BASIC ANALYSIS

• OVERLAY
• GRAPHIC MANIPULATIONS
• BUFFER ZONES

RECODE APPLICATION

AGRICULTURAL SOILS

SOILS

AGRICULTURAL SOILS

RECODE:
A - B = AGRICULTURAL
C - E = NON-AGRICULTURAL

OVERLAY APPLICATION

CROPS

SOILS

WHICH SOILS OCCUR WITH CROP 3?

OVERLAY

COINCIDENCE VIEW

ALL CROPS AND SOILS

OVERLAY APPLICATION

SOIL TYPE  CROP TYPE  FARM PRACTICE

YIELD POTENTIAL

OVERLAY
IDENTIFICATION OF FEATURES WITHIN POLYGONS

POINTS

HISTORICAL SITES  PARK DISTRICTS

OVERLAY + 1 2
3 4

POINT A = DISTRICT 1
POINT B = DISTRICT 4

LINES

OVERLAY + 1 2
3 4

RIVER X = DISTRICT 1.3
RIVER Y = DISTRICT 3.4

MAP ALGEBRA
MULTIPLICATION

OVERLAY USING MULTIPLY

<table>
<thead>
<tr>
<th>DISTRICTS</th>
<th>CROP LAND</th>
<th>COINCIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>1</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>3 4</td>
<td>0</td>
<td>0 1 2 3 4 5 6</td>
</tr>
</tbody>
</table>

MAP ALGEBRA
OVERLAY USING MAXIMUM

INPUT COVERAGE A

<table>
<thead>
<tr>
<th>1 2 3 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 3 4 4</td>
</tr>
<tr>
<td>4 0 1 0</td>
</tr>
<tr>
<td>7 2 4 6</td>
</tr>
</tbody>
</table>

RAINFALL: 1980

INPUT COVERAGE B

<table>
<thead>
<tr>
<th>1 2 3 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 5 5 6</td>
</tr>
<tr>
<td>7 4 8 9</td>
</tr>
</tbody>
</table>

RAINFALL: 1981

OUTPUT COVERAGE C

<table>
<thead>
<tr>
<th>4 3 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 5 5</td>
</tr>
<tr>
<td>4 4 6</td>
</tr>
</tbody>
</table>


OVERLAY USING WEIGHTS

ELEV
3 3 1
2 1 1
1 2 1

SOILS
3 3 3
2 2 2
1 1 1

SLOPE
2 2 2
1 1 1
3 1 1

WEIGHT
3 3 1
2 1 1
1 2 1

X 1 =

SLOPE

2 2 2
1 1 1
3 1 1

WEIGHT
6 6 6
4 4 4
2 2 2

X 2 =

OVERLAY USING ADD

HIGHEST VALUES HAVE HIGHEST RATING

11 11 9
7 6 6
6 5 4

1 = POOR
2 = MODERATE
3 = BEST

**CLIP and MASK**

**CLIP**

\[
\begin{array}{ccc}
A & B & C \\
D & E & F \\
G & H & I \\
\end{array}
\]

DISTRICTS

\[
\begin{array}{c}
\text{CLIP DRILL SITES} \\
\end{array}
\]

\[
\text{E-F SITES} \\
= \\
\text{ROADS} \\
= \\
\text{ROADS AND E-F DRILL SITES}
\]

**MASK**

\[
\begin{array}{ccc}
A & B & C \\
D & E & F \\
G & H & I \\
\end{array}
\]

DISTRICTS

\[
\begin{array}{c}
\text{MASK} \\
\end{array}
\]

\[
\text{ROADS AND E-F DRILL SITES} \\
= \\
\text{E-F ROADS AND DRILL SITES}
\]

HAZARD APPLICATION

FLOODING RISKS

RIVERS

FLOOD ZONES

OVERLAY

VILLAGES

ROADS

FLOOD-PRONE VILLAGES

FLOOD-PRONE ROADS

BUFFER DISTANCE

DISTANCE CALCULATION

POSTAL ZONES: COST OF MAILING

POLLUTION ZONES

SPATIAL ANALYSIS

OVERLAY

HOW MUCH CORN IS GROWN WITHIN 5 KM OF TOWN D?
WHAT IS THE TOTAL LENGTH OF ROADS WITHIN 5 KM OF D?

