Data Exploration

- Centered on the original data, data exploration allows a researcher to examine the general trends in the data, to take a close look at data subsets, and to focus on possible relationships between data sets.
- Data exploration takes advantage of interactive and dynamically linked visual tools. Maps, graphs, and tables are displayed in multiple windows and dynamically linked so that selecting records from a table will automatically highlight the corresponding features in a graph and a map.

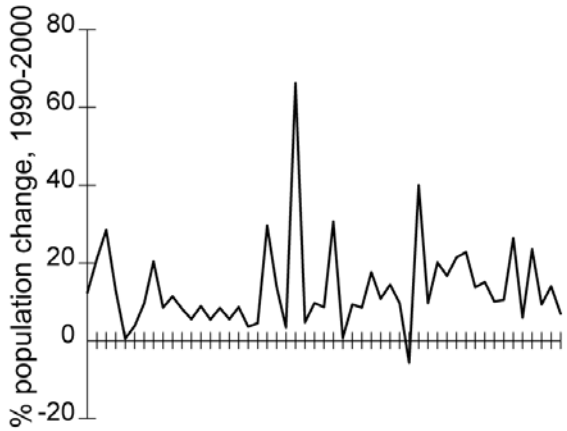


Figure 10.1
A line graph.

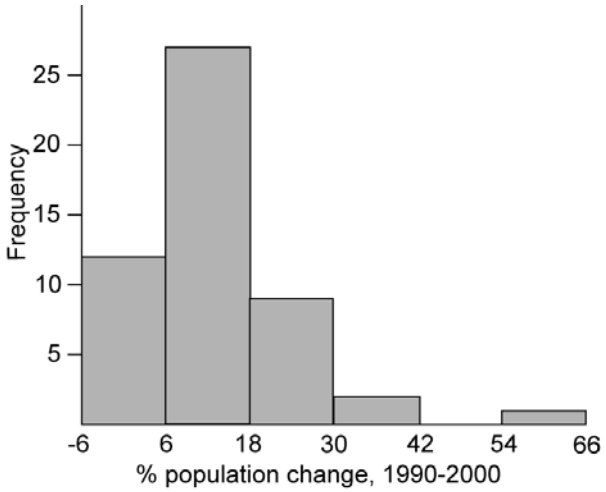


Figure 10.2
A histogram.

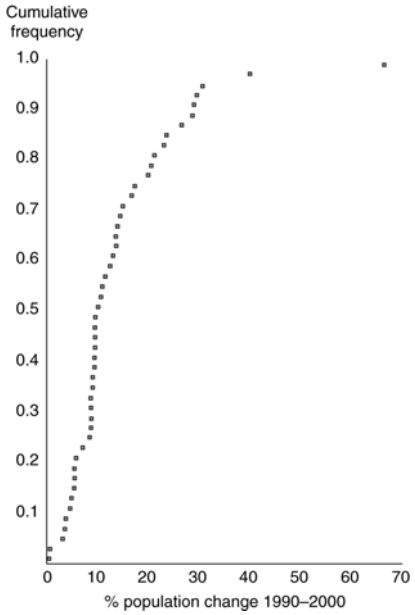


Figure 10.3
A cumulative distribution graph.

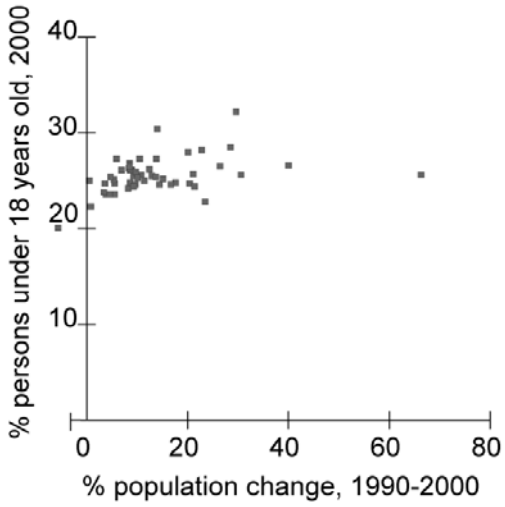


Figure 10.4
A scatterplot plotting % persons 18 years old in 2000 against % population change, 1990–2000. A weak positive relationship, with a correlation coefficient of 0.376, is present between the two variables.

