# Cartographic Symbolization and Design: ARC/INFO® Methods

**Alan Brenner**

The U. S. Environmental Protection Agency
Office of Information Resources Management
National Geographic Information Systems Program

The ARC Macro Language™ and C programs discussed in this guideline are available by anonymous ftp from sdcdg01.sdc.epa.gov and are in the files:

/pub/readme.cart
/pub/map_amls.tar.Z
/pub/map_design.tar.Z
/pub/map_post.Z
/pub/color_post.tar.Z
/pub/baw_post.tar.Z

For users not on the Internet, the programs can be obtained on hardcopy, 3.5 inch disks, QIC 150 or 8mm tape by contacting the National GIS Program at 703-235-5600, or:

401 M ST. SW, MS 3405R
Washington, DC 20460.

## Table of Contents

# List of Figures

# Chapter One
# The Utility of Maps and Graphics

Maps have a wide variety of uses, ranging from the recording of information (such as cadastral maps) to aiding navigation (such as road maps and naval charts). Two of the main uses of maps (and graphics) in environmental analysis, particularly as assisted by geographic information systems, are the analysis of data and the presentation of data. Neither of these uses precludes the other, and both can be aided by good graphic design techniques. Graphic design requires an understanding of the data (and the information represented by the data), an understanding of the methods of visual communication, and the ability to make use of the available means of communication. Although each of these components is described in other places (such as statistics books for data analysis, cartographic texts for visualization of spatial data, and software manuals for the actual production of maps), this guideline integrates these three areas in the context of the Environmental Systems Research Institute's geographic information system, ARC/INFO revision 6, as the available means of communication.

As such, chapter one contains a brief discussion of data and meta-data issues, and visualization and communication theory; chapter two consists of a discussion of issues involved in the design of maps, such as projections and data classification, as accomplished in ARC/INFO; and chapters three through nine present ARC Macro Language programs for specific symbolization techniques for point, line and area data. Chapter ten provides a brief conclusion.

## The Principles of Data

For geographic phenomena, there are two types of data that comprise information about a phenomena: attribute data (the measured characteristic of a location), and location data (the measured location of a characteristic). This pairing of information types is reflected in the two major approaches of geographic information systems: vector systems, which emphasize the attribute; and raster systems, which emphasize the location. ARC/INFO now combines both of these, allowing a wide variety of analysis and mapping, but the appropriate use of this information is still dependent on the analyst/cartographer.

Attribute data can be grouped into several categories--empirical levels--which influence the design requirements for representing information. Each empirical level contains all of the information of lower levels (and thus can be simplified) while adding additional information (see Figure 1.1 on page 2 for a comparison of attribute data levels with spatial characteristics and 'visual variables' for area data). The lowest empirical level is nominal data; this data only indicates that something is different than something else. Names of states are an example of nominal categorization. The next level of empirical data is ordinal data; this data indicates that something is more (or less) than something else, but no quantity can be given to that distance. The terms high, medium and low reflect ordinal differences. The third level of empirical data is interval data; a linear measurement of distance can be used to gauge the differences between instances. The highest level of empirical data is ratio data; this data indicates that something is more (or less) than something else (and thus different), that a linear distance can

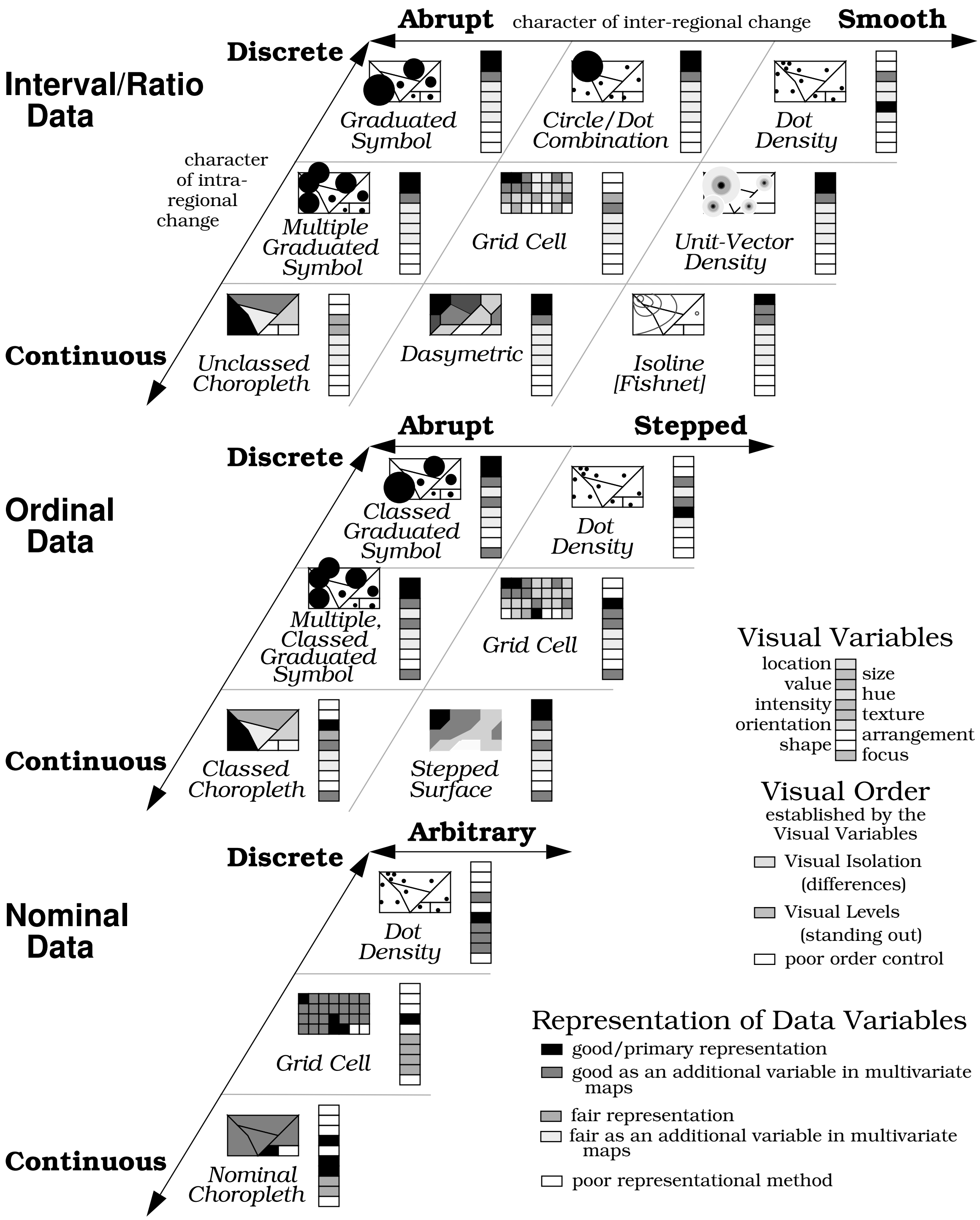



**Figure 1.1** Spatial data models for area data and their cartographic representations; developed from MacEachren and DiBiase (1991) and DiBiase, Krygier, Reeves, MacEachren and Brenner (1991). In a thematic map, the information to be displayed can be categorized by its empirical level (nominal to ratio), how the data changes within any region it is aggregated into, and how the data changes between adjacent regions. Once this categorization is done, visual variables can be selected on the basis of the ability to represent the information so categorized.

be used to gauge differences and that ratio comparisons (such as: this is twice that) can also be done. The essential difference between ratio and interval data is that for ratio data, a non-arbitrary zero point is intrinsic in the measurement; this difference is small enough that, for visualization, the methods used for representing both types of data are the same. An example of the difference between interval and ratio measurement is the difference between temperature measured in degrees Celsius and degrees Kelvin; the Celsius system has an arbitrary zero (the freezing point of water) but the Kelvin system's zero is not arbitrary (absolute zero: point at which the particle motion that is 'temperature,' stops).

Because of these differences in attribute data, care must be taken when comparisons between different levels of data area made (see Table 1.1 for the types of comparisons that can be performed on data of given levels). For example, ordinal data may be stored as integers that represent the order, but the values of these integers do not indicate any measurement of the variability of the data. This lack of measurement does not, however, preclude a software system from acting on the data as if it were ratio data and thereby calculating essentially meaningless statistics (or spatial patterns). If comparisons of data of different levels must be done, either the higher level of data must be reduced to the lower, or a transformation (an addition of information) must be done to raise the lower level to the higher. For nominal or ordinal data, the procedures of psychometrics can be used to recast ordinal information into interval or ratio data. Essentially this involves assigning utility values (such as money) to data levels that do not normally have this type of information associated with it (such as aesthetic values). This can be accomplished by conducting a survey to get individual assignments of value, and then using the collected data to assign overall values to the data.

| | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| Rank Comparison | Invalid | Valid | Valid | Valid |
| Addition, Subtraction | Invalid | Invalid | Valid | Valid |
| Multiplication, Division | Invalid | Invalid | Invalid | Valid |
| Statistics: | | | | |
| Parametric | Invalid | Invalid | Valid | Valid |
| Nonparametric | Valid | Valid | Valid | Valid |

**Table 1.1** Valid data comparisons. Because of the varying degree of numerical precision associated with different data levels (nominal through ratio) only certain operations can be applied to comparisons between two data sets; comparisons between data of two different levels must occur at the lower of the two classes (ESRI *Grid Class Notes* 1992, 11-22).

Although all information is classified due to the nature of measurement uncertainty in the recording of data (MacEachren 1992, 48), attribute data can be further classed into groups for both presentation and analysis (see Classifying Data on page 28 for a discussion of data classification in ARC/INFO). Classification involves the creation of range categories that individual instances fit into and thus take on that value range. This results in a loss of information and can be a reduction in the empirical level of a measurement (interval/ratio data is reduced to ordinal, for example). This loss of precision is offset by the simplification of the presentation of information. These classifications are accomplished by representing ranges of data as categories on the display medium, through the use of symbolization that is appropriate to the level of measurement (see Figure 1.1 on page 2).

Spatial data can be grouped into four categories. The first type of data represents point specific information. The second type of data represents linear information. The third type of data represents area information. The fourth type of data represents volume information. Each of these categories is scale specific; an area feature such as stream may require mapping as a linear feature if the total area under consideration is small enough that displaying the stream as having both length and breadth becomes too tedious or difficult to represent in the available media, for the additional information retained. Volumetric information is also dependant on the means of display--the appearance of three dimensions can only be approximated on a two dimensional surface. Although the visual variables that best represent different levels of information do not change for each of the types of spatial data, the ARC/INFO methods for accomplishing those representations change.