Spatial Analysis and Modeling Resources

Special thanks to Ron Briggs, UT-Dallas
Analysis Components

• analyses may be applied to (use as input):
  – tabular attribute data
  – spatial data/layers
  – combination of spatial and tabular

• results may be displayed as (produce as output):
  – table subsets, table combinations, highlighted records (rows), new variables (columns)
  – charts
  – maps/map features:
    » highlights on existing themes
    » new themes/layers
  – combination
Availability of Analytical Capabilities:

• **Basic: Desktop GIS packages**
  - ArcView
  - Mapinfo
  - Geomedia

• **Advanced: Professional GIS systems**
  - ARC/INFO, MapInfo Pro, etc.
  - provide data editing plus more advanced analysis
  Provided now through extra cost Extensions or professional versions of desktop packages

• **Specialized: modeling and simulation**
  - via scripting/programming within GIS
    » AMLs in Workstation ARC/INFO (v. 7)
    » Avenue scripts in ArcView 3.2
    » VB and ArcObjects in ArcGIS
    Write your own or download from ESRI Web site
  - via specialized packages and/or GISs
    » 3-D Scientific Visualization packages
    » transportation planning packages
    » ERDAS, ER Mapper or similar package for raster

Capabilities move ‘down the chain’ over time.

In earlier generation GISes, use of advanced applications often required learning another package with a different user interface and operating system (usually UNIX).
Advanced and Specialized Applications: 

*in comparison to basic applications*

Most ‘basic’ analyses are used to create *descriptive models* of the world, that is, representations of reality as it exists.

Most ‘advanced’ analyses involve creating a new *conceptual output* layer, or in some cases table(s) or chart(s), the values of which are some *transformation* of the values in the *descriptive input* layer.

e.g. slope or aspect layer

Most ‘specialized’ applications involve using GIS capabilities to create a *predictive model* of a real world process, that is, a model capable of reproducing processes and/or making predictions or projections as to how the world *might* appear.

e.g. fire spread model, traffic projections
Analysis Options: Basic
(Table of Contents)

• **Spatial Operations**
  - Vector
    - centroid determination
    - spatial measurement
    - buffer analysis
    - spatial aggregation
      » redistricting
      » regionalization
      » classification
    - Spatial overlays and joins
  - Raster
    - neighborhood analysis/spatial filtering
    - Raster modeling
  - geostatistics

• **Attribute Operations**
  - record selection
    » tabular via SQL
    » ‘information clicking’ with cursor
  - variable recoding
  - record aggregation
  - general statistical analysis
  - table relates and joins
Analysis Options: Advanced & Specialized

Advanced

• surface analysis
  – cross section creation
  – visibility/viewshed
• proximity analysis
  – nearest neighbor layer
  – distance matrix layer
• network analysis
  – routing
    » shortest path (2 points)
    » travelling salesman (n points)
  – time districting
  – allocation
• Thiessen Polygon creation

Specialized

• Remote Sensing image processing and classification
• raster modeling
• 3-D surface modeling
• spatial statistics/statistical modeling
• functionally specialized
  – transportation modeling
  – land use modeling
  – hydrological modeling
  – etc.
Spatial operations:  
**Centroid or Mean Center**

- balancing point for a spatial distribution
  - point representation for a polygon--analogous to the mean  
  \[ \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \]
  - single point summary for a distribution (point or polygon)
  - can be weighted by ‘magnitude’ at each point (analogous to weighted mean)
  - minimizes squared distances to other points, thus ‘distant’ points have bigger influence than close points (Oregon births more impact than Kansas births!)
  - is not the point of “minimum aggregate travel”--this would minimize distances (not their square) and can only be identified by approximation.

- useful for
  - summarizing change over time in a distribution (e.g US pop. centroid every 10 years)
  - placing labels for polygons

- for weird-shaped polygons, centroid may not lie within polygon

Note: many ArcView applications calculate only a “psuedo” centroid: the coordinates of the bounding box (the extent) of the polygon
Spatial measurements:
- distance measures
  - between points
  - from point or raster to polygon or zone boundary
  - between polygon centroids
- polygon area
- polygon perimeter
- polygon shape
- volume calculation
  - e.g. for earth moving, reservoirs
- direction determination
  - e.g. for smoke plumes

Spatial operations:

Comments:
- possible distance metrics:
  - straight line/airline
  - city block/manhattan metric
  - distance thru network
  - time/friction thru network
- shape often measured by:

\[
\frac{\text{perimeter}}{\sqrt{\text{area} \times 3.54}} = 1.0 \text{ for circle}
\]
\[
= 1.13 \text{ for square}
\]
\[
= \text{large number for irregular polygon}
\]

ArcGIS 8.1 geodatabases contain automatic variables:
- \textit{shape.length}: line length or polygon perimeter
- \textit{shape.area}: polygon area
(Much easier than AV 3.2—see appendix!)