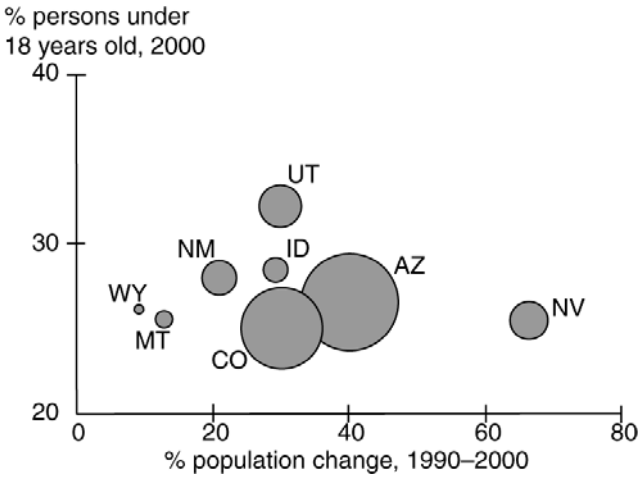


Figure 10.5
A bubbleplot showing % population change, 1990–2000, along the x-axis; % persons under 18 years old in 2000 along the y-axis; and state population in 2000 by the bubble size.

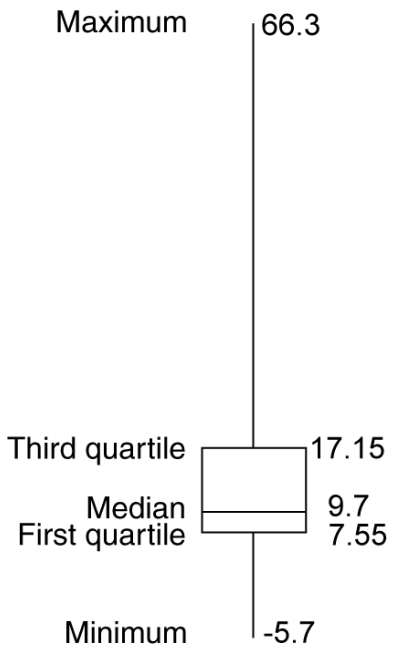


Figure 10.6
A boxplot based on the % population change, 1990–2000, data set.

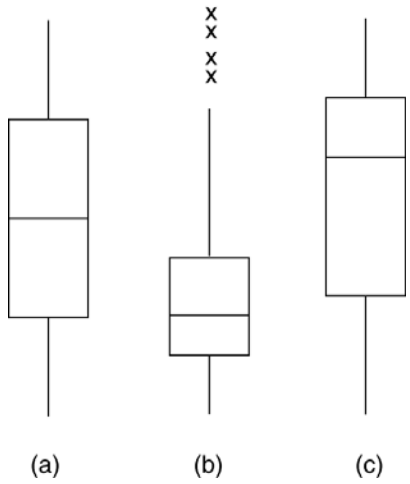


Figure 10.7
Boxplot (a) suggests that the data values follow a normal distribution. Boxplot (b) shows a positively skewed distribution with a higher concentration of data values near the high end. The x's in (b) may represent outliers, which are more than 1.5 box lengths from the end of the box. Boxplot (c) shows a negatively skewed distribution with a higher concentration of data values near the low end.

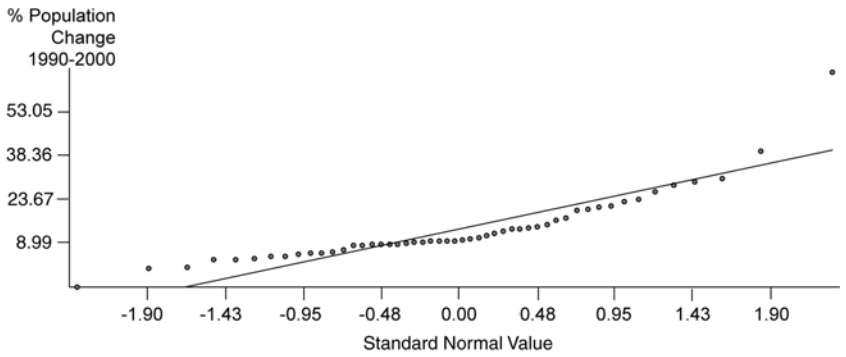


Figure 10.8
A QQ plot plotting % population change, 1990-2000 data value against the standardized value from a normal distribution.

