CHAPTER 11 VECTOR DATA ANALYSIS

Vector Data Analysis

- Vector data analysis uses the geometric objects of point, line, and polygon.
- The accuracy of analysis results depends on the accuracy of these objects in terms of location and shape.
- Topology can also be a factor for some vector data analyses such as buffering and overlay.

Buffering

- Based on the concept of proximity, buffering creates two areas: one area that is within a specified distance of select features and the other area that is beyond.
- The area that is within the specified distance is called the buffer zone.
- There are several variations in buffering. The buffer distance can vary according to the values of a given field. Buffering around line features can be on either the left side or the right side of the line feature. Boundaries of buffer zones may remain intact so that each buffer zone is a separate polygon.

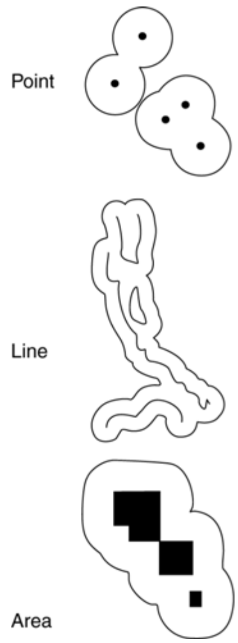

Figure 11.1
Buffering around points, lines, and areas.

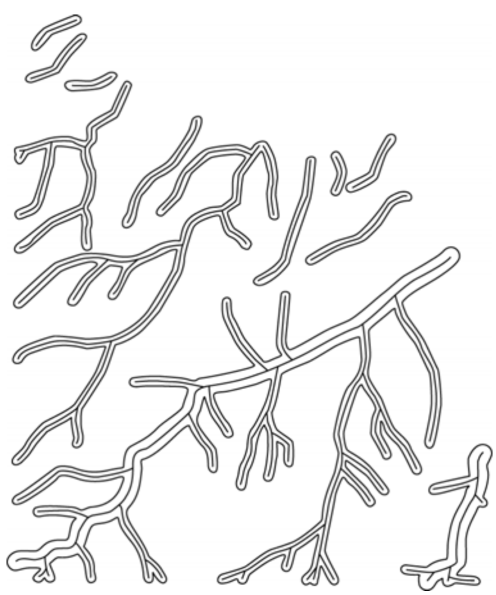

Figure 11.2
Buffering with different buffer distances.

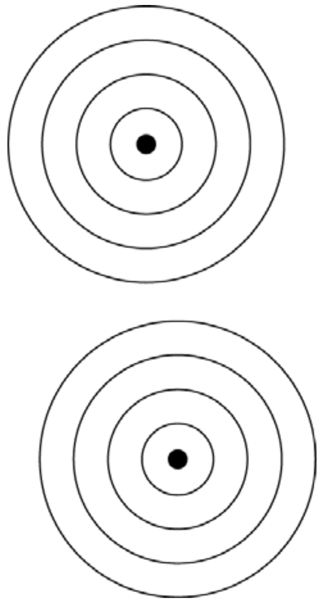

Figure 11.3
Buffering with four rings.

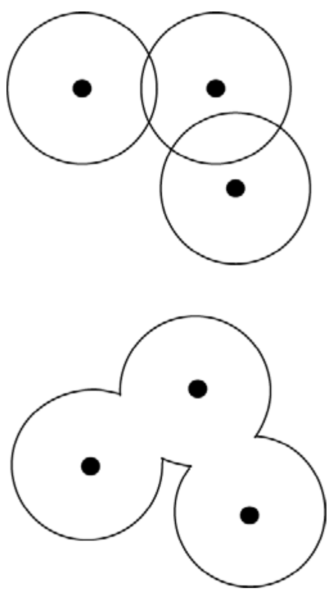

Figure 11.4
Buffer zones not dissolved (top) or dissolved (bottom).

Overlay

- An overlay operation combines the geometries and attributes of two feature layers to create the output.
- The geometry of the output represents the geometric intersection of features from the input layers.
- Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbors.