For environmental data, Mark Monmonier and Branden Johnson (1990, 5-7) have characterized spatial data into: single location; single location and affected area; and multiple locations and the pattern of distribution. Single location answers not only 'what' but 'where'; this can be applied to all of the basic types of spatial data and allows the map user to relate environmental information to his or her own experience (this is what Edward Tufte (1990) calls micro/macro readings). Building on single location, single location and affected area adds information concerning how a 'what' influences its surroundings (because 'influence' may be more subject to interpretation than a measurement, presentation of meta-data can become very important in the presentation of data). Finally, multiple locations and the pattern of distribution integrates more than one instance of single location and affected area into one map.

Area phenomena can also be categorized on the basis of its spatial grouping and how it changes over space (MacEachren and DiBiase 1991) (see Figure 1.1). Grouping ranges from continuous (no grouping) to discrete (complete grouping). This reflects the degree of spatial autocorrelation within areas. Changes over space can be smooth to abrupt. This reflects the degree of spatial autocorrelation between areas. These changes suggest that appropriate symbolization choices be made that accurately reflect the nature of the data. The possible symbolization choices include, but are not limited to: graduated symbols for abrupt, discrete data; dot density for smooth, discrete data; isopleth (or the '3-D' equivalent fishnet) for smooth, continuous data; and choropleth for abrupt, continuous data; this is in contradiction with the all-to-common practice of making choropleth maps for all types of area data.

## Meta-Data and Uncertainty

The uncertainty of information is becoming an important topic with the increased use of computers for data processing, presentation and analysis. This is being addressed with position statements such as the Environmental Protection

Agency's *Locational Data Policy* (1991) and the National Center for Geographic Information and Analysis' Visualization of Data Quality initiative (MacEachren 1992, 47). Yet, these cannot eliminate uncertainty, which exists at the most basic levels of measurement according to Heisenberg's Uncertainty Principle (Capra 1983). Because of this, David Rejeski (1991) suggests that uncertainties should be addressed openly in order to ensure that decisions have both utility and believability. He, as well as Granger Morgan and Max Henrion, recognize that uncertainty can be valuable information. Morgan and Henrion (1990, 3) give three specific reasons for the inclusion of uncertainty in policy oriented research:

1. A central purpose of policy research and policy analysis is to help identify important factors and the sources of disagreement in a problem, and to help anticipate the unexpected. An explicit treatment of uncertainty forces us to think more carefully about such matters, helps us identify which factors are most and least important, and helps us plan for contingencies or hedge our bets.
2. Increasingly we must rely on experts when we make decisions. It is often hard to be sure we understand exactly what they are telling us. It is harder still to know what to do when different experts appear to be telling us different things. If we insist they tell us about the uncertainty of their judgments, we will be clearer about how much they think they know and whether they really disagree.
3. Rarely is any problem solved once and for all. Problems have a way of resurfacing. The details may change but the basic problems keep coming back again and again. Sometimes we would like to be able to use, or adapt, policy analyses that have been done in the past to help with the problems of the moment. This is much easier to do when the uncertainties of the past work have been carefully described, because then we can have greater confidence that we are using the earlier work in an appropriate way.

Uncertainty has dictionary definitions such as, "uncertain in respect of duration, continuance, occurrence, etc.," "liability to chance," "indeterminate as to magnitude or value" (Simpson and Weiner, *Oxford English Dictionary*, 1989, 899).[1] A more useful interpretation for use in environmental risk analysis would be that uncertainty is the information contained in the data about data (that is, meta-data). By defining uncertainty this way, meta-data can be used as another piece of information in the analysis and presentation of data, including risk-based policy making.

Uncertainty has a taxonomy that should be useful in delimiting the origins of uncertainty and the reliability of data at any given point in an analysis (see Figure 1.2 on page 6). Although Morgan and Henrion (1990) discuss "The Nature and Sources of Uncertainty" as chapter 4 of *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* in detail, this presentation is a distillation based on several sources, which include Morgan and Henrion and others sources, as noted. The taxonomy can be broken down into three groups: uncertainties of the physical world; uncertainties of the computer world; and

---

[1]. Uncertainty begins as the vagueness of duration, etc with Wyclif in 1382. By 1982, Oxford (1989, 900) reports: "What the uncertainty principle asserts is that for *no* state of *any* system can *all* dynamical variables be arbitrarily well-determined."

uncertainties of the human world. Problems of the measurement of natural phenomena constitute the first category; Alan MacEachren (1992, 48) reports that the National Center for Geographic Information and Analysis calls this "data quality". The uncertainty of the physical world can be further split into measurement uncertainty and parameter uncertainty.

**Uncertainty**

- **Physical World**
  - **Measurement**
    - Location
    - Attribute
  - **Parameter**
    - Aggregation
    - Generalization
    - Time
    - Consistency
- **Computer World**
  - **Descriptive**
    - Numeric
    - Spatial Delineation
  - **Computational**
    - Rounding
    - Significant Digit Shift
    - Over/Underflow
  - **Propagational**
    - Compounded Computational
    - Polygon Overlay
  - **Modeling**
    - Robustness
    - Validity
- **Human World**
  - **Sending Meanings**
    - Open Presentation of Data and Meta-Data
  - **Receiving Meaning**
    - Ignoring Meta-Data
    - Misunderstanding Meta-Data

**Figure 1.2** A taxonomy of uncertainty.

### Uncertainties of the Physical World

Measurement uncertainty for geographic data includes both location and attribute uncertainties. Location uncertainties (Rejeski and Kapuscinski 1990, 10) can be considered the accuracy (closeness to a 'true' value) of the instruments and the reliability, or precision, (repeatability of a measurement) of the methods used to calculate a site's position (see Figure 1.3 on page 7). For example a location of a phenomenon can be determined by use of many methods, such as a professional surveyor's analysis, use of a global positioning system, or terrain analysis estimation. Each of these methods is of varying accuracy and precision, and the meta-data that should be recorded includes the manner in which a site location was first defined, the method use to derive its location, the time it was derived, and an estimate of its accuracy.

Attribute uncertainty can be considered the accuracy and reliability of the instruments used to take a measurement of the environment. For example, an instrument may be set up to measure atmospheric concentrations of carbon monoxide, measuring levels in parts per million and with an accuracy of plus or minus one part per million. The data in this situation is the part per million measurement and the meta-data includes the plus or minus one part per million accuracy of the instrument.

Parameter uncertainty (Rejeski and Kapuscinski 1990, 11; MacEachren 1992, 47-8) involves the problem of the aggregation and generalization of point samples to areas for spatial data, or trend lines for linear data; the temporal variability between data items and between measurement times and data usage; and the logical consistency and completeness of data. This is the question of whether the measurements that are recorded and then used for analysis are adequate measurements of what was intended to be measured; it entails the consequences of the assumption of autocorrelation. Before a model is constructed to provide an explanation or a projection of environmental phenomena, it must be recognized that there are few (if any) phenomena that can be precisely defined and measured for all possible occurrences. Because of this, interpolation and extrapolation--whether linear (as in a measurement), spatial (as in an area generalization), or temporal (as in data from one time as estimates for some other time)--must be done even though the process introduces uncertainty into the analysis.

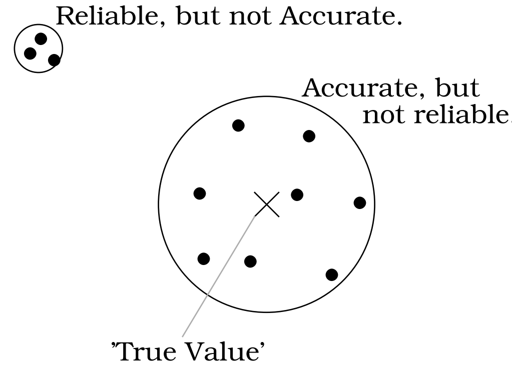



**Figure 1.3** An example of the difference between accuracy and reliability in spatial location measurements.

There have been suggestions made for reducing and representing parameter uncertainty. For generalizing spatial data, Cort, Rowe and Philpot (1985; in MacEachren 1992, 44) suggest that interpolation in spherical, rather than planar, coordinates will introduce less uncertainty in the creation of area information from point samples (particularly for large areas). MacEachren and Davidson (1987; in MacEachren 1992, 46) demonstrate that increasing sampling frequency will also reduce (but not eliminate) the uncertainty of interpolation; this should also be true for linear and temporal data as well. For, representing parameter uncertainty, Rejeski and Kapuscinski (1990, 10) suggest the use of transitional buffer zones to represent "fuzzy" boundaries, rather than one line demarcating "hard edges" (see Figure 1.4 on page 8).

### Uncertainties of the Computer World

Problems of the use of computers for storing and manipulating data constitute the second category of uncertainty. This can be subdivided into four groups: descriptive uncertainty, computational uncertainty, propagational uncertainty and modeling uncertainty. Thoughtful use of programs and programming techniques can reduce the amount of uncertainty introduced to data by computer manipulation. Herman Knoble (1990, 2) states that correct and accurate computer programs must be an ethical responsibility when dealing with numerical algorithms, all of which, "at the bottom line...affect people."