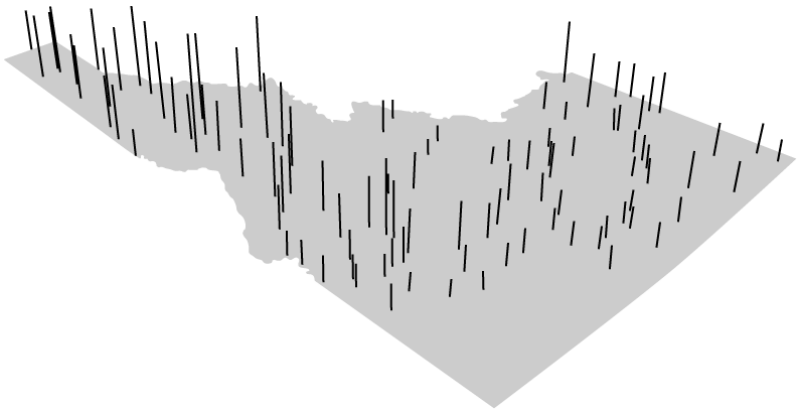


Figure 10.9
A 3-D plot showing annual precipitation at 105 weather stations in Idaho. A north to south decreasing trend is apparent in the plot.

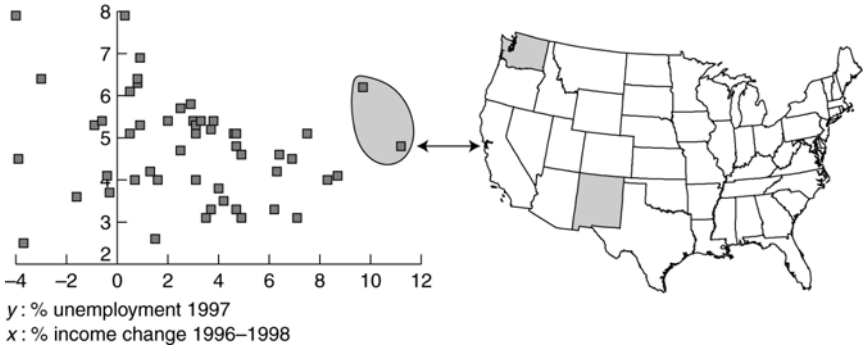


Figure 10.10
The scatterplot on the left is dynamically linked to the map on the right. The "brushing" of two data points in the scatterplot highlights the corresponding states (Washington and New Mexico) on the map.

Map-Based Data Manipulation

- Maps are an important part of GIS operations including data exploration.
- Map-based data manipulation includes data classification, spatial aggregation, and map comparison.

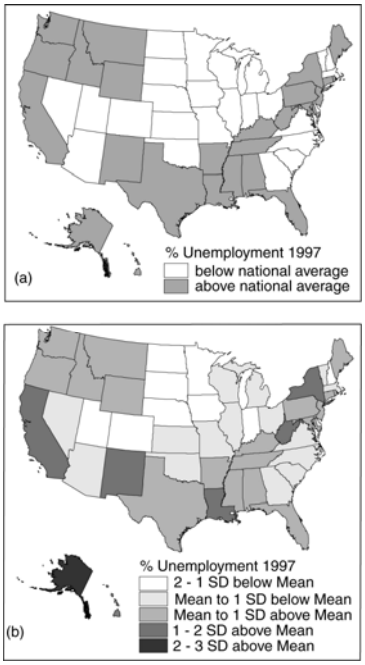


Figure 10.11
Two classification schemes: above or below the national average (a), and mean and standard deviation (SD) (b).

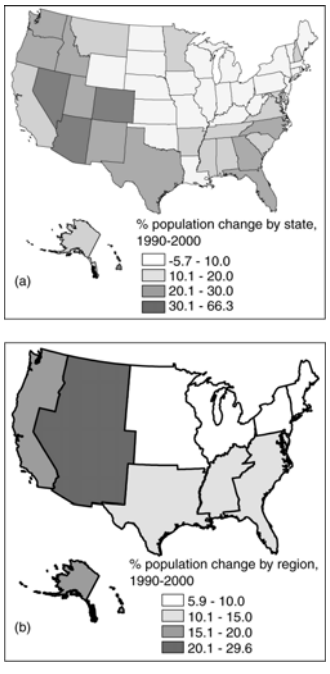


Figure 10.12
Two levels of spatial aggregation: by state (a), and by region (b).

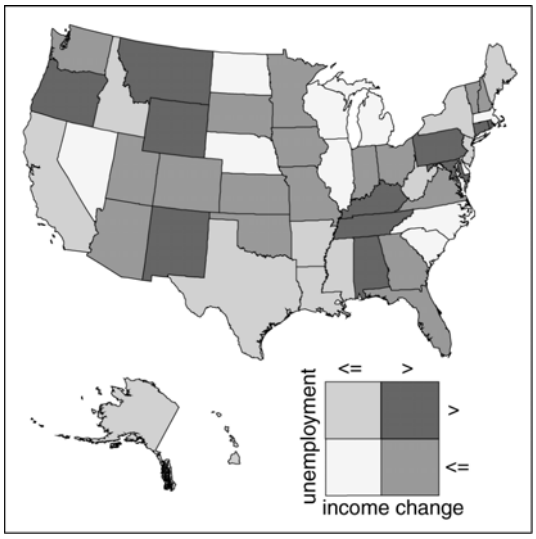


Figure 10.13
A bivariate map: (1) rate of unemployment in 1997, either above or below the national average, and (2) rate of income change, 1996–1998, either above or below the national average.

