APPLICATIONS: DATA EXPLORATION

This applications section covers five tasks. Task 1 presents an overview of data exploration in ArcView, including use of summary statistics, selecting features by graphics, and making charts. Both Tasks 2 and 3 deal with attribute data query. Task 4 combines spatial and attribute data queries. Task 5 covers raster data query.

DOWNLOAD "DATA EXPLORATION" DATA FROM WEBSITE

Task 1. An Overview of Data Exploration in ArcView

What you need: *idcities.shp*, with 654 places in Idaho; *idcounty.shp*, showing county boundaries in Idaho; and *snowsite.shp*, with 206 snow stations in Idaho and the surrounding states.

Task 1 is loosely organized around the question of finding a good skiing spot within 40 miles of Sun Valley, Idaho. This task is intended to familiarize you with several aspects of data exploration using ArcView: getting summary statistics, selecting map features by graphic, and creating charts. To keep it organized, each aspect of data exploration is included in a sub-section.

- **Derive summary statistics**

1. Start ArcView, open a new view, and add *idcities.shp*, *idcounty.shp*, and *snowsite.shp* to view. All three themes are based on the Idaho statewide coordinate system and measured in meters. Select Properties from the View menu. In the View Properties dialog, set the Map Units as meters and the Distance Units as miles.
2. Make *snowsite.shp* active, and choose Table from the Theme menu to open the attribute table of *snowsite.shp*. In the *snowsite.shp* theme table, click on the field Swe-max to make it active. Then select Statistics from the Field menu. Swe-max stands for maximum snow water equivalent.

- **Select map features with graphics**

1. Make *idcities.shp* the active theme. Click the Query Builder button. Set the query expression as [City_name] = "Sun Valley" and click New Set. Use the Zoom To Select button and then the Zoom In button to zoom in the area around Sun Valley.
2. Press and hold the Draw Point tool, and choose Draw Circle from the pull-down menu.
3. This step is to draw a circle around Sun Valley and to use the circle as a graphic to select snow stations within the circle. Click on Sun Valley as closely as possible at its location and drag the cursor to make a circle with a radius of 40 miles. The message at the bottom of the computer screen shows you how closely the circle radius is to 40 miles. If you cannot get exactly 40 miles interactively, you can select Size/Position from the Graphics menu and type 40 miles as the radius value.
4. Now select snow stations that fall within the circle. Make sure *snowsite.shp* is active, and click the Select Features Using Graphics button. Those selected snow stations are highlighted.
5. Select Table from the Theme menu. Click Promote to move to the top of the table the highlighted records, that is, records for the selected snow stations.

- **Make charts**

1. While the *snowsite.shp* table with its selected records is still active, click the Create Chart button. This part of Task 1 shows you how you can make charts using attribute data of the selected snow stations from the previous section.
2. In the Chart Properties dialog, do the following: rename the chart as Swe-max, click Swe-max as the Field, click Add to move Swe-max to the Group, change Label Series to Swe-max, and click OK. The chart

is basically a histogram depicting the maximum snow water equivalent of each selected snow station. Make another chart using the field of Elev. Elev lists the elevation of the snow station in feet.

3. To see the relationship between Swe-max and Elev, you can make a scatterplot. Click the Create Chart button to open a new chart. Add both Swe-max and Elev to Groups.
4. Click the *xy* Scatter Chart Gallery button to open options of scatterplots. Select the option in the upper left with the linear scaling of *x* and *y* and click OK. The other options are based on the logarithmic scaling of *x*, or *y*, or both *x* and *y*. The scatterplot shows a positive relationship between Swe-max and Elev: the higher the snow station is, the more the maximum snow water equivalent is expected.

Task 2. Attribute Data Query

What you need: *wp.shp*, a vegetation stand coverage; *wpdata.dbf*, a dBASE file containing stand data for *wp.shp*.

As explained in the text, query or data selection is the most important element of data exploration. Approached from either attribute data or spatial data, results of query are displayed in the linked windows of view, table, and chart. Task 2 focuses on attribute data query.