Descriptive uncertainty deals with the representation of data in computers, including both numeric and spatial problems. Numeric uncertainty arises from the method computers use to store data. For example, the binary number system does not have an exact representation of some common decimal numbers, such as 1/100. Another type of numeric uncertainty arising from number storage is the shifting of significant digits. If a number is input into a computer that has the ability to store a longer number than is input, the computer will 'pad' the extra

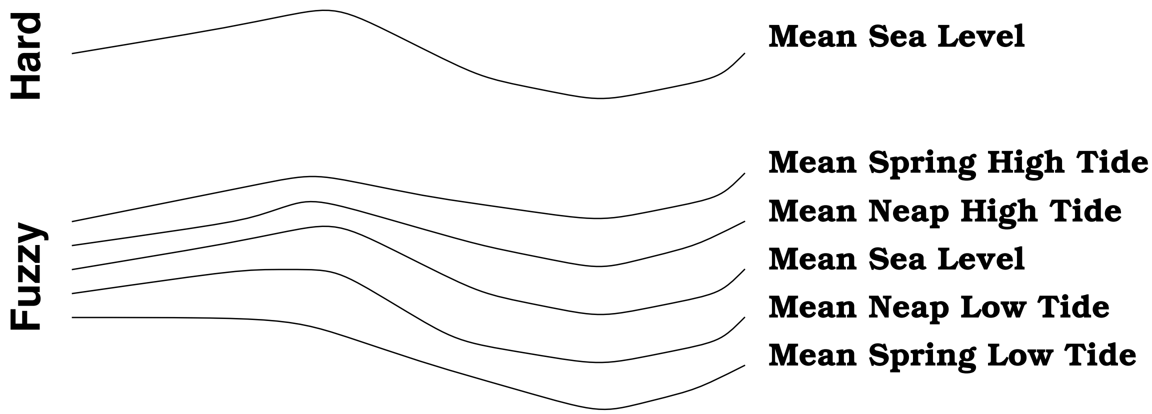



**Figure 1.4** An example of a change from a hard to a 'fuzzy' boundary.

space with zeros. These digits will be available for computation in numeric modeling even though they add no information and are meaningless. Significant digit shift can occur in the other direction as well. If a number is input into a computer that stores fewer digits than is input, the computer will round or truncate the number to fit its numeric scheme. These problems can be controlled, but not eliminated, by specifically programming the computer for the required operations (Knoble 1990), but for generally available software this is not possible.

Spatially, data is generally stored as either a table of vectors that define sharp boundaries between regions, or in regular tessellations (square, 'raster' grids) that force a predefined spatial pattern on an area (Rejeski and Kapuscinski 1990, 10). Both of these methods introduce uncertainty: the vector representation forces boundary lines where transition zones may be; and tessellations assume the entire grid cell is homogeneous. These problems can be reduced, but again not eliminated, by using smaller polygons or grid cells, but this forces a trade off between data file size and computational time, and decreased uncertainty, which may not be pragmatically feasible.

Computational uncertainty deals with the problems of numeric modelling by using computers (Knoble 1990; Rejeski and Kapuscinski 1990, 13). Once data are stored in a digital format, any further processing can introduce uncertainty. Type shifting (such as from an integer format to a real number format, or from reals to integers) can introduce uncertainty by forcing rounding to occur. More commonly, rounding occurs within real numbers when numbers that are not similar in value are arithmetically joined. Knoble (1990, 4-5) gives an example of the results of this type of rounding: an IBM 3090 Model 600, VS FORTRAN 2.4 program that for the formula:

```
P=((A+X)**2-A**2-2.*A*X)/X**2
```

generates an answer of -4999.99609 when A is equal to 1000.0 and X is equal to 0.01, despite the fact that the formula simplifies to P equalling one for all X's not equal to zero.

Significant digit shift can also occur in arithmetic operations, particularly in operations involving subtraction of similar values. Subtracting $1.23456787 * 10^7$ from $1.23456789 * 10^7$ yields 0.2, but a computer can store this value as 2.00000000

* $10^{-1}$, and as with data input, will allow computation on all of the digits to the right of the 2, which is the only meaningful digit in the number.

Two additional, similar problems that may occur are overflow and underflow. These result when a number is incremented, or decremented beyond the storage type's ability to represent numbers. Depending on the operating system this may cause an error or may be ignored with the value remaining the same or drastically changing. Borland's Turbo C++ 2.0, running under MS-DOS 5.0 on an IBM AT, compiles programs that allow adding one to the thirty-two bit, 'unsigned long' integer 4294967295 (which is equal to $2^{32}$ - 1, and is the largest integer that can be represented in thirty-two bits), changing the value of the variable to 0. Turbo C++ compiled programs will give an overflow error when thirty-two bit real numbers (type 'float') are incremented outside of type float's range. When these computational uncertainty errors occur in a program, and are not handled well, the program could continue and generate an apparently correct answer purely by chance (Knoble 1990, 2).

Propagational uncertainty deals with the problem of how uncertainty from a physical world measurement or another computer-related uncertainty moves through successive iterations of a model. It can be tested by varying the input to a numeric model in small steps to see if small changes can make large differences in the output of a computation, which in principle should not occur. For example, by making the value of A equal to 100.0 and X equal to 0.01, Knoble's (1990, 5) program generates a value for P equal to -39.0624847; by changing X to 0.0078125, the program generates a value of P equal to 0.0000000. Propagational uncertainty can also be tested if the computer program can be rewritten to an algebraically equivalent, but computationally different manner, which would allow comparison between programs that should generate the same output. By simplifying the formula in Knoble's example to a command line such as:

```
if X not equal to 0 then P = 1 else print "DIVISION BY ZERO"
```

the problems of floating point arithmetic can be avoided.

Propagational uncertainty not only involves the accuracy of real number computations for numeric data, but also the generation of polygon overlay slivers within vector spatial data, such as that used by ARC/INFO (see Figure 1.5 on page 10). The slivers may be handled by "fuzzy tolerances" (ARC Command References, Commands J-Z, 1991, UNION 1), which would allow the shifting of close lines so that they merged in the output. This could cause the shifting of data from one layer of a known high accuracy (such as a surveyor's cadastral data) to correspond with a layer of lower or unknown accuracy (such as information digitized from a medium or small scale paper map). All future use of the merged data layer would include the uncertainty created by the overlaying of two data layers of varying accuracy, and the meta-data that should be attached to the new data layer would have to reflect an estimate of how much shifting occurred. This type of shifting can be eliminated by avoiding fuzzy overlays, although this can cause the generation of sliver regions along boundaries that will cause increased storage and processing time for use of the new layer, and may not hold any useful information other than the indication of the difference between the two layers used to create the new layer.

Modeling uncertainty culminates the types of uncertainty associated with computers. Although there are several types of modelling (verbal, graphic, physical, and mathematical), each of which is subject to questions of robustness and validity versus the goal of the model (such as description or prediction), it is only mathematical models that tend to be dependent on computers for execution and are thus subject to the other computer uncertainties. The robustness of a model is reflected in a model's ability to handle all appropriate input and produce a reasonable output. It is thus tied to propagational uncertainty, but robustness also entails that not only do small changes in input not cause inappropriate changes in output, but that the model's output, in practice, is also within the expected range of the mathematical model, in principle. The validity of a model is the question of whether a model, in practice, actually represents the phenomena being modeled, in principle. This questions the methods used to operationalize a numerical model, such as the validity of using 'if…then' statements to ensure the apparent robustness of a procedure that would otherwise produce non-robust output, when no such statements are apparent in the 'real world' phenomena.

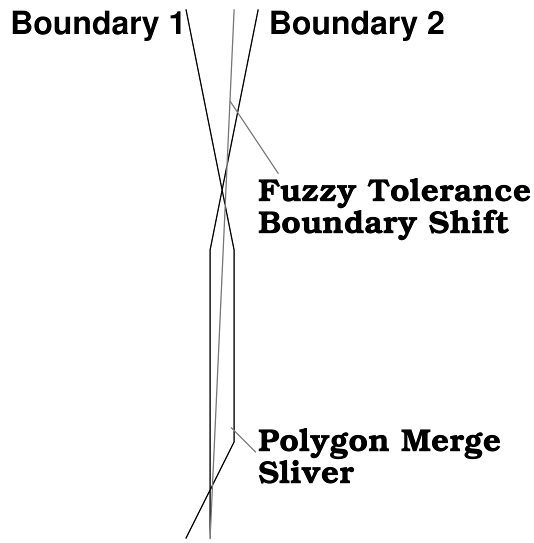



**Figure 1.5** Polygon overlay can cause shifting in boundary lines, or can create boundary slivers.

### Uncertainties of the Human World

Problems of the human communication of information and meanings constitutes the third area of uncertainty (Rejeski and Kapuscinski 1990, 12). In attempting to communicate information meta-data can be lost in two ways: the sender of information can give data without the meta-data information, or the receiver of information does not understand that part of the message. The first way, not giving meta-data with data, can be the result of restrictions of space within publication materials, the desire for the appearance of greater accuracy (Star 1985), and until recently a lack of awareness of the potential importance of uncertainty information, particularly in the areas of human and environmental risk analysis (Rejeski and Kapuscinski 1990). The second way that meta-data can be lost is when the receiver of information does not receive that part of a message. This can result from the ignoring of meta-data or the inability to interpret the meta-data through lack of experience in dealing with the way it is conveyed. This type of failed effective communication can be the result of the variable interpretations of both words and images. The meaning of words constitutes one of the major problems the U. S. Environmental Protection Agency has had to deal with in risk analysis (Rejeski and Kapuscinski 1990, 12). Effective communication of meta-data has been studied in a strictly human realm (body language, etc.), but little has been done in the area of communication of uncertainty within the realm of scientific communication, particularly with spatial data (Rejeski and Kapuscinski 1990, MacEachren 1992).

This taxonomy should prove useful, particularly for reducing human communication uncertainties. By recognizing that uncertainty exists in measurement and is propagated in computer manipulation of data, these uncertainties can be dealt with openly and honestly, as Rejeski (1991) suggests. This, aided by visual communication techniques, should reduce human uncertainty by increasing the amount of information (by communicating meta-data) given in the presentation of data.

## The Idea of Visual Communication

The effective communication of environmental data and uncertainty as meta-data requires an understanding of the principles of visual communication and map design. David DiBiase (1990) has developed a model of information display, in a research setting, as a means of communication in a continuum from communication to self through communication to others (see Figure 1.6). Communication to self can be thought of as "visual thinking", and includes data exploration, hypothesis generation and confirmation. At this level of data visualization, maps and other graphics are used to "prompt insight, reveal patterns in data, and highlight anomalies" (MacEachren 1992, 1). The goal in the creation of these images should be to assist these goals. Because of these goals, it is with this type of visual communication that the possibility of visualization error is greatest. MacEachren and Ganter (1990) describe these errors as seeing wrong (similar to the type I error in hypothesis testing--identifying a pattern that is not there) and not seeing (similar to the type II error--not identifying a pattern that is there). But, since these graphics are generally rough and intended only for viewing by the researcher(s), searching for one 'optimal' display is less important; more views on the same data may be more helpful and reduce the chances of not seeing or seeing wrong.