Perimeter to area ratio differs
Spatial operations:

**Spatial Measurement--example**

<table>
<thead>
<tr>
<th>SHAPE</th>
<th>AREA</th>
<th>PERIMETER</th>
<th>CNTY_</th>
<th>CNTY_ID</th>
<th>NAME</th>
<th>FIPS</th>
<th>Shape Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polygon</td>
<td>0.265</td>
<td>2.729</td>
<td>2605</td>
<td>2605</td>
<td>Anderson</td>
<td>48001</td>
<td>1.50</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.368</td>
<td>2.564</td>
<td>2545</td>
<td>2545</td>
<td>Andrews</td>
<td>48003</td>
<td>1.19</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.209</td>
<td>2.171</td>
<td>2680</td>
<td>2680</td>
<td>Angelina</td>
<td>48005</td>
<td>1.34</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.072</td>
<td>2.642</td>
<td>2899</td>
<td>2899</td>
<td>Aransas</td>
<td>48007</td>
<td>2.78</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.233</td>
<td>1.941</td>
<td>2335</td>
<td>2335</td>
<td>Archer</td>
<td>48009</td>
<td>1.14</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.233</td>
<td>1.941</td>
<td>2103</td>
<td>2103</td>
<td>Armstrong</td>
<td>48011</td>
<td>1.14</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.299</td>
<td>2.278</td>
<td>2870</td>
<td>2870</td>
<td>Atascosa</td>
<td>48013</td>
<td>1.18</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.224</td>
<td>1.900</td>
<td>2471</td>
<td>2471</td>
<td>Dallas</td>
<td>48113</td>
<td>1.13</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.222</td>
<td>1.889</td>
<td>2481</td>
<td>2481</td>
<td>Dawson</td>
<td>48115</td>
<td>1.13</td>
</tr>
<tr>
<td>Polygon</td>
<td>0.368</td>
<td>2.580</td>
<td>2106</td>
<td>2106</td>
<td>Deaf Smith</td>
<td>48117</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Shape Index can be calculated from area and perimeter measurements. These are available in *Attributes of ..... file* if source is a coverage or a geodatabase. If source is a shapefile, you need an *ArcScript* to calculate area and perimeter (or convert to geodatabase) (Note: “Shape index” is unrelated to “shapefile.”)
Spatial Operations: *buffer zones*

- region within ‘x’ distance units
- buffer any object: point, line or polygon
- use multiple buffers at progressively greater distances to show gradation
- may define a ‘friction’ or ‘cost’ layer so that spread is not linear with distance
- Implement in Arcview 3.2 with Theme/Create buffers in ArcGIS 8 with Tools/Buffer Wizard

Examples

- 200 foot buffer around property where zoning change requested
- 100 ft buffer from stream center line limiting development
- 3 mile zone beyond city boundary showing ETJ (extra territorial jurisdiction)
- use to define (or exclude) areas as options (e.g. for retail site) or for further analysis
- in conjunction with ‘friction layer’, simulate spread of fire

Note: only one layer is involved, but the buffer can be output as a new layer
Spatial Operations: spatial aggregation

- **districting/redistricting**
  - grouping contiguous polygons into districts
  - original polygons preserved

- **Regionalization (or dissolving)**
  - grouping polygons into contiguous regions
  - original polygon boundaries dissolved

- **classification**
  - grouping polygons into non-contiguous regions
  - original boundaries usually dissolved
  - usually ‘formal’ groupings

Grouping/combining polygons—is applied to one polygon layer only.

**Criteria may be:**
- formal (based on in situ characteristics)
  e.g. city neighborhoods
- functional (based on flows or links):
  e.g. commuting zones

**Groupings may be:**
- contiguous
- non-contiguous

**Boundaries for original polygons:**
- may be preserved
- may be removed (called dissolving)

**Examples:**
- elementary school zones to high school attendance zones (*functional districting*)
- election precincts (or city blocks) into legislative districts (*formal districting*)
- creating police precincts (*funct. reg.*)
- creating city neighborhood map (*form. reg.*)
- grouping census tracts into market segments—yuppies, nerds, etc (*class.*)
- creating soils or zoning map (*class*)

Implement in ArcView 8 thru Tools/Geoprocessing Wizard, using *dissolve features based on an attribute*
**Districting**: elementary school attendance zones grouped to form junior high zones.

**Regionalization**: census tracts grouped into neighborhoods

**Classification**: cities categorized as central city or suburbs soils classified as igneous, sedimentary, metamorphic
Proximity Calculations

- **POINTDISTANCE** computes the distances between point features in one coverage to all points in another coverage that are within the specified search radius.