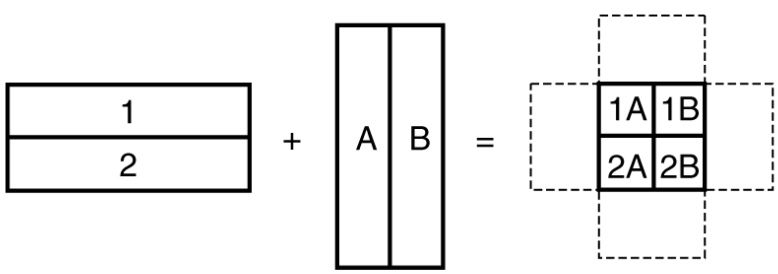

Figure 11.5
Overlay combines the geometry and attribute data from two layers into a single layer. The dashed lines are not included in the output.

Feature Type and Overlay

Overlay operations can be classified by feature type into point-in-polygon, line-in-polygon, and polygon-on-polygon.

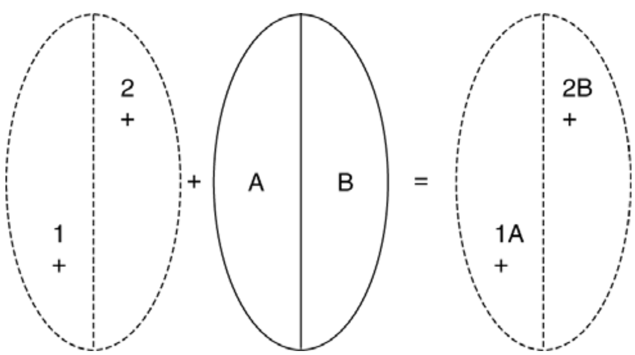

Figure 11.6
Point-in-polygon overlay. The input is a point layer (the dashed lines are for illustration only and are not part of the point layer). The output is also a point layer but has attribute data from the polygon layer.

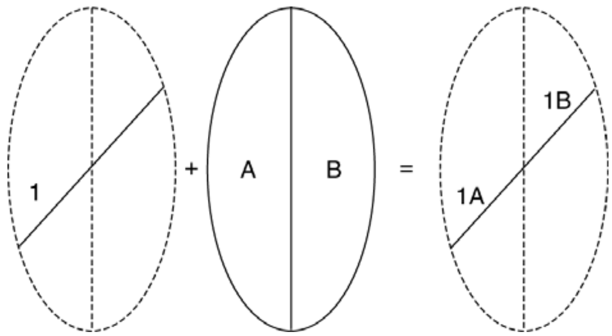

Figure 11.7
Line-in-polygon overlay. The input is a line layer (the dashed lines are for illustration only and are not part of the line layer). The output is also a line layer. But the output differs from the input in two aspects: the line is broken into two segments, and the line segments have attribute data from the polygon layer.

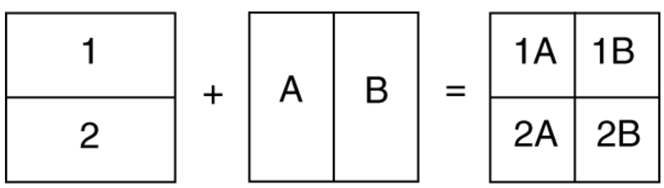

Figure 11.8
Polygon-on-polygon overlay. In the illustration, the two layers for overlay have the same area extent. The output combines the geometry and attribute data from the two layers into a single polygon layer.

Overlay Methods

- All overlay methods are based on the Boolean connectors of AND, OR, and XOR.
- An overlay operation is called Intersect if it uses the AND connector.
- An overlay operation is called Union if it uses the OR connector.
- An overlay operation that uses the XOR connector is called Symmetrical Difference or Difference.
- An overlay operation is called Identity or Minus if it uses the following expression: [(input layer) AND (identity layer)] OR (input layer).

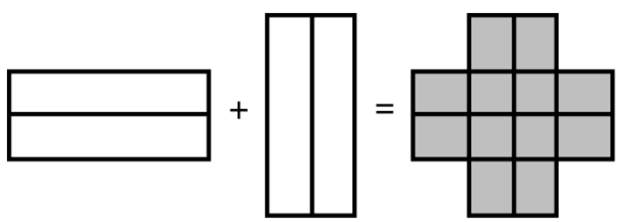

Figure 11.9
The Union method keeps all areas of the two input layers in the output.