Communication to others can be thought of as "visual communication" and is the realm of presentation graphics. At this level of data visualization, maps and other graphics are used to synthesize data into an "abstract statement concerning patterns and relationships" (MacEachren 1992, 6) and finally to present the information to others to persuade them of the accuracy of the data assessment (MacEachren 1992, 7). Edward Tufte's (1983, 77) principles of graphic excellence can help

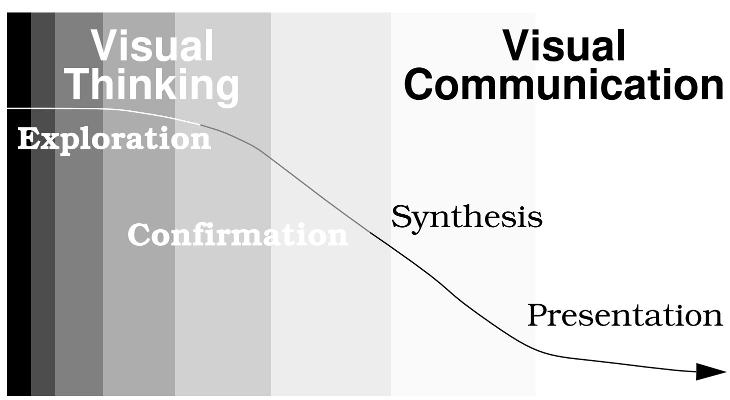



**Figure 1.6** DiBiase's (1990) model of the function of graphics in research. Visualization begins with exploring an idea. The idea then moves through confirmation and synthesis in a larger group of both ideas and people, and ends with presentation (although each stage can spawn new ideas).

ensure the most information is presented in a minimum of space, and that this information is conveyed ethically:

The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
Show data variation, not design variation.
In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
Graphics must not quote data out of context.

Tufte also discusses the concept of data-carrying ink--that is, don't put more ink on the page than is necessary to convey the information. These principles are similar to those that Morgan and Henrion (1990) propose for designing graphics for the presentation of uncertainty information. Judy Olson (1981) also suggests the need for clear and accurate legends in her guidelines for the production of bivariate maps.

Monmonier and Johnson (199077) have proposed a guideline for the communication of environmental risk, which can aid in the making of maps for visual communication. Their multistep process is not a waterfall type model (when one level is completed it cannot be returned to), but rather a guideline for iterative refinement for the presentation of data. The steps they include are: setting up the design team; identifying the communication goal; the issue profile; the audience; the messages of environmental maps; methods; and evaluation. Setting up the design team acknowledges that one person may not have all of the knowledge necessary to adequately design a graphic, and that, as necessary, each of the following steps should involve each person who has or can contribute to the map. Identifying the communication goal is simply that a focus should be selected for the map; this will facilitate inclusion of important information and removal of extraneous data. The issue profile deals with the history of the problem to be mapped and the constraints on producing the map; that is, the 'environment' of the map and map design process.

The audience must also be considered when designing a graphic. This includes consideration of who will be viewing the map (politicians, scientists, the public, etc.). These different audiences will generally have varying degrees of map-interpretation skills; this influences the amount of explanatory information that should be included, the appropriateness of bi- or multivariate maps, and the appropriateness of 'eye catchers' such as bright colors. Consideration of the audience should also include an acknowledgment of how the map will be presented; this is part of the issue profile and influences the choices of methods.

Monmonier and Johnson (1990) present a list of several categories of messages that environmental maps generally present. The first of these is "What we found/What we know/What we think we know" (p12). This is the presentation of information at its most basic level, but even at this level meta-data can be presented--'what we think we know.' The second of these messages is "What you can do/What you should do" (p13). This is very dependent on the choice of audience--a lawmaker's set of choices of what to do can be quite distant from a concerned citizen's, for example. The third message is "What we're doing/ What we want to do" (p13). For this type of message Monmonier and Johnson suggest that an overview map with several smaller maps of detail maps may aid the presentation. The last category is "Why we're doing what we're doing/Why you should do what we're asking you to do" (p13). This type of map should present the reasoning behind a choice or plan of action; this can include the presentation of the history side of the issue profile.

The final two steps of Monmonier and Johnson's strategy are methods and evaluation. Methods need to address questions such as the need for one or several maps, how these maps will be presented (large size color maps, 8.5x11 black and white maps, slides, video, etc.) and whether or not additional information such as non-map graphics should be included. Although evaluation is presented as the last step in map design, it is a part of each of the earlier steps. It is the last step as a review of the, possible, final design. Evaluation can be formal (a survey) or informal (a telephone call to someone who has seen the map and can suggest any possible improvements).

## The Means of Visual Communication

Jacques Bertin proposed a group of 'visual variables' that has since been added on to by other cartographers (such as: Morrison, 1974; McCleary, 1983; and Woodward, 1991). This group of variables constitutes those representational techniques that a cartographer or illustration designer has in the creation of an image. The list includes: location, size, value, hue, intensity, orientation, shape, texture, arrangement, and focus (see Figure 1.7 on page 14). This list is not necessarily all inclusive, but the list is useful in the design of maps and graphics. Each of the variables can be used to establish visual isolation (difference from surroundings) and visual levels (greater noticeability) (see Figure 1.1 on page 2), and according to Bertin, these visual variables have certain levels of measurement that are commonly associated with them, and thus allow representations that convey the character of the data.

In a static map, the use of location is limited to the signification of the spatial place of an item, although in orthogonal displays, location is a flexible tool because of the possibility of specifying the viewpoint on a simulated three-dimensional surface. In multiple maps, or dynamic mapping, change in position can be used to show movement of a feature. It is inherently an interval/ratio variable, but can be used to depict all levels of measurement.

Size is most often used to depict an ordinal variable, although interval/ratio variables can also be depicted using size. Because a larger symbol is almost always associated with 'more,' this is the context that it should generally be used in.

Of the 'color' variables, Bertin (1983, 42) only identified value and hue. Value, like size, has a distinct range from more to less and is therefore good for mapping ordinal data and can also be used for interval/ratio data. ARC/INFO refers to value as lightness. Hue can be used to depict ordinal or interval/ratio data because the frequencies that constitute hue are ordered, but these orders are not always readily remembered and used; hue (at a constant value and intensity) is therefore best used for nominal data. Intensity, which is also called saturation, has a distinct range from more to less and is therefore good for ordinal data.

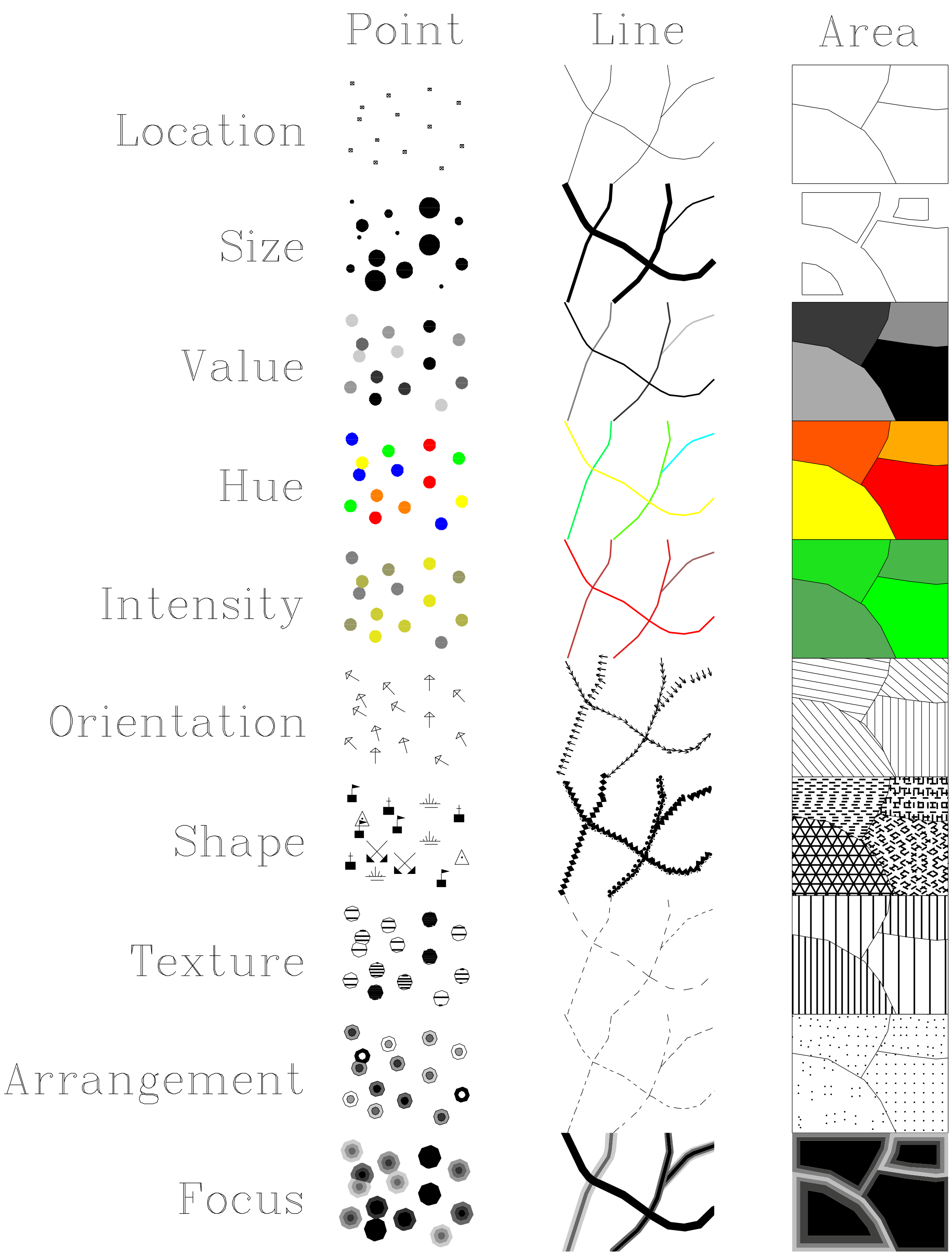



**Figure 1.7** The visual variables as presented by DiBiase, Krygier, Reeves, MacEachren, and Brenner (1991). All have been done in ARC/INFO 6.0, but some (such as texture for point symbols, orientation for line symbols, and arrangement for area symbols) are more difficult and, thus, less useful than others.

As availability increases for the means of creation of high resolution color displays, use of the color visual variables will become an even greater part of cartography. Color can be specified in several ways (Dent 1985). One common and useful way is the Munsell color system (see Figure 1.8); it is based on the human perception of color (and is similar to the Tektronix color cone, and a color specification system in ARC/INFO, Hue-Lightness-Saturation). This system divides color into the three visual variable categories value, hue and intensity.