POINTDISTANCE <from_cover> <to_cover> <out_info_file> {search_radius}
Proximity Calculations, page 2

- **NEAR** computes the distance from each point in a coverage to the nearest arc, point or node in another coverage.

  ```
  NEAR <in_cover> <near_cover> {LINE | POINT | NODE} {search_radius} {out_cover} {NOLOCATION | LOCATION}
  ```

  **Arc:**
  ```
  location
  ```

  The PAT of the output coverage will have two additional items: one to store the distance to the nearest feature in the near_cover, and a second to store the internal number of the nearest feature. **LOCATION** adds items for x and y coordinates.
Spatial Operations:

Spatial Matching: Spatial Joins and Overlays

- **combine two (or more) layers to:**
  - select features in one layer, &/or
  - create a new layer

- **used to integrate data having different spatial properties (point v. polygon), or different boundaries (e.g. zip codes and census tracts)**

- **can overlay polygons on:**
  - points *(point in polygon)*
  - lines *(line on polygon)*
  - other polygons *(polygon on polygon)*
  - many different Boolean logic combinations possible
    » Union (A or B)
    » Intersection (A and B)
    » A and not B ; not (A and B)

- **can overlay points on:**
  - Points, which finds & calculates distance to nearest point in other theme
  - Lines, which calculates distance to nearest line

**Examples**

- assign environmental samples (points) to census tracts to estimate exposure per capita *(point in polygon)*
- identify tracts traversed by freeway for study of neighborhood blight *(polygon on lines)*
- integrate census data by block with sales data by zip code *(polygon on polygon)*
- Clip US roads coverage to just cover Texas *(polygon on line)*
- Join capital city theme to all city theme to calculate distance to nearest state capital *(point on point)*
POLYGON OVERLAY

- POLYGON OVERLAY commands involve three coverages: an input coverage, an overlay coverage, and an output coverage created as a result of the overlay.
- Input coverage features can be polygons, lines, or points.
- The overlay coverage feature must be polygons.
- Output coverage features are of the same class as the input coverage features.
  - polygon-on-polygon overlay - output is a polygon coverage
  - line-on-polygon overlay - output is an arc coverage
  - point-on-polygon overlay - output is a point coverage

- CLIP, ERASE, SPLIT, UPDATE, UNION, INTERSECT, IDENTITY
CLIP

- CLIP - extracts those features from an input coverage that overlap with a clip coverage. This is the most frequently used polygon overlay command to extract a portion of a coverage to create a new coverage.

```
CLIP <in_cover> <clip_cover> <out_cover> {POLY | LINE | POINT | NET | LINK | RAW} {fuzzy_tolerance}
```

The <clip_cover> must have polygon topology.
Boundaries of interior polygons in the <clip_cover> are not used in CLIP.
CLIP uses the clip coverage as a **cookie cutter**; only those input coverage features that are within the clip coverage are stored in the output coverage.
Topology is built for the output coverage.
Only the attributes of the in_cover are retained in the out_cover
Clipping an Image

• Clipping of images is fundamentally different from clipping coverages. This is because images are a raster data format while coverages are a vector format. There are several ways to clip an image. The most straightforward way is given herein.

• Use the RECTIFY command, specifying a clip BOX or clip_cover on the command line. Note that the clip_cover area is defined by the BND of the coverage, not the polygon features in the coverage.

• An image is always rectangular, so it can only be clipped with a rectangular box or BND.
ERASE

- **ERASE** - erases the input coverage features that overlap with the erase coverage polygons.

ERASE <in_cover> <erase_cover> <out_cover> {POLY | LINE | POINT | NET | LINK | RAW} {fuzzy_tolerance}

The <erase_cover> must have polygon topology. Boundaries of interior polygons in the <erase_cover> are not used in ERASE.

The polygons of the erase coverage define the erasing region. Input coverage features that are within the erasing region are removed. The output coverage contains only those input coverage features that are outside the erasing region.

Erase is the opposite of clip: it leaves you with “the left over dough”

Topology is rebuilt for the output coverage.
ERASE

POLY option

LINE option

POINT option
SPLIT

- **SPLIT - breaks a single coverage into many coverages.**
  - `SPLIT <in_cover> <split_cover> <split_item> {POLY | LINE | POINT | NET | LINK | RAW} {fuzzy_tolerance}`
    - `<split_item>` - the item in `<split_cover>` which will be used to split the `<in_cover>`.
    - You will be prompted to enter the names of output coverages and `<split_item>` values
    - Up to 50 output coverages can be specified.
    - The `<split_cover>` must have polygon topology.

`Arc: split maptile index tilename poly 1.0`

When done entering coverages, type END or a blank line.