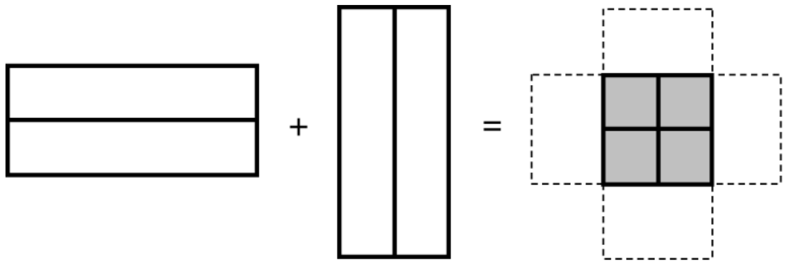

Figure 11.10
The Intersect method preserves only the area common to the two input layers in the output. (The dashed lines are for illustration only; they are not part of the output.)

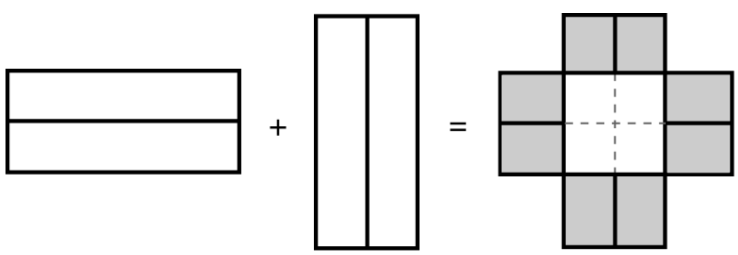

Figure 11.11
The Symmetric Difference method preserves only the area common to only one of the input layers in the output. (The dashed lines are for illustration only; they are not part of the output.)

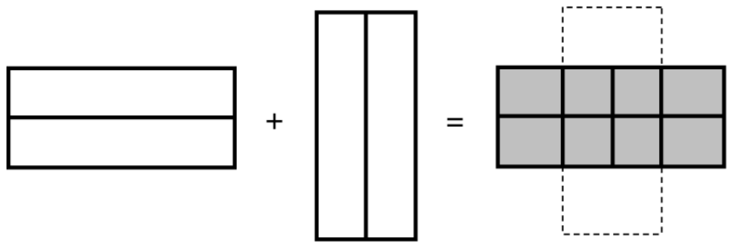

Figure 11.12
The Identity method produces an output that has the same extent as the input layer. But the output includes the geometry and attribute data from the identity layer.

Slivers

- A common error from overlaying polygon layers is slivers**,** very small polygons along correlated or shared boundary lines of the input layers.
- To remove slivers, ArcGIS uses the cluster tolerance, which forces points and lines to be snapped together if they fall within the specified distance.

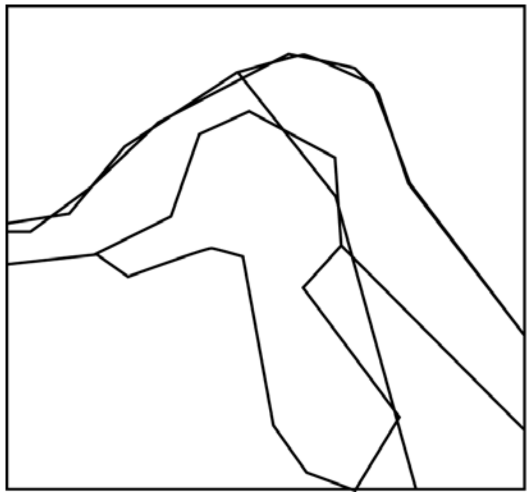

Figure 11.13
The top boundary has a series of slivers. These slivers are formed between the coastlines from the input layers in overlay.