Value is the measure of the lightness or darkness of a surface; it is the total amount of light that is reflecting or emitting from a surface measured relative to the human ability to discriminate black from gray up to white. It is the only measure of light along the white-gray-black continuum, because all frequencies of light should be equally present.

Hue is the term most often meant when the word 'color' is used. The Munsell color system records hue as an angular measure around a color space. It represents the modal value of the frequency of light that is reflecting or being emitted from a surface. When a light has no modal frequency, the perceived light is along the white-gray-black continuum, which constitutes the central axis of the Munsell color space.

Intensity is a measure of the purity of the spectral frequency of light. It represents the variance of the frequency of light; greater variance means that a larger range of frequencies of light are being emitted or reflected from a surface--that the surface has more gray in it. A feature of intense colors is that they tend to be more noticeable (stand out more) than less intense colors, even when the total amount of light reflected or emitted may be the same, which makes intensity good for establishing visual hierarchies. Because of the human visual system, the value of the most intense color of different hues varies, with yellow having the highest value for its maximum intensity. This change is accounted for in the Munsell color system; the Tektronix color cone assumes that maximum intensity for any color occurs at the midpoint of the value scale.

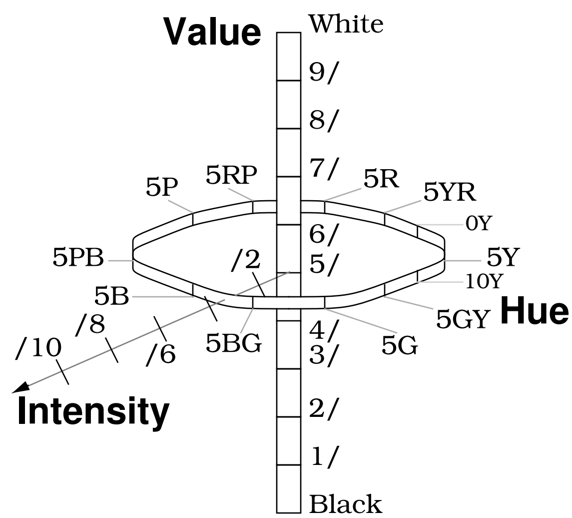



**Figure 1.8** The Munsell color system; for example, an intense yellow could be specified as 5Y7/14.

For cartographic use in displaying ordered data, color hues are generally arranged to allow ease of interpretation. There are many possible color schemes, many having been suggested for terrain shading. For other data, the spectral ordering of hues may be the most obvious for use in mapping ordered data, but because colors can shift in intensity and value, this may lead to misleading maps--yellow will stand out in the spectral pattern. Two suggestions are of note for remedying this problem. The first is the use of a part-spectral scheme: yellow through orange to red; or, yellow through green to blue. This can allow redundancy in the visual variables of intensity and value, which reinforces the hue progression. The second is the use of hues ordered on the basis of value. This may be a viable alternative, but could prove confusing to those who know the spectral sequence of hues.

The ability to discriminate colors has received some study, particularly in the discrimination of values. The Munsell color system divides the range from black to white into eleven steps, as does the system for black and white photography developed by Ansel Adams (Upton and Upton 1989, 313). For most cartographic use, eleven steps would be both difficult to create and difficult to interpret; most cartographic research indicates that the use of half that range (five or six) is better for map design. If value gives an indication for the amount of discriminability that can be expected for intensities, four intensity values is probably the most that should be used. Discrimination of hues for cartographic use is affected by the apparent changes in intensity and value that occur as hue is changed, but for univariate symbolization, five steps should be easily discriminable in one of the part-spectral sequences.

Colors are produced by two methods: color addition and color subtraction. The color addition process is most commonly seen in color monitors for computer displays and television. It involves the mixture for red, green and blue to create colors on a black surface with all three being used for white. The color subtractive process is used for the production of printed material, such as paper maps. It involves the use of cyan, yellow and magenta (and black) to subtract colors that would reflect from the surface of white paper. Because the overprinting of cyan, yellow, and magenta generally leaves muddy brown, black is often used as a fourth color in the printing process. Colors are obtained by overprinting the four separates, with each offset slightly in order to allow the apparent mixing of colors in a dither pattern.

The production processes must be kept in mind when choosing color schemes on a color monitor that will be printed on a paper output device. Colors on a computer display may not appear the same on a printed sheet, because of the change in creation method. In addition, on a computer monitor, a point can take on any of the possible colors generatable by the graphics system, and resolution is independent of color. On paper output, the dither patterns that are used to create the appearance of many colors may cause a change in apparent color and, more significantly, output resolution. ARC/INFO allows the use of color lookup tables in commands such as HPGL2 (ESRI *ARC Command References*, 1991) to enable redefining screen colors to pretested printer colors to help compensate for changes in color; changes in resolution must be accounted for by the cartographer.

A final consideration for color (particularly hue) is its social interpretation. Colors can have meanings that must be taken in to consideration when designing maps and graphics, for example: red as stop; yellow as caution; green as go. For environmental maps, use of red (particularly, intense red) can indicate imminent danger, and with dark shades of other colors (blues, grays and browns, for example) can evoke a sense of foreboding or futility (this is demonstrated in *The Nuclear War Atlas* and movies such as *Blade Runner*). Other color schemes can invoke other reaction: pastels (colors of low intensity and high value) can indicate serenity (unclassified maps by the Central Intelligence Agency often make use of pastels); intense colors such as yellow and cyan grab attention, and can indicate happiness--they are often used in maps for children.

The final three variables that Bertin identified are orientation, shape and texture. Orientation is readily discriminated by the human visual system and is therefore good for indication of nominal data, although by using common symbols such as clock faces, ordinal and interval/ratio data can be displayed with orientation. Shape should be used for nominal data categories, because shapes in general do not have an apparent order and are not as readily distinguished as other visual variables. Texture is the relative coarseness of an area fill, and because rough textures appear closer than fine textures, texture is good for establishing visual hierarchies. Texture should generally be used for nominal data, but it can also be used for ordinal and interval/ratio data.

Unlike orientation and texture, shape has a continuum of possible representations: from mimetic to abstract (see Classifying Space on page 24). Mimetic symbols convey the appearance of what they represent (an animal's outline represents sightings of that animal). Abstract symbols must be defined (a legend indicates that triangles represent sightings). ARC/INFO provides many abstract symbols, a few that are less abstract (for example, an anchor that might be used to represent a marina), and the ability to define new symbols (see *Map Display and Query*). Selection of appropriate levels of abstractness must be considered in designing a display--use of mimetic symbols may allow less dependence on the legend, but too many mimetic symbols may give the appearance of a map oriented toward children.

Two additional visual variables that have been identified since Bertin proposed his list are arrangement and focus. Arrangement is the order of symbols in an area fill (from regular to random to clustered), and nominal and ordinal data can be represented by changes in arrangement. Focus is the last visual variable; it is the crispness of the edges of symbols. Because this has an apparent order, focus can be used to display ordinal data and could be used for interval/ratio data, although too much variability of focus may make a map hard to read.

## Methods for Visually Communicating Data and Meta-Data

The communication of uncertainty has been studied by Morgan and Henrion (1990) with Harold Ibrekk, although their concern was the communication of the uncertainty of linear data. Because of this, the methods they present cannot be readily used in the communication of mapped geographic data, except for point symbolization (even this is difficult in ARC/INFO), but their study highlights some of the difficulties that can be encountered in the visual communication of uncertainty and their conclusion are useful.

Their study analyzed nine methods for displaying uncertainty in linear data: a point estimate with an error bar; a discrete density function); a pie chart; a probability density function; a half-height probability density function mirrored on the x-axis; a dot density horizontal bar a vertical line density horizontal bar; a modified Tukey box plot (minimum and maximum points are not included, and a mean point is included); and a cumulative density function (see Figure 1.9 on page 18). Their recommendations include using displays that specifically show the information that is to be extracted (such as a point for a mean value, if mean values are important), and using multiple displays (particularly, display the cumulative density function and the probability density function one over the other, with a