Attribute Data Query

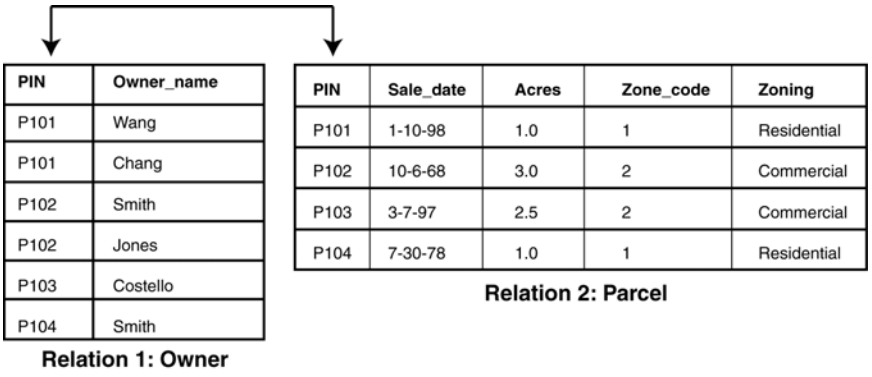
- Attribute data query retrieves a data subset by working with attribute data.
- The selected data subset can be simultaneously examined in the table, displayed in charts, and linked to the highlighted features in the map.
- The selected data subset can also be saved for further processing.

SQL

SQL (structured query language) is a data query language designed for relational databases.

The basic syntax of SQL includes the following: *select* <attribute list>, *from* <relation>, and *where* <condition>.

ArcGIS has already prepared the keywords in the dialog for querying a local database. Therefore, we only have to enter the *where* clause (commonly called the query expression) in the dialog box.



PIN	Owner_name
P101	Wang
P101	Chang
P102	Smith
P102	Jones
P103	Costello
P104	Smith

Relation 1: Owner

PIN	Sale_date	Acres	Zone_code	Zoning
P101	1-10-98	1.0	1	Residential
P102	10-6-68	3.0	2	Commercial
P103	3-7-97	2.5	2	Commercial
P104	7-30-78	1.0	1	Residential

Relation 2: Parcel

Figure 10.14
PIN relates the owner and parcel tables and allows use of SQL with both tables.

Query Expressions

- Query expressions consist of Boolean expressions and connectors.
- A simple Boolean expression contains two operands and a logical operator such as Parcel.PIN = ‘P101’ .
- Boolean connectors are AND, OR, XOR, and NOT, which are used to connect two or more expressions in a query statement.
- Boolean connectors of NOT, AND, and OR are actually keywords used in the operations of Complement, Intersect, and Union on sets in probability.

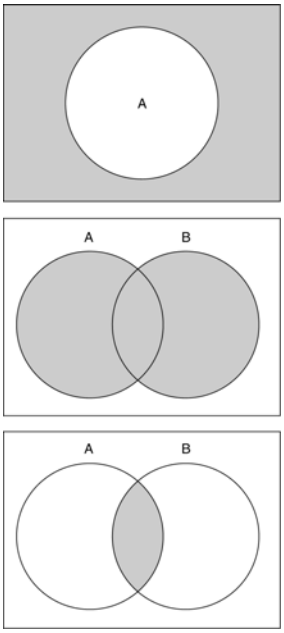


Figure 10.15
The shaded portion represents the complement of data subset A (top), the union of data subsets A and B (middle), and the intersection of A and B (bottom).

Cost	Soiltype	Area	Cost	Soiltype	Area
1	Ns1	500	6	Tn4	300
2	Ns1	500	7	Tn4	200
3	Ns1	400	8	N3	200
4	Tn4	400	9	N3	100
5	Tn4	300	10	N3	100

TABLE 10.1 A Data Set for Query Operation Examples

Type of Operation

Attribute data query begins with a complete data set. A basic query operation is to select a subset. Given a selected data subset, three types of operations can act on it: add more records to the subset, remove records from the subset, and select a smaller subset