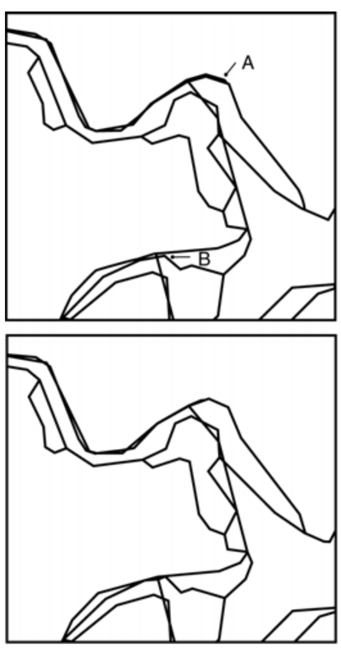

Figure 11.14
A cluster tolerance can remove many slivers along the top boundary (A) but can also snap lines that are not slivers (B).

Areal Interpolation

One common application of overlay is to help solve the areal interpolation problem. Areal interpolation involves transferring known data from one set of polygons (source polygons) to another (target polygons).

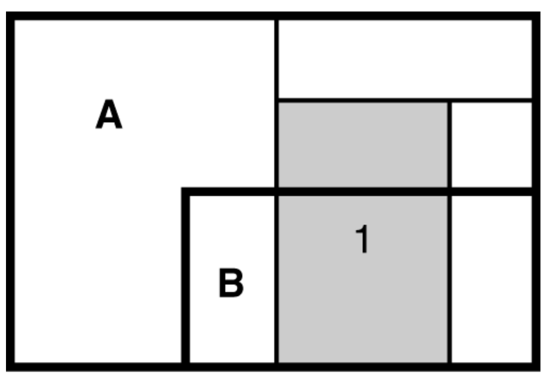

Figure 11.15
An example of areal interpolation. Thick lines represent census tracts and thin lines school districts. Census tract A has a known population of 4000 and B has 2000. The overlay result shows that the areal proportion of census tract A in school district 1 is 1/8 and the areal proportion of census tract B, 1/2. Therefore, the population in school district 1 can be estimated to be 1500, or [(4000 x 1/8) + (2000 x 1/2)].

Pattern Analysis

- Pattern analysis refers to the use of quantitative methods for describing and analyzing the distribution pattern of spatial features.
- At the general level, a pattern analysis can reveal if a distribution pattern is random, dispersed, or clustered.
- At the local level, a pattern analysis can detect if a distribution pattern contains local clusters of high or low values.
- Pattern analysis includes point pattern analysis (nearest neighbor, Ripley's K-function), Moran's I for measuring spatial autocorrelation, and G-statistic for measuring high/low clustering.

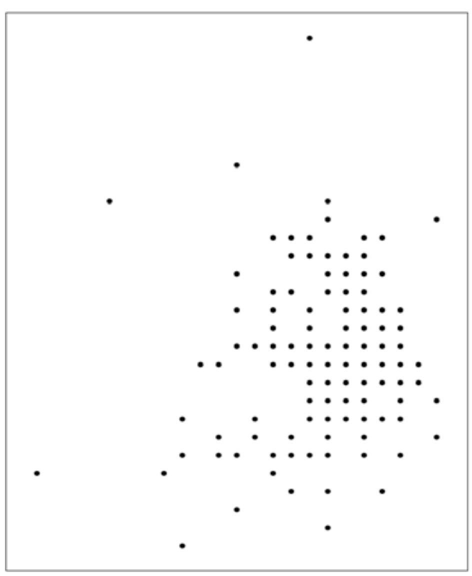

Figure 11.16
A point pattern showing deer locations.

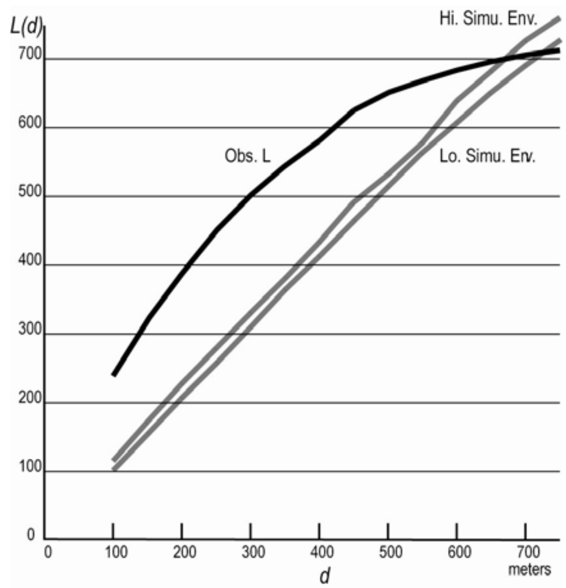

Figure 11.17
The computed $L(d)$ and the lower and upper simulation envelopes.