1. Start ArcView, open a new view, and add *wp.shp* to view. Open the theme table of *wp.shp*.
2. Activate the Project window. Click Tables and Add to open the Add Table dialog. Make sure the file type is dBASE. Double-click on *wpdata.dbf* to select it.
3. At this point, you have opened two tables: the theme table (Attributes of *wp.shp*) and *wpdata.dbf*. To join the data from *wpdata.dbf* to the theme table, do the following: click on Id in *wpdata.dbf*, click on Id in the *wp.shp* theme table, and then click the Join button to join the two tables. Id is the key relating the two tables.
4. Make sure the *wp.shp* theme table includes attribute data from *wpdata.dbf* and is active. Click the Query Builder button to open the Query Builder dialog. The top part of the dialog box shows, from left to right, fields in the attributes of *wp.shp* table, logical operators and Boolean connectors, and values of the selected field. Notice that the name of each field is enclosed in a pair of square brackets. The bottom part of the dialog box has the display area of logical expressions on the left and three buttons for different query methods on the right. The buttons are New Set, Add To Set, and Select From Set. New Set selects a new data subset from the theme table. Add To Set adds a new data subset to the previously selected records. Select From Set selects a new data subset from the previously selected records.
5. Double-click the field of Origin in the Query Builder dialog, click the > operator, and double-click the value of 0. A logical expression, ([Origin] > 0), is now shown in the display area. This is the first logical expression. Click the connector AND, double-click the field of Origin, click the <= operator, and type 1900 to complete the second logical expression. The completed logical expressions should read: ([Origin] > 0) AND ([Origin] <= 1900). Now click on New Set. Records in the theme table that satisfy the logical expressions are highlighted. The upper left corner of the ArcView window shows "175 of 856 selected." Do not dismiss the Query Builder dialog because you will use it again for refining the query operation.
6. The field Origin represents the origin of trees in a stand, expressed in the year trees were planted. The value 0 means that the origin is unknown. Therefore, the logical expressions in step 5 selected stands with trees known to be at least 100 years old. Click the Promote button to bring the selected records to the top of the theme table. Examine the Origin

values of the selected records to see if any of them are after 1900. Now view the map. Highlighted polygons correspond to the selected records.

7. Finally, narrow the selected records by including aspect as an additional criterion. Return to the Query Builder dialog. Drag and highlight the logical expressions between the outer parentheses and delete them. Construct the following logical expressions: ([As] = "N") OR ([As] = "NE") OR ([As] = "NW"). Then click on Select From Set. The number of records selected, as shown in the upper left corner of the ArcView window, is reduced from 175 to 44. The reduced data subset shows only old-growth stands that have the aspects of north, northeast, and northwest. Again, you can verify that the selected records do meet both the origin and aspect criteria. View the map to see where those stands are located.

Task 3. Relational Database Query

What you need: *mosoils.shp*, a soil coverage; *comp.dbf*, *forest.dbf*, and *plantnm.dbf*, three dBASE files from the National Map Unit Interpretation Record (MUIR) database maintained by the Natural Resources Conservation Service (NRCS).

Task 3 lets you work with the MUIR database. By linking the tables in the database properly, you can explore many soil attributes in the database from any table. And, because the tables are linked to the soil map, you can also see where selected records are located.