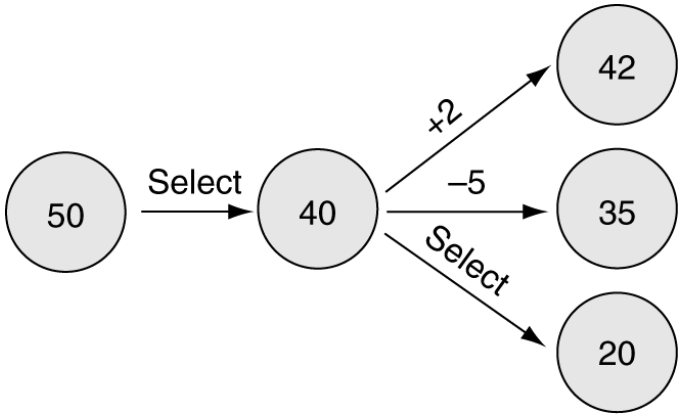


Figure 10.16
Three types of operation may be performed on the subset of 40 records: add more records to the subset (+2), remove records from the subset (-5), or select a smaller subset (20).

Relational Database Query

- Relational database query works with a relational database. A query of a table not only selects a data subset in the table but also selects records related to the subset in other tables.
- To query a relational database, we must be familiar with the overall structure of the database, the designation of keys in relating tables, and a data dictionary listing and describing the fields in each table.

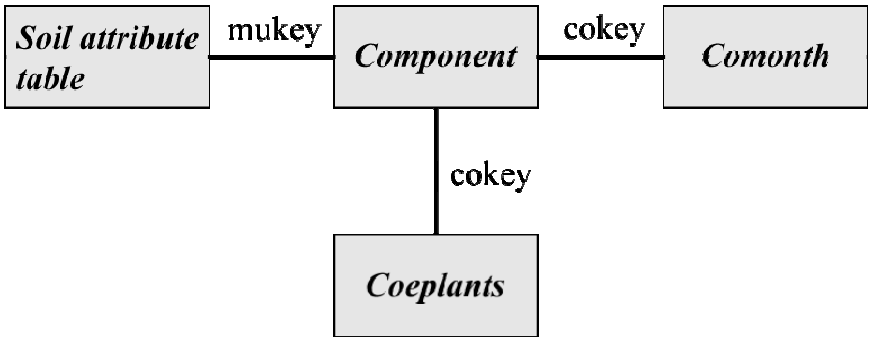


Figure 10.17
The keys relating three dBASE files in the SSURGO database and the soil map attribute table.

Spatial Data Query

- Spatial data query refers to the process of retrieving a data subset from a layer by working directly with features.
- We may select features using a cursor, a graphic, or the spatial relationship between features.
- Spatial relationships used fro query include containment, intersect, and proximity.

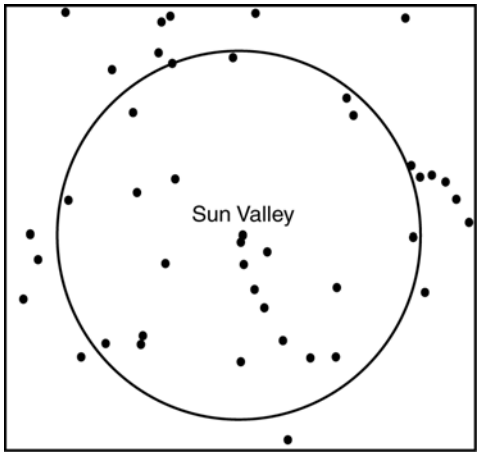


Figure 10.18
Select features by a circle centered at Sun Valley.

Raster Data Query

Raster data can be queried by cell value and select features.

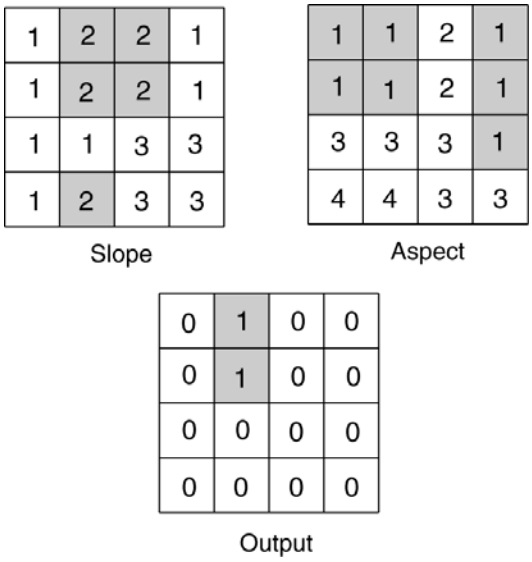


Figure 10.19
Raster data query: slope = 2 and aspect = 1. Selected cells are coded 1 and others 0 in the output raster.

gnuplot
http://www.gnuplot.info/
U.S. Census Bureau
http://www.census.gov/