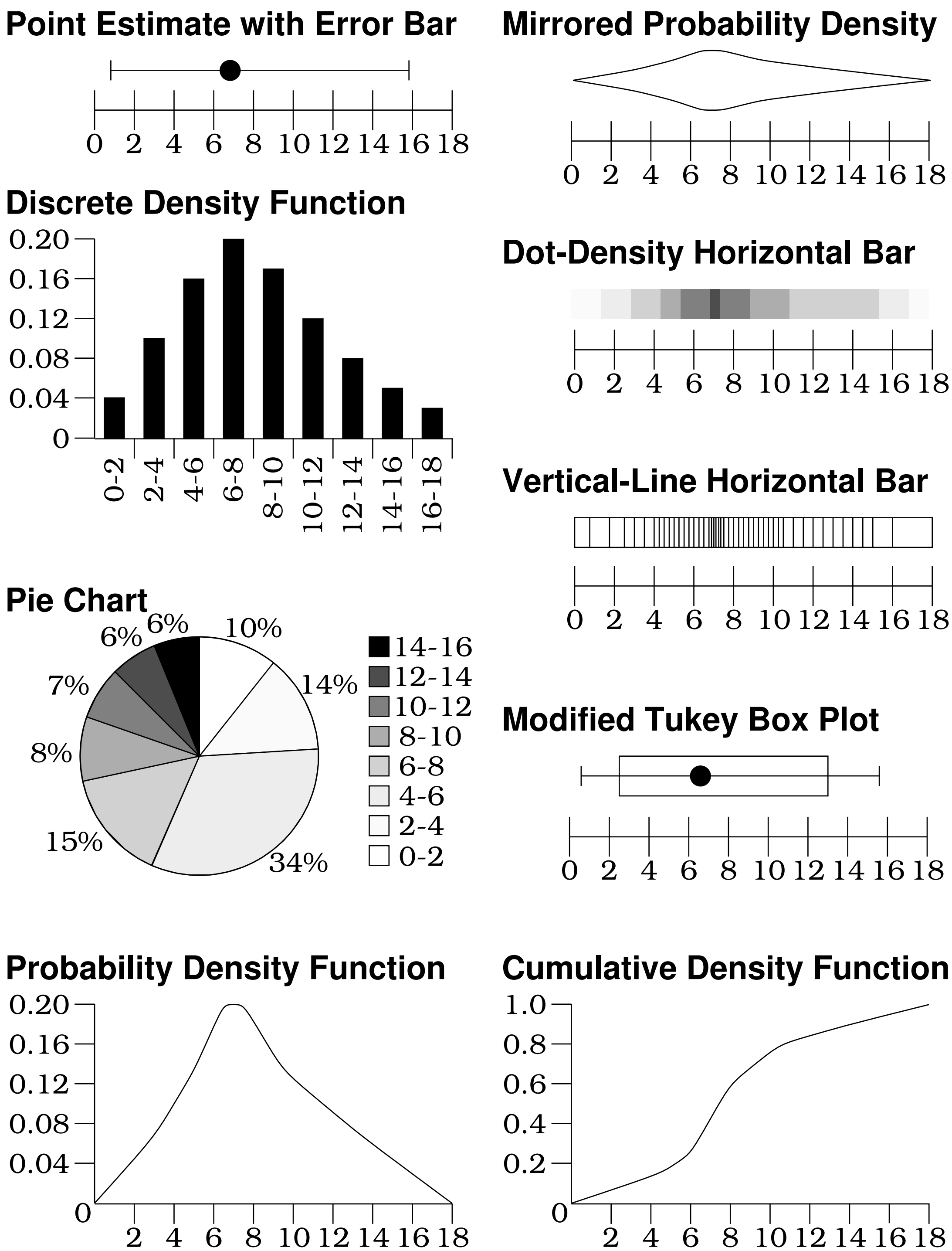



**Figure 1.9** The methods of graphically communicating uncertainty tested by Morgan, Henrion and Ibrekk (redrawn from Morgan and Henrion 1990, 228-9; not to scale). For single variable graphs Morgan and Henrion suggest using a combined probability density/cumulative density display, and labeling important points in the data (for example, the mean value).

common horizontal axis).

For two- and multidimensional uncertainty in non-spatial displays, Morgan and Henrion give several examples of methods that can be used for graphing (see Figure 1.10 on page 20). These include: multiple lines in a probability density/ cumulative density pair; multiple Tukey box plots; linear graphs with error bars; orthogonal displays; and triangle plots alone and in multiples.

They conclude with a list of factors that should go in to design of uncertainty displays (Morgan and Henrion 1990, 252):

finding a clear, uncluttered graphic style and an easily understood format,
making decisions about what information to display,
making decisions about what information to treat in a deterministic form and what to treat in a probabilistic form,
making decisions about what kind of parametric sensitivities will provide key insights.

They also suggest that display design often involves the reduction of a multi-dimensional model into the two dimensions of a paper or monitor display (as does Tufte 1991), and that the intended audience's experience in interpreting graphs must be considered when creating a display.

For the representation of uncertainty for spatial variables, there are two possible cartographic routes. The first of these is the creation of two maps, one for displaying the data, and one for displaying the meta-data. Olson (1981) and Laurence Carstensen (1986) have tested this choropleth map arrangement against two bivariate mapping techniques (Olson: spectral encoding; Carstensen: intersecting lines) in the representation of statistical correlation.

Olson finds that map readers can initially interpret value shaded, monovariate map pairs more readily than bivariate maps. On the other hand, she reports that over half of those who could interpret bivariate maps at a significantly better than guessing level, did better with a bivariate map than two separate maps. She then suggests that the bivariate maps may be more readily interpreted once the "cognitive hurdle" (Olson 1981, 269) of the bivariate mapping technique is overcome.

Carstensen also finds that map readers find interpreting value shaded, monovariate map pairs easier than bivariate maps. Because his test compared a classed map pair and unclassed bivariate maps, the problem he notes with the use of map pairs (poorer statistical residual scores) would not be as likely to occur if map pairs are compared with classed bivariate maps.

The map pair technique should be a useful tool for communication of two variables, when the data must be classed (as is done for pragmatic reasons in ARC/INFO). Since meta-data values may not be correlated with data values, the use of two monovariate maps should be an effective tool for communicating uncertainty versus techniques that were designed to highlight spatial correlation.

The second cartographic route for the visual representation of uncertainty is bivariate mapping (the mapping of two variables onto the same map). This would ensure that meta-data is presented to the map reader with the data, but this mapping method generally increases the difficulty with which data can be extracted from a map. Because of the variety of visual variables, there are several possible methods of bivariate mapping, some of which have been tested for communication effectiveness; most of these have been oriented toward the

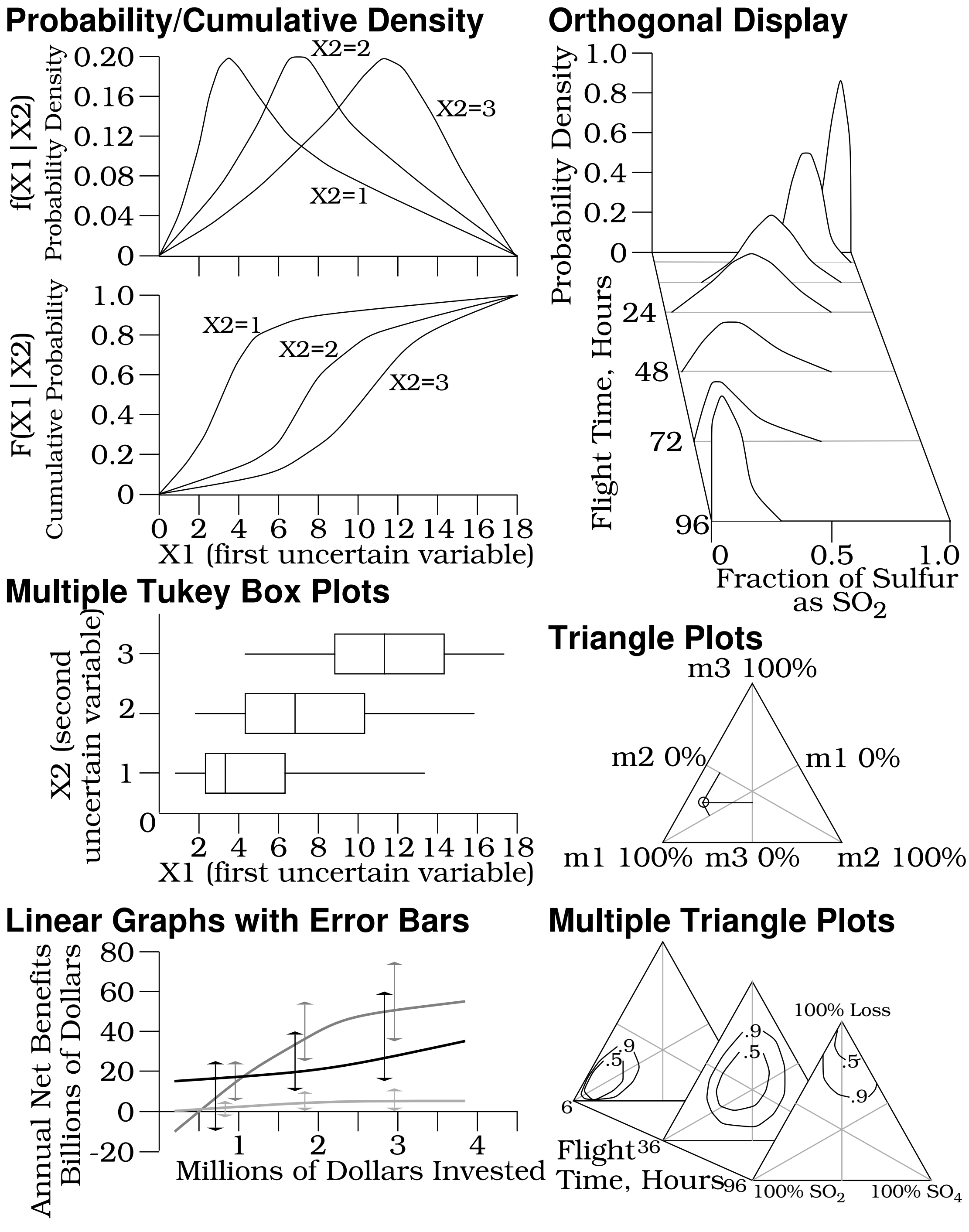



**Figure 1.10** Two- and multidimensional representations of uncertainty (redrawn from Morgan and Henrion 1990, 240-51; not to scale). Values in triangle plots are read by going from each of the bisecting lines out perpendicularly to the point of intersection; the position on the bisecting line is the value of the point for that measure, and the sum of the measures from the three bisecting lines will equal 100%. The 50% and 90% probability areas for the most likely location for a measured point indicate peaks in the trivariate probability density function.

representation of spatial correlation. This includes the testing of the intersecting lines method (which uses texture), color maps made by the U. S. Census Bureau (which uses a full hue range to achieve bivariate representations--spectral encoding), color maps that rely on two complementary hues, and the equiprobability ellipse (which also uses the complementary hues).

In a technique that allows for bivariate mapping in monochrome displays, Carstensen (1982) has tested the communication effectiveness of the intersecting lines method of bivariate mapping (see Figure 5.4b on page 73). These maps use horizontal and vertical lines for representing two variables. Carstensen suggests that this scheme be used in an unclassed map, but this scheme can be used for classed representations of data and meta-data. Although this technique can be used in monochrome displays, colored lines can be used to distinguish the two variables (this has not been tested for communication effectiveness, though).

This technique has one disadvantage: the method of producing the area symbolization may lead to a conflict between two visual variables. Both texture and value change with the mixing of the lines; both establish visual hierarchies with coarse textures and dark values standing out in the display. This is a problem because in the representation scheme these two points (coarseness and darkness) are at opposite ends of the data ranges. This can cause value to be used for identifying relationships even though squareness in the texture is the intended visual variable (Carstensen 1986).

The U. S. Census products, originally published for 1970 census data, that show two variables were studied for communication effectiveness by Olson (1981) (see Figure 5.5c and d on page 75). These maps use two hue patterns for representing two variables, with the part-spectral ranges of yellow to blue and yellow to red being used on the x and y axes, respectively. Olson reports that some of these maps (such as education and income) convey information well, especially for homogeneous regions. She also indicates that these maps are thought to be more authoritative and more innovative than two separate maps showing the same information. In concluding, she suggests that prominent and clear legends are necessary for accurate interpretation of bivariate maps; that both the monovariate map pair and the bivariate map be shown, with the monovariate map pair in a monochrome format; and explanatory notes should be include to the types of information presented. These guidelines are in keeping with both Morgan and Henrion, and Tufte, and are feasible with ARC/INFO's ability to rapidly generate small, monovariate maps.