1. Start ArcView, open a new view, and add *mosoils.shp* to view. Open the theme table of *mosoils.shp*.
2. Next, add the dBASE files to the computer screen. Activate the Project window. Click on Tables and Add. Navigate to the three dBASE files and add them as tables. You should now have four tables and the soil map on the monitor. Arrange them so that you can work with each one of them.
3. The next step is to link the tables. The idea is to keep the four tables separate but dynamically linked rather than joined. To link two tables, you need to know which fields to use as keys. As illustrated in the chapter, musym can link the theme table and *comp.dbf*, muid can link *comp.dbf* and *forest.dbf*, and plantsym can link *forest.dbf* and *plantnm.dbf*. Linking tables is directional, that is, from the source table to the destination table. In data exploration, you want to be able to search soil attributes from any table. Therefore, you need to perform link twice between every two tables. To link the theme table to *comp.dbf*, click on musym in the theme table and musym in *comp.dbf*, and select Link from the Table menu. Then, repeat the same process in the opposite direction: click on musym in *comp.dbf*, musym in the theme table, and select Link from the Table menu. Now, you have completed the two-way linking between the theme table and *comp.dbf*. Do the same between *comp.dbf* and *forest.dbf*, and between *forest.dbf* and *plantnm.dbf*.
4. At this point, the four tables are linked in both directions. The chapter asked a question about what types of plants are found in areas where annual flooding frequency is rated as either frequent or occasional. You can now answer the question by doing the following. Make *comp.dbf* active. Click on the Query Builder button. In the Query Builder dialog, prepare the query expression as ([Anflood] = "FREQ") OR ([Anflood] = "OCCAS"), and click New Set. Records in *comp.dbf* that meet the criteria are highlighted as the corresponding records in the other three tables and the corresponding polygons in the soil map. You can now see where areas with frequent or occasional flooding are located, the forest types in these areas, and the common names of plant species.
5. You can try another query with the tables. You probably want to first clear the selected records by clicking the Select None button. Now, make *plantnm.dbf* active and click the

Query Builder button. Prepare the query statement as ([Comname] = "lupine"). Click New Set. The selected record in *plantnm.dbf* and its corresponding records in the other tables are highlighted. You can also see where lupine can be found on the map.

Task 4. Combining Spatial and Attribute Data Queries

What you need: *thermal.shp*, a coverage with 899 thermal wells and springs; *idroads.shp*, showing major roads in Idaho.

Task 4 assumes that you are asked by a company to locate potential sites for a hot spring resort in Idaho. You are given two criteria for selecting potential sites:

- The hot spring must be within 2 miles of a major road.
- The temperature of the hot spring must be greater than 60°C.

The field type in *thermal.shp* uses *s* to denote springs and *w* to denote wells. The field temp shows the water temperature in °C.

1. Start ArcView, open a new view, and add *thermal.shp* and *idroads.shp* to view. Select Properties from the View menu. In the View Properties dialog, set the Map Units as meters and the Distance Units as miles. Both *thermal.shp* and *idroads.shp* have meters as the map units.
2. Activate *thermal.shp* theme. Click Select By Theme from the Theme menu. In the dialog, set the query statement to read: "Select features of active themes that Are Within Distance Of the selected features of *idroads.shp*," and the Selection Distance to be 2 miles. Click New Set. Those highlighted thermal springs and wells are within 2 miles of major roads in Idaho.
3. Next, narrow the selection of map features by using the second criterion. Select Tables from the Theme menu. Use Promote to move the selected records to the top of the table. Click on the Query Builder button. Prepare the query expression as: ([Type] = "s") AND ([Temp] > 60). Because you want to select from the previously selected records, click on Select From Set. Dismiss the Query Builder dialog.
4. Again, use Promote to move the selected records to the top. The Type value of the 15 selected records should be all *s* for springs, and the Temp value should be all above 60. In fact, one of the selected records has the name of "Zim's Resort," a hot spring that has already been developed into a resort area. The map shows you where those 15 hot springs are located.
5. As explained in the chapter, this task can also be solved by first selecting hot springs with water temperatures above 60°C through attribute data query and then selecting those springs that are within 2 miles of major roads through spatial data query. The final answer should be the same.

Task 5. Raster Data Query

What you need: *slope_gd*, a slope grid; and *aspect_gd*, an aspect grid.

Task 5 shows you different methods for querying a single grid or multiple grids.