Table 11.1 Expected $L(d)$, Observed $L(d)$, and their difference for deer location data

Expected $L(d)$	**Observed $L(d)$**	**(2) – (1)**
100	239.3	139.3
150	323.4	173.4
200	386.8	186.8
250	454.5	204.5
300	502.7	202.7
350	543.9	193.9
400	585.1	185.1
450	621.5	171.5
500	649.5	149.5
550	668.3	118.3
600	682.9	82.9
650	697.1	47.1
700	704.9	4.9
750	713.7	-36.3

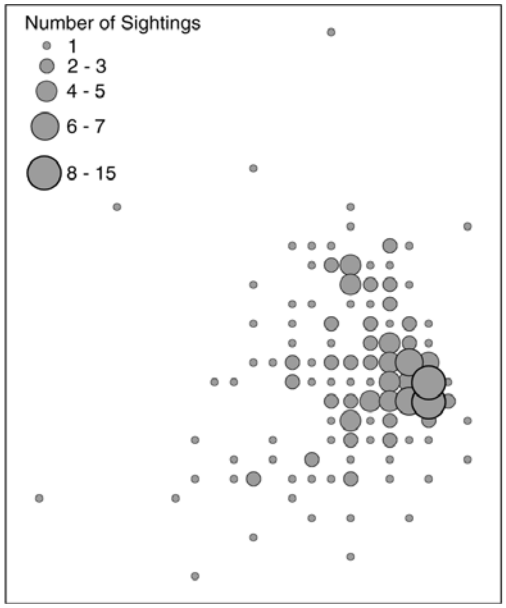

Figure 11.18
A point pattern showing deer locations and the number of sightings at each location.

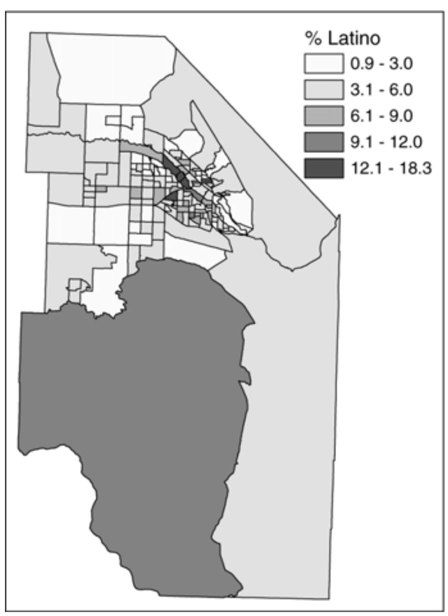

Figure 11.19
Percent Latino population by block group in Ada County, Idaho. Boise is located in the upper center of the map with small sized block groups.

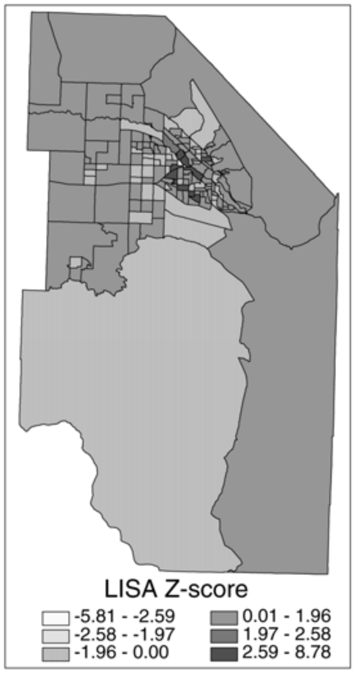

Figure 11.20
Z scores for the Local Indicators of Spatial Association (LISA) by block group in Ada County, Idaho.