Enter the 1st coverage: tilesplit1
Enter item value: tile3
Enter the 2nd coverage: tilesplit2
Enter item value: tile4
Enter the 3rd coverage: end
<table>
<thead>
<tr>
<th>Before SPLIT</th>
<th>After SPLIT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coverage EXCOV</strong></td>
<td><strong>OUTCOV1 (EXCOV split by TILE3)</strong></td>
</tr>
<tr>
<td><strong>TILE1</strong></td>
<td><strong>TILE2</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TILE3</strong></td>
<td><strong>TILE4</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Split Coverage by INDEX</strong></td>
<td><strong>OUTCOV2 (EXCOV split by TILE4)</strong></td>
</tr>
</tbody>
</table>
UPDATE

- **UPDATE** - replaces the input coverage areas with the update coverage polygons using a cut-and-paste operation.
  - UPDATE \(<\text{in\_cover}>\) \(<\text{update\_cover}>\) \(<\text{out\_cover}>\) \{POLY | NET\} \{fuzzy\_tolerance\} \{KEEPBORDER | DROPBORDER\}
    » \{KEEPBORDER | DROPBORDER\} - specifies whether or not the outside border of the \(<\text{update\_coverage}>\) will be kept after it is inserted into the \(<\text{in\_cover}>\).
  - UPDATE uses the updating extent in a ‘**cut-and-paste**’ operation; update coverage features replace the area they overlap in the input coverage. The result is stored in the output coverage.
  - Both the input and update coverages must have polygon topology.
  - Topology is rebuilt for the output coverage.
  - Attributes are also updated. Items in the PAT are merged using the old internal number of each polygon.
UPDATE, page 2

KEEPBORDER option

Input Coverage

Update Coverage

Output Coverage

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>
UPDATE, page 3

DROPBORDER option

Input Coverage

Update Coverage

Output Coverage

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>D</td>
</tr>
</tbody>
</table>
Polygon Overlay with Attribute Merging

- There are three commands that perform polygon overlay that merge attribute data: UNION, INTERSECT, and IDENTITY

- UNION combines all the features of both coverages
- INTERSECT Only those features in the area common to both coverages will be preserved in the output coverage. Any data that lie outside the common area are deleted (clipped) from the output coverage.
- IDENTITY All features of the input coverage, as well as those features of the identity coverage that overlap the input coverage, are preserved in the output coverage.
Union

- UNION - computes the geometric intersection of two polygon coverages. All polygons from both coverages will be split at their intersections and preserved in the output coverage.

- UNION <in_cover> <union_cover> <out_cover> {fuzzy_tolerance} {JOIN | NOJOIN}
  - {JOIN | NOJOIN} - specifies whether all items in both the <in_cover> PAT and <union_cover> PAT will be joined into the output coverage PAT.
INTERSECT

- INTERSECT - computes the geometric intersection of two coverages. Only those features in the area common to both coverages will be preserved in the output coverage.

- INTERSECT <in_cover> <intersect_cover> <out_cover> {POLY | LINE | POINT} {fuzzy_tolerance} {JOIN | NOJOIN}
IDENTITY

• **IDENTITY** - computes the geometric intersection of two coverages. All features of the input coverage, as well as those features of the identity coverage that overlap the input coverage, are preserved in the output coverage.

• **IDENTITY** <in_cover> <identity_cover> <out_cover> {POLY | LINE | POINT} {fuzzy_tolerance} {JOIN | NOJOIN}
Solving Problems Through Spatial Analysis

• Extract features from a coverage to create a new coverage
• Create zones around features
• Combine coverages to create new data relationships
• Develop models
• Identify trends
• Make better decisions
Typical Steps in Spatial Analysis

- Establish analysis objectives and criteria
- Prepare data for spatial operations
- Perform spatial operations
- Prepare derived data for tabular analysis
- Perform tabular analysis
- Evaluate and interpret results
- Refine the analysis as necessary
- Produce final maps and tabular reports of the results
Spatial Analysis Example, page 1

• Your firm has been contracted to identify sites that are suitable for a landfill. An ideal site would meet the following criteria:
  – At least 1/4 mile from any local street
  – Soil suitable for landfill development
  – Zoned industrial or agricultural
  – Vacant or agricultural land use
  – Not prone to flooding
  – Not within a fault zone
  – Area greater than 1,400,000 square feet

• Coverages available:
  – 1. streets
  – 2. soil
  – 3. zoning
  – 4. landuse
  – 5. flood
  – 6. faultzone
Spatial Analysis Example, page 2

• STEP 1: BUFFER streets for 1/4 mile (1320 feet)
• STEP 2: UNION the street buffer coverage with soils, flood, zoning, landuse, and faultzone (all polygon on polygon with join). Add an item to the final coverage called suitable.
• STEP 3: RESELECT (in Arcplot) from the resulting coverage, buffer item INSIDE = 1 (to select areas outside the buffer polygon), soiltype = clay, flood = outside, zone = industrial or zone = agricultural, landuse = vacant or landuse = agricultural, and faultzone = outside.
• STEP 4: Once the polygons are selected, calculate suitable to be equal to 1. Combine polygons for which suitable = 1 (dissolve on suitable). Reselect polygons with area>1400K
Spatial Analysis Example #2

• **Identify suitable sites (from 6 possible locations) for a branch bank:**
  – more than 10,000 population within 2 miles
  – no other bank within 1 mile
  – on a parcel adjacent to a major thoroughfare

• **Available coverages:**
  – 6parcels
  – census
  – banks
  – roads
Spatial Analysis Example #3

• Design a sensitive habitat preservation zone for the purple-throated tutsi bird (very rare), which needs:
  – grassland, area over 1,000 acres
  – adjacent to forest, area over 500 acres
  – no major road within 5 miles.