1. Start ArcView, and load Spatial Analyst. Open a new view, and add *slope_gd* and *aspect_gd* to view. *Slope_gd* has the following slope classes in degree: 1 (0–10), 2 (10–20), 3 (20–30), and 4 (30–40). *Aspect_gd* has the following aspect classes: 1 (flat), 2 (north), 3 (east), 4 (south), and 5 (west).
2. Select Properties from the View menu, and set the Map Units to meters and the Distance Units to kilometers. First query *slope_gd* using the graphic method. Click the Draw Circle tool. Then click a point in *slope_gd* and drag the cursor to make a circle with a radius of 1.5 kilometers. If you cannot get exactly 1.5 kilometers interactively, select Size/Position from the Graphics menu and type 1.5 kilometers as the radius value.

3. Click on the Histogram button. The histogram shows the cell values and their frequencies within the circular area. You can also click on the Identify tool and then the bar graph of a cell value to find its exact frequency (count). To remove the circle, make it active and select Delete Graphics from the Theme menu.
4. Map Query is the tool for querying a grid by its cell values. Select Map Query from the Analysis menu. In the *Map Query 1* dialog, set the query expression as: ([Slope_gd] = 2.AsGrid). The request of AsGrid is automatically added to the expression. Click Evaluate. *Map Query 1* in the Table of Contents shows areas with slopes between 10 and 20 degrees as True.
5. Map Query can also query both *slope_gd* and *aspect_gd* to find areas having slopes between 10 and 20 degrees and the south aspect. Select Map Query from the Analysis menu. In the *Map Query 2* dialog, set the query expression as: ([Slope_gd] = 2. AsGrid) AND ([Aspect_gd] = 4.AsGrid). Click Evaluate. *Map Query 2* shows areas that satisfy the logical expression.
6. To save the result of a map query, first activate the output to be saved and then select Save Data Set from the Theme menu. In the next dialog, specify the name and the path for saving the data set.

REFERENCES

Anselin, L. 1999. Interactive Techniques and Exploratory Spatial Data Analysis. In P.A. Longley, M.F. Goodchild, D.J. MaGuire, and D.W. Rhind (eds.). *Geographical Information Systems*, 2d ed. New York: John Wiley & Sons, pp. 253–66.

Batty, M., and Y. Xie. 1994. Modelling Inside GIS: Part I. Model Structures, Exploratory Spatial Data Analysis and Aggregation. *International Journal of Geographical Information Systems* 8: 291–307.

Becker, R.A., and W.S. Cleveland. 1987. Brushing Scatterplots. *Technometrics* 29: 127–42.

Cleveland, W.S. 1993. *Visualizing Data.* Summit, NJ: Hobart Press.

Cleveland, W.S., and M.E. McGill, eds. 1988. *Dynamic Graphics for Statistics.* Belmont, CA: Wadsworth.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*, revised edition. New York: John Wiley & Sons.

DiBiase, D., A.M. MacEachren, J.B. Krygier, and C. Reeves. 1992. Animation and The Role of Map Design in Scientific Visualization. *Cartography and Geographic Information Systems* 19: 201–14, 265–66.

Haslett, J., G. Wills, and A. Unwin. 1990. SPIDER—An Interactive Statistical Tool for The Analysis of Spatially Distributed Data. *International Journal of Geographical Information Systems* 3: 285–96.

MacEachren, A.M. 1995. *How Maps Work: Representation, Visualization, and Design.* New York: The Guilford Press.

Meyer, M.A., F.R. Broome, and R.H.J. Schweitzer. 1975. Color Statistical Mapping by the U.S. Bureau of the Census. *American Cartographer* 2: 100–17.

Monmonier, M. 1989. Geographic Brushing: Enhancing Exploratory Analysis of the Scatterplot Matrix. *Geographical Analysis* 21: 81–84.

Olson, J. 1981. Spectrally Encoded Two-Variable Maps. *Annals, Association of American Geographers* 71: 259–76.

Peterson, M.P. 1995. *Interactive and Animated Cartography.* Englewood Cliffs, NJ: Prentice Hall.

Scott, L.M. 1994. Identification of GIS Attribute Error Using Exploratory Data Analysis. *The Professional Geographer* 46: 378–86.

Tomlin, C.D. 1990. *Geographic Information Systems and Cartographic Modeling.* Englewood Cliffs, N.J.: Prentice Hall.

Tufte, E.R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.