As a response to the problems of interpreting a spectrally encoded bivariate maps, Steiner (1979, in Eyton 1984) has proposed a complementary color scheme (see Figure 5.6 on page 77). This color system makes use of the mixing of complementary colors (such as red and cyan) to produce a central gray region that highlights the diagonal that represents correlation in a bivariate map. This can be done for unclassed or classed representations of data, and by flipping the order of one tint, negative correlations can be shown as well as positive correlations.

Building on Steiner's proposal, J. Ronald Eyton (1984) developed the complementary color bivariate map with an equiprobability ellipse (see Figure 5.7a and b on page 79). These maps use a modified, 2x2 complementary color range, with an additional class that occurs in the middle of the matrix and represents the central cluster of data. By plotting the two variables to be mapped on a scatter diagram, linearly correlated data will have an ellipse-shaped central cluster. By selecting a percentage of the total number of observation to be included in the central ellipse, a category for the central cluster can be created with the formula (Eyton 1984, 488):

$$X^2 = \frac{\dfrac{(X - X_m)^2}{\sigma_x^2} - \dfrac{(Y - Y_m)^2}{\sigma_y^2}}{1-r^2} - \frac{2r(X - X_m)(Y - Y_m)}{\sigma_x \bullet \sigma_y}$$

$X^2$ = chi-square (1.386 for 50%)
X, Y = the X and Y observations
$X_m$, $Y_m$ = the means of X and Y
$\sigma_x$, $\sigma_y$ = the variances of X and Y
r = coefficient of correlation (Pearson's r)

When this ellipse is displayed in gray, surrounded by the four corners of the data set in white, black, red and cyan, a bivariate map that portrays the central data cluster explicitly (without having a staircase effect) can be created.

There are other possible methods of bivariate mapping that have been postulated as effective communicators of (un)certainty. Two of these, which have not been tested, are color intensity and focus. Neither of these methods place emphasis on a correlation of the data and meta-data variables, but rather use visual variables to highlight portions of the data set, as determined by the meta-data information.

Use of color intensity may prove useful for the cartographic representation of uncertainty, because it allows the highlighting of certain (or uncertain) areas by specifying intense shades for those areas, and less intense for others (see Figure 5.5b on page 75). Because intense colors stand out in an image, this technique would allow the creation of a distinct visual hierarchy that emphasizes certain (or uncertain) values. Generally, for maps that will be used for data communication (as in the DiBiase model), intense colors should represent certain areas and less intense (that is, more gray) colors should represent uncertain areas.

Focus is potentially a useful means of representing uncertainty. MacEachren (1992) presents several possible variations on focus: edge crispness (for external boundaries of points, lines and areas), fill clarity (for internal boundaries within point, line or area symbols), fog (by imposition of an interposing, translucent layer over another symbol), and resolution (point thinning for vector databases or aggregation in to larger area units for raster databases). All of these involve the blending of a symbol with the surrounding parts of the image, thereby eliminating clearly defined regions. In ARC/INFO edge crispness can be accomplished by buffering regions and careful assignment of color in the buffer areas; resolution can be accomplished by thinning points manually or with the ARCEDIT command GENERALIZE.

Another method of bivariate mapping is the use of a fishnet, orthogonal, view to display one data set, with another data set used to color the display (see Figure 9.2 on page 101). This is commonly used as a technique for displaying terrain elevation (with land use/land cover draped over the net by specifying the net's color with the land use symbolization scheme), although there is no restriction on its use for other types of data. When a data layer represents information that is known to be continuous and smoothly changing (such as elevation, air temperature, or air pressure), this type of representation is appropriate. If the uncertainty of a spatial data layer can be shown or assumed to be continuous and smoothly changing, a fishnet representation of that statistical surface (with data values used to specify the color of the net) should be an effective method of conveying meta-data.

Finally, another representation of continuous and smoothly changing data that could be used to display data and meta-data is the use of isolines (see Figure 9.1a on page 99). With this technique data and/or meta-data can be represented, with the use of another display technique if only one is to be displayed with isolines. Like fishnets, isolines are often used for displaying terrain information (as in topographic maps), but isolines can also be used to show data and meta-data. This could be accomplished by using different hues, values, intensities, sizes, or textures to indicate which lines represent data and which represent meta-data.

# Chapter Two
# Producing Displays

As Monmonier and Johnson indicate, the process of designing and creating maps and graphics is a multistep process. There are many things that must be considered in the design process, including the classification of both space and data, and issues such as page layout. This chapter addresses these issues in the context of ARC/INFO and leads into the methods of representing data that are presented in later chapters.

## Classifying Space

A fundamental process in the classification of space is the abstraction of data. MacEachren and Ganter (1990) have describe this as a shifting along a continuum ranging from images to graphics. At one end of this continuum is information in its rawest form: for spatial data, aerial photographs and other remote sensing products (see Figure 2.1, left). Further along the continuum, maps provide an abstraction of images (Figure 2.1, middle). Some of the information that is present in an image is dropped in order to allow symbolization of information that may not be directly perceivable in the image. For example, replacing a dark line through a green area with symbols for a road and a forested area; the road can then be given a label, as well as provide an indication of the number of lanes and access. At the other end of the continuum, graphics allow the use of position in the display to symbolize any variable (Figure 2.1, right). For example, a distance decay graph uses the X axis to indicate distance from a point in any direction, and the point may not be tied to any one geographic location.

A similar continuum is the range of symbols that can be used in maps and graphics (see Figure 2.2 on page 25). At the end of the continuum most similar to images are mimetic symbols. At the end of the continuum most similar to graphics are abstract symbols. Except for a few symbols that are toward the mimetic end of this continuum, most of the symbols that are provided with ARC/INFO are

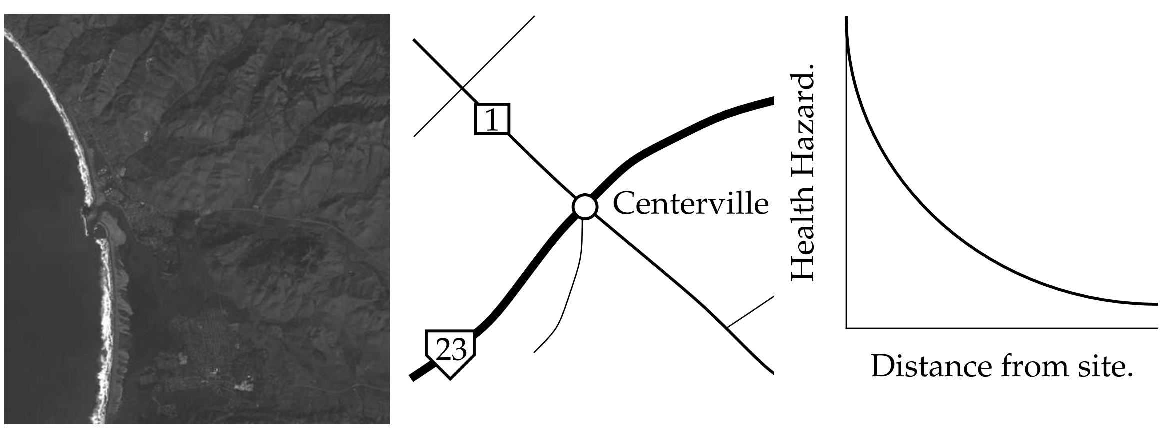



**Figure 2.1** The Image to Graphic continuum. Images approximate what we see, graphics provide abstract representations of, possibly, invisible relations (MacEachren and Ganter, 1990).

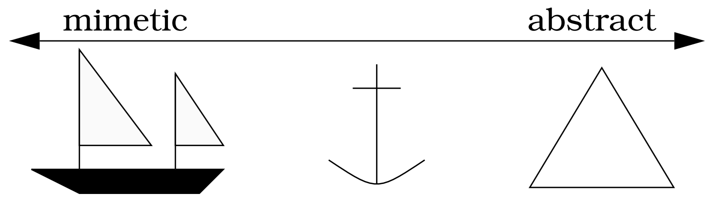



**Figure 2.2** The Image to Graphic continuum applied to symbolization. Symbols can vary from mimetic to abstract. All of these symbols could be used to represent a marina--the audience must be considered in selecting which to use.

abstract--these tend not to have commonly accepted meanings, and thus can be defined by the cartographer as needed.

The degree of abstraction that should be used is dependent on several factors. Most importantly, the data that needs to be mapped must be effectively shown, with a degree of abstraction that is appropriate to the data. For example, if a distance decay model is developed from a set of sample sites, showing the values measured at individual sites may not convey the distance decay as clearly as a graph. If the site locations are important, a map can be generated with the data and an inset showing the graph can be made using ARCPLOT's graphing tools. As mentioned in Chapter One, the audience must also be considered when determining symbolization abstractness. Highly mimetic symbols can convey a sense of simpleness, which may need to be avoided in order to project an image of authority. A third consideration is the means of display; mimetic symbols can be used to minimize the need for/use of a legend. This can be helpful for slides and overhead displays.

## Projections

There are several classes of projections, each with strong points. Equal area projections show the same amount of space on the earth's surface for any given area of the map. Equidistant projections show true distance from a point or along given lines. Conformal projections have a constant scale in every direction from any given point, and because of this latitude and longitude lines meet at right angles. Some projections, such as Robinson's, are not mathematical transformations, but rather, tabular. These projections have been designed for specific purposes (Arthur Robinson developed his projection for world maps that are more visually appealing than others, like the Mercator projection). Other projections include the Mercator and gnomic projections; the Mercator projection shows lines of constant compass directions as straight, and the gnomic projection shows all great circles as straight lines. These two projections are best used for navigation and not for general reference or environmental maps.