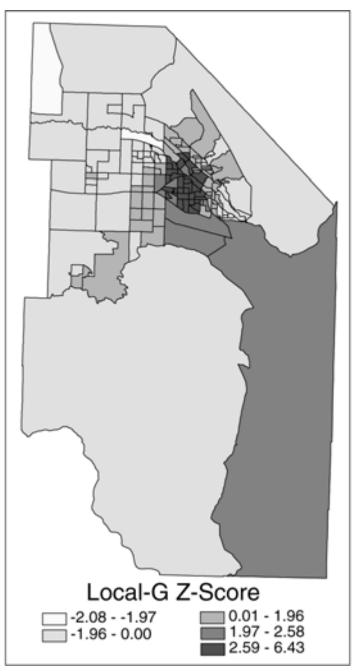

Figure 11.21
Z scores for the local G-statistics by block group in Ada County, Idaho.

Feature Manipulation

- Tools are available in a GIS package for manipulating and managing maps in a database.
- These tools include Dissolve, Clip, Append, Select, Eliminate, Update, Erase, and Split.

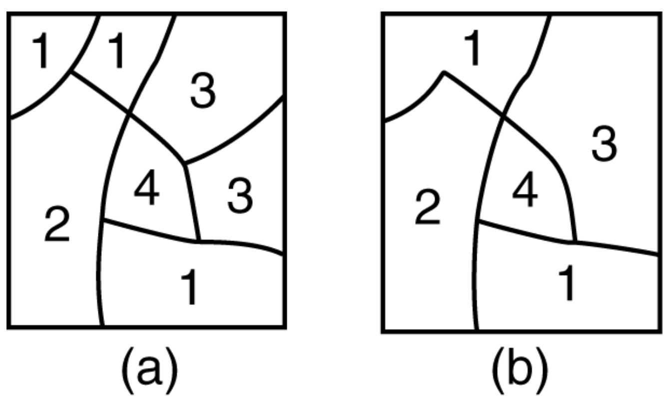

Figure 11.22
Dissolve removes boundaries of polygons that have the same attribute value in (a) and creates a simplified layer (b).

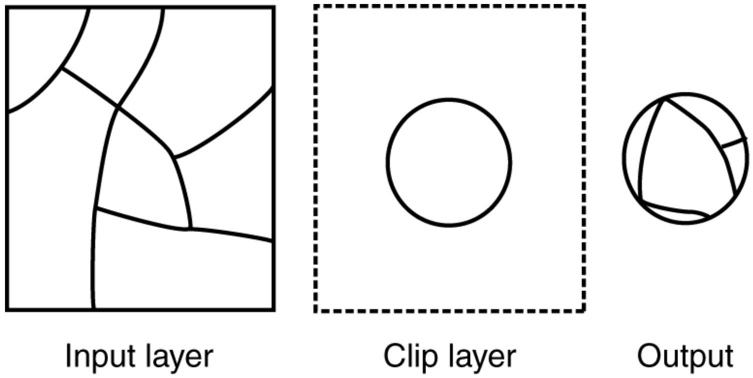

Figure 11.23
Clip creates an output that contains only those features of the input layer that fall within the area extent of the clip layer. (The dashed lines are for illustration only; they are not part of the clip layer.)

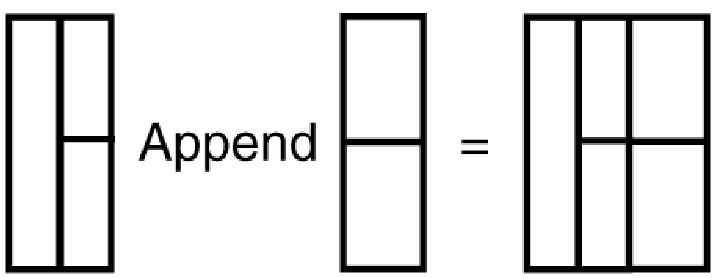

Figure 11.24
Append pieces together two adjacent layers into a single layer but does not remove the shared boundary between the layers.

