true or false result. For example, a user could try to find where weather is sunny or warm exclusively (but not both) across Mexico using XOR.

Logical operators include functions such as >, <, or = that compare values and generate a true or false result. For example, a user could compare the 2020 population of Malibu, California to the 2020 population of Santa Monica, California.

Spatial operators are a set of functions that evaluate spatial relationships, including intersection, containment, and proximity between sets of features. For example, a user could use containment to find that Las Vegas is within Nevada.

9. Briefly summarize five of the six different data quality issues addressed by good metadata. 5pts

Geometric accuracy describes how close the x-y values of a data set correspond to the actual locations on the earth's surface.

Thematic accuracy refers to the attributes stored in the table. Some types of data are relatively straightforward to record, such as the name of a city or the number of lanes in a road.

Resolution refers to the sampling interval at which data are acquired. Resolution may be spatial, thematic, or temporal.

Precision refers to either the number of significant digits used to record a measurement or the statistical variation of a repeated single measurement.

Logical consistency evaluates whether a data model or data set accurately represents the real-world relationships between features.

10. What is the primary purpose of domains? Explain the difference between a coded domain and a range domain. 5pts

Domains provide rules about the values placed in attribute fields to help avoid errors and foster consistent data entry. Coded domains include lists of values to pick from. Range domains specify the maximum and minimum allowed values for numeric fields.

11. In what ways does the Clip tool differ from the Select Layer By Location tool? 5pts

A clip works like a cookie cutter to truncate the features of one file based on the outline of another. The features being extracted come from the input layer. Only the outside boundary of the clip layer is used for clipping; internal boundaries, if present, have no effect on the output. In clip, any features crossing the boundary will be truncated, potentially changing areas, perimeters, or lengths of features.

The Select Layer By Location tool provides the means to set up a spatial query. The Relationship identifies the spatial operator used to compare the layers. The Selecting Features identifies the layer to which the input features are compared.

The difference is that a clip creates a new feature, a Select Layer By Location does not.

12. What is a dangle and how are dangles prevented when digitizing new lines? 5pts

A dangle is a line that fails to connect to another line when it is supposed to.

The Snapping tool ensures that the nodes of lines and the vertices of polygons match (and that dangles are prevented when digitizing new lines).

13. Before any raster algebra expression can be executed, what must be true of the resolution and structure of the input rasters?

To perform raster analysis, all of the input rasters must have the same cell size and extent so that the cells in each stack align. If a tool receives inputs with different cell sizes or extents, the rasters will automatically be resampled to match each other. By default, all rasters are resampled to the most coarse cell size present in the input rasters.

14. Describe the difference between a join and a relate. When is a relate the only option?

A join is the temporary combination of data from two tables based on a common attribute field or location. A relate is a temporary association between two tables based on a common field whereby fields may be selected based on whether they match selected records in the other table. The difference here is that a join actually results in a new data table, a relate does not.

A relate is the only option when working with a many-to-many relationship.

15. Why is logical consistency a potential issue during spatial queries? 5pts

Without logical consistency, a spatial query may not be executable if there are slivers or angles. If an area is double counted, queries and operators may be executable, but the results will be erroneous.

The expected relationships between regions with shared boundaries is not reflected in the database. Furthermore, because feature classes vary in precision and geometric accuracy, it can often happen that two objects which coincide in the real world will not match when the GIS coordinates are compared.

16. What primary characteristics distinguish a spatial join from selecting by location? 5pts

A spatial join a function that combines the attributes of features in two layers based on a spatial relationship. Selecting by location is simply selecting elements from one layer only based on location.

Spatial joins may be performed on any two feature classes. A user can join points to points, polygons to polygons, lines to points, and any combination of the three geometry types,

Cardinality is an important issue for spatial joins, whereas cardinality is not necessarily an issue for selecting joins.

When selecting by location, the geometry type never changes. Polygons remain polygons and point features remain as such. In spatial joins, when the target is a point feature layer and the join is a polygon feature layer, the output layer stays as a point feature layer, but it incorporates information from the polygon feature layer.

17. If a data layer is in North America Equidistant Conic coordinates and the data frame is set to UTM coordinates, in which coordinate system will the edits be stored? 10pts

If the output is placed in a feature dataset, the coordinate system will always match that of the feature dataset.

If the coordinate system is set in the Environment settings and the output will not become part of a feature dataset, the Environment settings coordinate system is used.

** If the Environment setting is not set, the default rule applies—that the output will match the coordinate system of the first input to the tool.**

That means that in this case, the edits to the coordinate system will be match the data frame (UTM coordinates).

18. Explain what happens when a feature class is generalized or dissolved. What are some reasons why it would be done? 5pts

The dissolve tool is used to merge features when they share the same attribute value(s).

The generalize tool is used to simplify the spatial complexity of line or polygon features by removing vertices that are closer than a specified tolerance value, such as simplifying this coastline.

Sometimes a data set may be too detailed for effective use. A data set with a large source scale may take too long to draw because there are too many vertices for the scale at which it will be used. In these situations, simplifying the data may result in better performance for the intended task.

19. You have a table of cities and a table of counties, both with a state abbreviation field. Can you join them if cities is the target table? If counties is the target table? Explain your answer. 5pts

If the cities are the target table, the counties table can be joined to cities table as the county of each city will be added as a new column.

If the counties are the target table, the cities table cannot be joined to the counties table as the user would have to store multiple values (the city names) in each cell and then try to add this nonsensical column.

20. Why is it best to store rasters in the same coordinate system in which they will be analyzed? Which types of projections work best? 5pts

The reason why is it best to store rasters in the same coordinate system in which they will be analyzed is because rasters are always rectangles. If two different rasters are of different resolution levels, one or both rasters will need to be resampled (and this could lead to the loss of important details and information). Rasters must be stored in the same coordinate system in which they will be analyzed because that coordinate system and level of sampling determines the level of detail at which map algebra and raster calculations will be carried out.

Additionally, many raster analysis calculations use distance and angle formulas. The results would be inaccurate if a different projection is used, causing a significant distortion.

The Universal Transverse Mercator (UTM) is the best type of projection as it is ideal for any single zone like a county.