• What data do I need?
Spatial Matching: 
**Clipping and Erasing**
(sometimes referred to as *spatial extraction*)

- **CLIP** - extracts those features from an input coverage that overlap with a clip coverage. This is the most frequently used polygon overlay command to extract a portion of a coverage to create a new coverage.

- **ERASE** - erases the input coverage features that overlap with the erase coverage polygons.
Example: Spatial Matching via

*Polygon-on-Polygon Overlay: Union*

The two themes (land use & drainage basins) do not have common boundaries. GIS creates combined layer with all possible combinations, permitting calculation of land use by drainage basin.
Implementing Spatial Matching in ArcGIS 8

Available in three places
• via Selection/Select by Location
  – this selects features of one layer(s) which relate in some specified spatial manner to the features in another layer
  – if desired, selected features may be saved later to a new theme via Data/Export Data
• via Spatial Join (right click layer in T of C, select Join/Joins and Relates, then click down arrow in first line of Join Data window---see Joining Data in Help for details)
  – Use for: points in polygon
    lines in polygon
    points on lines (to calculate distance to nearest line)
    points on points (to calculate distance to “nearest neighbor” point)
• via Tools/Geoprocessing Wizard
  – Creates a new layer (e.g. shape file) & combines attribute tables from 2 or more input themes
  – Five options available for different types of matching (see below)

Options in Geoprocessing Wizard
• Dissolve features based on an attribute
  – Use for spatial aggregation/dissolving
• Merge layers together
  – Use for edge matching
• Clip one layer based on another
  – Use one theme to limit features in another theme
    (e.g. limit a Texas road theme to Dallas county only)
• Intersect two layers (extent limited to common area)
  – Use for polygon on polygon overlay
• Union two layers (covers full extent of both layers)
  – Use for polygon on polygon overlay

--Selection: simply selects (“highlights”) entire spatial features in the target layer, but doesn’t modify these features
--joins: operate on tables and normally creates a new table with additional variables, but again does not modify spatial features themselves
--geoprocessing wizard: modifies geographic features thus creates new spatial file
Spatial Operations:

*neighborhood analysis/spatial filtering*

- **spatial convolution or filter**
  - applied to one raster layer
  - value of each cell replaced by some function of the values of itself and the cells (or polygons) surrounding it
  - can use ‘neighborhood’ or ‘window’ of any size
    - 3x3 cells (8-connected)
    - 5x5, 7x7, etc.
  - differentially weight the cells to produce different effects
  - **kernel** for 3x3 mean filter:
    - \[
    \begin{array}{ccc}
    1/9 & 1/9 & 1/9 \\
    1/9 & 1/9 & 1/9 \\
    1/9 & 1/9 & 1/9 \\
    \end{array}
    \]
    - weights must sum to 1.0

- **low frequency (low pass) filter:**
  - **mean filter**
    - cell replaced by the mean for neighborhood
    - equivalent to weighting (multiplying) each cell by \(1/9 = .11\) (in 3x3 case)
    - smooths the data
    - use larger window for greater smoothing
  - **median filter**
    - use median (middle value) instead of mean
    - smoothing, especially if data has extreme value outliers
Spatial Operations:
spatial filtering -- high pass filter

high frequency (high pass) filter

\textit{negative weight filter}

– exaggerates rather than smooths local detail
– used for edge detection

\textbf{standard deviation filter}
\textit{(texture transform)}

– calculate \textit{standard deviation} of neighborhood raster values
– high SD=high texture/variability
– low SD=low texture/variability
– again used for edge matching
– neighborhoods spanning border have large SD ‘cos of variability

\begin{align*}
\text{cell values (} v_i \text{) on each side of edge} & \\
\text{filtered values for highlighted pixel} & \\
2 & 5 \\
1(5)(9)+5(5)(-1)+3(2)(-1) = 14 \\
1(2)(9)+5(2)(-1)+3(5)(-1) = -7 \\
1(2)(9)+8(2)(-1) = 2 \\
1(5)(9)+8(5)(-1) = 5 \\
f_i \cdot v_i \cdot w_i & \\
-1 & -1 & -1 \\
-1 & 9 & -1 \\
-1 & -1 & -1
\end{align*}
Spatial Operations:  
*raster–based modelling*