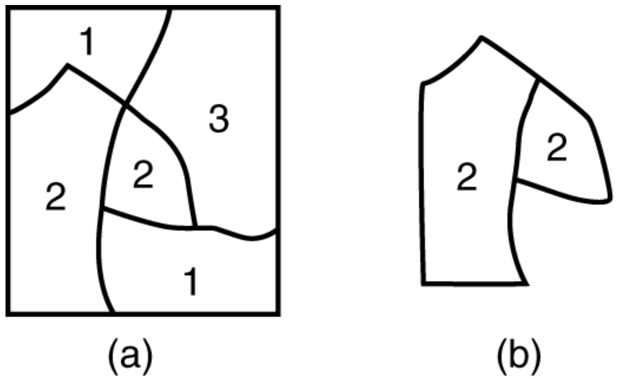

Figure 11.25
Select creates a new layer (b) with selected features from the input layer (a).

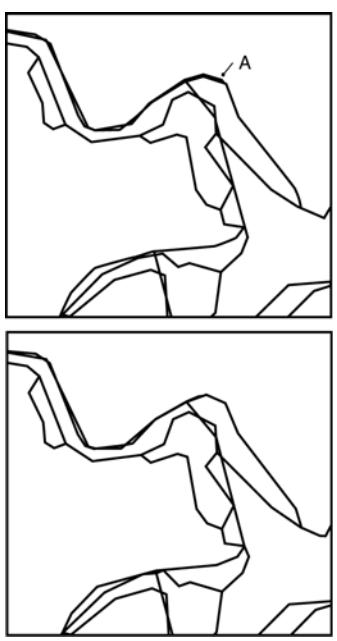

Figure 11.26
Eliminate removes some small slivers along the top boundary (A).

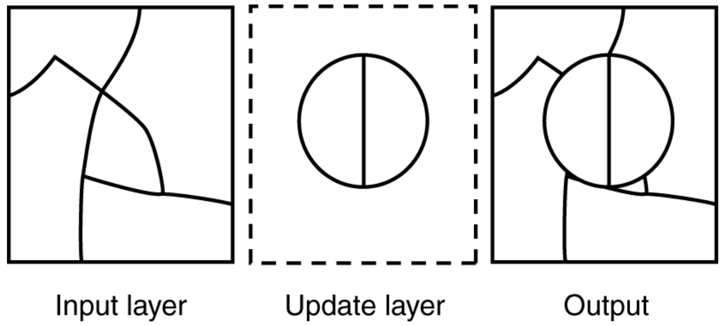

Figure 11.27
Update replaces the input layer with the update layer and its features. (The dashed lines are for illustration only; they are not part of the update layer.)

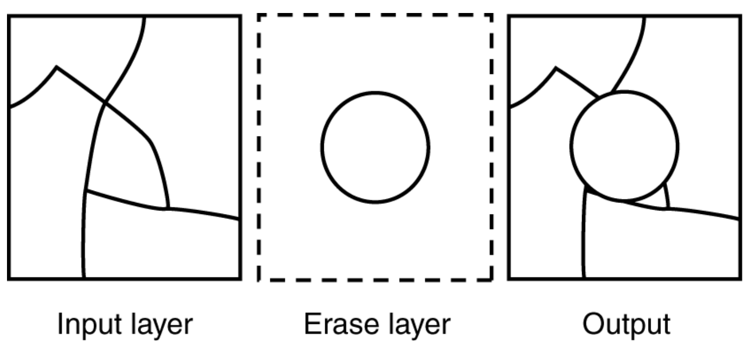

Figure 11.28
Erase removes features from the input layer that fall within the area extent of the erase layer. (The dashed lines are for illustration only; they are not part of the erase layer.)

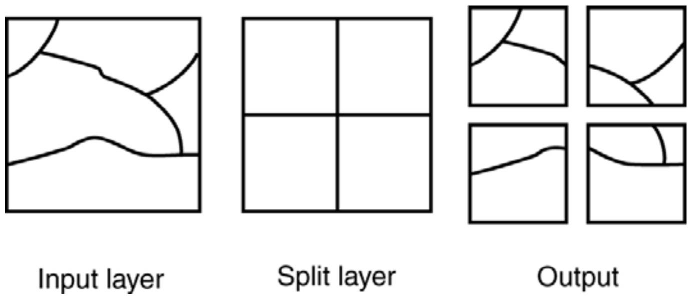

Figure 11.29
Split uses the geometry of the split layer to divide the input layer into four separate layers.

CrimeStat
http://www.icpsr.umich.edu/NACJD/crimestat.htm