- **Relating multiple rasters**
- **Processes may be:**
  - Local: one cell only
  - Neighborhood: cells relating to each other in a defined manner
  - Zonal: cells in a given section
  - Global: all cells
- **ArcGIS implementation:**
  - All raster analyses require either the Spatial Analyst or 3-D Analyst extensions (extra cost)
  - Base ArcView can do no more than display an image (raster) data set

- **Suitability modelling**
  \[
  \text{soil} + \text{slope} + \text{for sale} = \text{Site options}
  \]

- **Diffusion Modelling**
  \[
  \text{Incidence matrix} \times \text{Probability mask} = \text{System at time t+1}
  \]

- **Connectivity Modelling**
  \[
  \text{Initial State} + \text{Connectivity matrix} = \text{Resultant State}
  \]
Attribute Operations: record selection or extraction

Tabular (select by attributes)
- Independent selection by clicking table: right click in TofC, select Open Attribute Table & click on grey selection box at start of row; hold ctrl key for multiple rows)
- Create SQL query: use Selection/Select by Attribute
- use table relates /joins to select specific data

Graphic (select by location)
- with manual cursor on map (use Select Features tool)
  - at a point
  - within a rectangle
  - within a radius (circle) around a point
- By using another layer (spatial extraction)

Hot Link
- Click on map to ‘hot link’ to pictures, graphs, or other maps

Outputs may be:
- Simultaneously highlighted records in table, and features on map
- New tables and/or map layers

Examples
- Use SQL query to select all zip codes with median incomes above $50,000 (tabular)
- identify zip codes within 5 mile radius of several potential store sites and sum household income (graphic)
- show houses for sale on map, and click to obtain picture and additional data on a selected house (hot link)
Attribute Operations: variable recoding

- establishing/modifying number of classes and/or their boundaries for continuous variable. Options for ArcGIS
  - natural breaks (default) (finds inherent inherent groups via Jenks optimization which minimizes the variances within each of the classes).
  - quantile (classes contain equal number of records--or equal area under the frequency distribution)
  - equal interval (user selects # of classes) (equal width classes on variable)
  - Defined interval (user selects width of classes) (equal width classes on variable)
  - standard deviation (categories based on 1,2, etc, SDs above/below mean)
  - Manual (user defined)
    » whole numbers (e.g. 2,000)
    » meaningful to phenomena (e.g zero, 32°)

- aggregating categories on a nominal (or ordinal) variable
  - pine and fir into evergreen

No change in number of records (observations).
Attribute Operations:

*record aggregation*

- combining two or more records into one, based on common values on a key variable
- the attribute equivalent of regionalization or classification
- equivalent of PROC SUMMARY in SAS
- interval scale variables can be aggregated using mean, sum, max, min, standard deviation, etc. as appropriate
- ordinal and nominal require special consideration
- example: aggregate county data to states, or county to CMSA

Record count decreases (e.g. from 12 to 2)

<table>
<thead>
<tr>
<th>Fips</th>
<th>PMSA_90</th>
<th>PMSA_93</th>
<th>Pop90</th>
<th>Pop95_est</th>
<th>Pop90-95%</th>
<th>MedInc89</th>
<th>Suburb</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>48085</td>
<td>1920</td>
<td>1920</td>
<td>264036</td>
<td>346232</td>
<td>5.93</td>
<td>46020</td>
<td>1</td>
<td>Collin</td>
</tr>
<tr>
<td>48113</td>
<td>1920</td>
<td>1920</td>
<td>1852810</td>
<td>1959281</td>
<td>1.09</td>
<td>31605</td>
<td>0</td>
<td>Dallas</td>
</tr>
<tr>
<td>48121</td>
<td>1920</td>
<td>1920</td>
<td>273525</td>
<td>334070</td>
<td>4.22</td>
<td>36914</td>
<td>1</td>
<td>Denton</td>
</tr>
<tr>
<td>48139</td>
<td>1920</td>
<td>1920</td>
<td>85167</td>
<td>94223</td>
<td>2.03</td>
<td>30553</td>
<td>1</td>
<td>Ellis</td>
</tr>
<tr>
<td>48213</td>
<td>1920</td>
<td>1920</td>
<td>58543</td>
<td>64293</td>
<td>1.87</td>
<td>20747</td>
<td>1</td>
<td>Henderson</td>
</tr>
<tr>
<td>48231</td>
<td>1920</td>
<td>1920</td>
<td>64343</td>
<td>66972</td>
<td>0.78</td>
<td>25317</td>
<td>1</td>
<td>Hunt</td>
</tr>
<tr>
<td>48257</td>
<td>1920</td>
<td>1920</td>
<td>52220</td>
<td>60114</td>
<td>2.88</td>
<td>27280</td>
<td>1</td>
<td>Kaufman</td>
</tr>
<tr>
<td>48397</td>
<td>1920</td>
<td>1920</td>
<td>25604</td>
<td>32725</td>
<td>5.30</td>
<td>42417</td>
<td>1</td>
<td>Rockwall</td>
</tr>
<tr>
<td>48221</td>
<td>2800</td>
<td>2800</td>
<td>28981</td>
<td>33384</td>
<td>2.89</td>
<td>31627</td>
<td>1</td>
<td>Hood</td>
</tr>
<tr>
<td>48251</td>
<td>2800</td>
<td>2800</td>
<td>97165</td>
<td>106181</td>
<td>1.77</td>
<td>30612</td>
<td>1</td>
<td>Johnson</td>
</tr>
<tr>
<td>48367</td>
<td>2800</td>
<td>2800</td>
<td>64785</td>
<td>73794</td>
<td>2.65</td>
<td>30592</td>
<td>1</td>
<td>Parker</td>
</tr>
<tr>
<td>48439</td>
<td>2800</td>
<td>2800</td>
<td>1170103</td>
<td>1278606</td>
<td>1.77</td>
<td>32335</td>
<td>0</td>
<td>Tarrant</td>
</tr>
</tbody>
</table>

Source: US Bureau of the Census

MedInc=Median Household Income. Pop90 as of April 1. Pop95 as of July 1.

Type of processing: sum, sum re-calc. → count average of medians!
Attribute Operations: *Joining and Relating Tables*
associating spatial layer to non-spatial table

**Join:** one to one, or one to many, relationship, appends attributes

Associate table of country capitals with country layer: only one capital for each country (one to one)

<table>
<thead>
<tr>
<th>Country Code</th>
<th>Country</th>
<th>Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>France</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>Saudi Arabia</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Chad</td>
<td></td>
</tr>
<tr>
<td>248</td>
<td>Spain</td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>Venezuela</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>UK</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Philippines</td>
<td></td>
</tr>
</tbody>
</table>

Layer Attribute Table  | NonSpatial Table  | Layer Attribute Table after Join

Associate country layer with type of government: one gov. type assigned to many countries--but each country has only one gov. type (one to many)

<table>
<thead>
<tr>
<th>Gov. Code</th>
<th>Country</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>France</td>
<td>Republic</td>
</tr>
<tr>
<td>30</td>
<td>Vietnam</td>
<td>Communist State</td>
</tr>
<tr>
<td>15</td>
<td>UK</td>
<td>Const. Monarchy</td>
</tr>
<tr>
<td>20</td>
<td>Argentina</td>
<td>Communist State</td>
</tr>
<tr>
<td>10</td>
<td>Saudi Arabia</td>
<td>Absolute Monarchy</td>
</tr>
<tr>
<td>15</td>
<td>Sweden</td>
<td>Const. Monarchy</td>
</tr>
<tr>
<td>45</td>
<td>Portugal</td>
<td>Parliamentary Democracy</td>
</tr>
</tbody>
</table>

Layer Attribute Table  | NonSpatial Table  | Layer Attribute Table after Join
Attribute Operations: *Joining and Relating Tables*

*associating spatial layer to non-spatial table (contd.)*

**Relate:** many to one relationship, attributes not appended

Associate country layer with its multiple cities (many to one)

<table>
<thead>
<tr>
<th>Country Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>France</td>
</tr>
<tr>
<td>68</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>106</td>
<td>Chad</td>
</tr>
<tr>
<td>248</td>
<td>Spain</td>
</tr>
<tr>
<td>199</td>
<td>Venezuela</td>
</tr>
<tr>
<td>9</td>
<td>UK</td>
</tr>
<tr>
<td>96</td>
<td>Philippines</td>
</tr>
</tbody>
</table>

Layer Attribute Table

<table>
<thead>
<tr>
<th>Country Code</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>129</td>
<td>Mombasa</td>
</tr>
<tr>
<td>129</td>
<td>Nairobi</td>
</tr>
<tr>
<td>29</td>
<td>Paris</td>
</tr>
<tr>
<td>29</td>
<td>Lyon</td>
</tr>
<tr>
<td>29</td>
<td>Marseille</td>
</tr>
<tr>
<td>60</td>
<td>Katmandu</td>
</tr>
<tr>
<td>248</td>
<td>Madrid</td>
</tr>
<tr>
<td>248</td>
<td>Barcelona</td>
</tr>
<tr>
<td>248</td>
<td>Valencia</td>
</tr>
</tbody>
</table>

NonSpatial Table

If *joined* Paris to France, for example, we lose Lyon and Marseille, therefore use *relate*

**Note:** if we flip these tables we can do a join since there is only one country for each city (one to many)

**For both Joins and Relates:**
- Association exists only in the map document
- Underlying files not changed unless export data