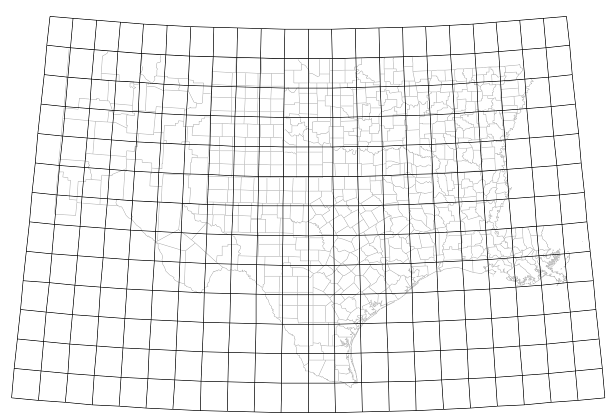


**Figure 2.3** EPA Region Six shown in Alber's conic equal area projection.

For environmental maps, in general, equal area projections should be used. This insures that symbols, especially those based on size, are not distorted on the basis of the base cartographic data (such as county, or state outlines). For the 48 contiguous states, Alber's conic equal area projection should be used; Alber's conic should also be used for smaller regions that are east-west oriented, such as EPA Region Six (see Figure 2.3). The sinusoidal equal area projection should be used for regions that are north-south in extent, such as EPA Region One (see Figure 2.4). Lambert's azimuthal equal area should be used for areas that have the same extent in all directions from a center point (see Figure 2.5).

Because the distortions of any given projection are scale dependant, the use of any specific projection becomes less critical as the map's scale increases (and thus the area shown becomes smaller). For example, a map of a hazardous waste site may need the Universal Transverse Mercator grid for locations within the site; at this large a scale (1:50,000 or larger) use of the UTM projection is more appropriate than an equal-area projection--there will be little, if any, perceivable distortion of areas regardless of the projection used, as long as the projection is centered on the site.

The ARC command PROJECT allows transformation of coverages between projections, as well as realignment of projections. If data is obtained from a national database, the projection (if not included with the database) may be:

```
Project: projection albers
Project: units meters
Project: parameters
1st standard parallel: 29 30 0
2nd standard parallel: 45 30 0
central meridian: -96 0 0
latitude of projections origin: 23 0 0
false easting (meters): 0
false northing (meters): 0
```

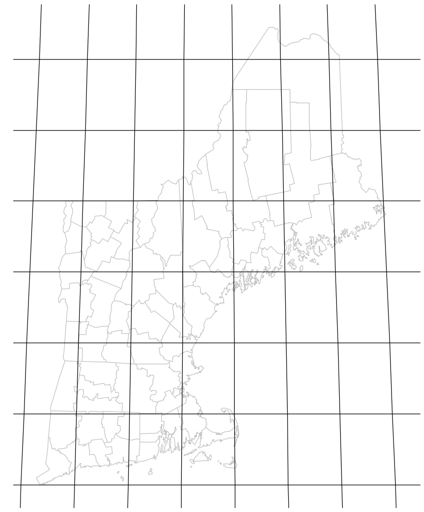


**Figure 2.4** EPA Region One shown in the sinusoidal equal area projection.

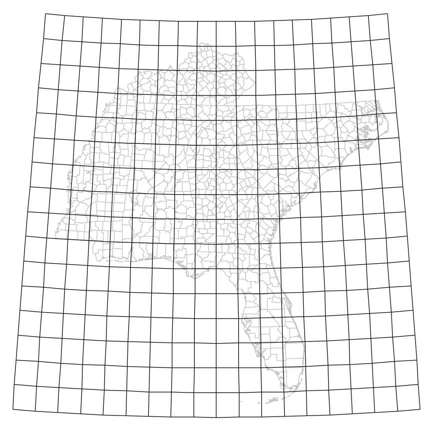


**Figure 2.5** EPA Region Four shown in Lambert's azimuthal equal area projection.

Because this is designed for the contiguous 48 states, any subset of this data should be reprojected for the subset, even if the output projection is Alber's. In particular, the standard parallels should be chosen such that the new parallels divide the region into three equal east-west bands, and the central meridian select bisects the region. This minimizes distortion away from the parallels and prevents an appearance of the whole map leaning in one direction, which results from an off-center central meridian.

## Scale and Generalization

As noted above, selection of scale is critical in the selection of an appropriate projection. Scale is also important in the selection of total area to be mapped, and the relation of the area to the data to be mapped. By using a smaller scale, and thus showing more area, the apparent seriousness of a problem will be reduced. This is because the apparent extent of a problem is reduced by showing more of the surrounding area. Zooming in on a site has the opposition effect--the appearance of symbolization over a large part of the display conveys the idea that the problem is everywhere. Use of zoomed in areas, which allow presentation of detail, with a locator map can combine these two extremes (see Figure 2.6 on page 27). The locator map allows a wide perspective, which, when combined with the large-scale, detailed map, conveys both a micro and a macro reading, as Tufte (1990) suggests is a key component of good presentation graphics.

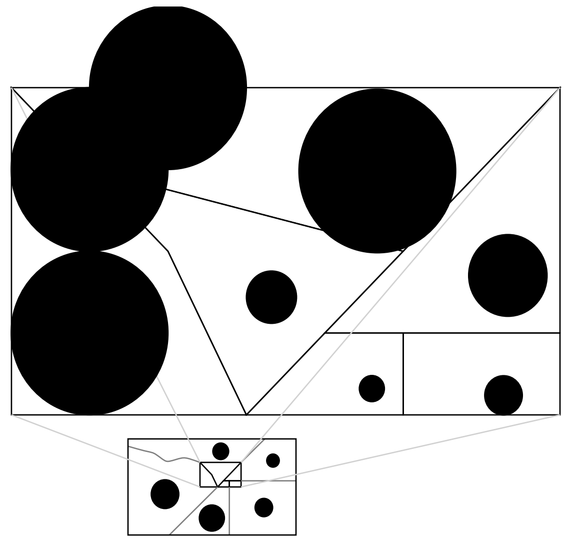


**Figure 2.6** Insets allow focusing on detail and give a 'big picture.' When possible, place them in otherwise unused space.

## Space and Time

With the increased processing speed of single user workstations, and the increased flexibility of ARC/INFO, animation of data--both for analysis and presentation--is becoming feasible. Animation has several possible approaches. Change in time can be used to depict existence, attributes, or change in existence or attributes. Change can also be broken into: looking at different parts of a data set, one after the other; looking at a data set that shows variation in time, in time sequence; or use of progression in time to show another data variable (this is comparable to use of an axis of a graph to indicate change in time rather than change in space).

There are two primary methods of creating animations in ARC/INFO, and both require AML programming. The most flexible method is the use of AML driven interactive map composition. This allows the display of an object (for example a point location, such as the population center of the United States) that changes over time. The object can be drawn at one location, erased with the MDELETE command, and redrawn at a new location. The other method of animation is the use of GRID to display changes in area data. The cliche 'Raster is Faster' is still true enough to make a difference. Separate grid layers can be generated, each of which shows the change in areal extent of a phenomenon. Each layer can then be draw with successive calls to a display routine.

## Classifying Data

Data to be mapped for presentation should generally be classified (into five or six groups) in order to aid ease of interpretation. When drawing features in ARCPLOT, the commands: ARCLINES, LABELMARKERS, POINTMARKERS, and POLYGONSHADES, as well as AML driven RESELECT's, allow the use of lookup tables to define a data to symbol relationship for feature display (the CLASS command also allows grouping of data). These tables allow user defined ranges for cartographic products, rather than the default of directly relating the data item identifications with the symbol set identifications. These ranges must be determined by the cartographer and range types include: quantiles, equal intervals, geometric progressions, mean and standard deviation intervals, and natural breaks (the Jenks' Optimal classification) (see Figure 2.7). Certain data sets may need to be classified on the basis of predetermined breaks (such as maximum allowable concentration of a pollutant). This can be accomplished by manually specifying all breakpoints, or by use of another classification scheme, with the externally defined break added in. For example: run the Jenks' optimal AML for five classes, note the breakpoints, then run the manual specification AML on a new lookup table and specify the Jenks' breakpoints plus the predefined point.

### Manual Classification

For both the CLASS command and lookup tables, break points can be chosen manually, by exporting the data to a statistical package, or with the assistance of the ARCPLOT command STATISTICS. The command syntax is:

```
STATISTICS <cover> <feature_class>
        <cover> - contains the items to be classed;
        <feature_class> - includes POINTS, ARCS and POLYS.
```

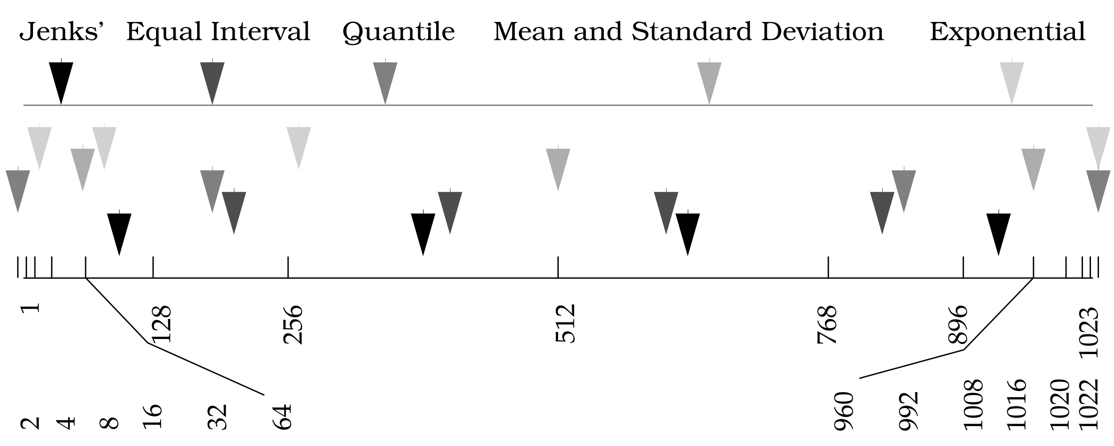



**Figure 2.7** Breakpoints set by different classification schemes. Equal Interval is close to the statistically optimal (Jenks'), for this example. Quantile separates relatively similar values at the extremes of the distribution. Mean and Standard Deviations fail because the distribution is bimodal, not normal; Exponential fails by grouping all of the items greater than 256 into one category.

STATISTICS has several subcommands:

SUM <item>
MEAN <item>
MINIMUM <item>
MAXIMUM <item>
STANDARDDEVIATION <item>
<item> is a field in the attribute table of <cover>.
END - indicates all subcommands have been entered.

These values can be used to calculate mean and standard deviation break points, as well as quantiles and geometric progressions.

There are several methods of defining classes within ARC/INFO, as well as several methods of determining class breaks. As MacEachren (1992) notes, interval and quantile classification systems, which are automated parts of the CLASS command, generally are not the best methods for grouping data and therefore will not be discussed here. The syntax for manually specifying intervals with the ARCPLOT command CLASS is:

CLASS MANUAL <#classes> <break...break>
<#classes> - the number of classes that will be generated;
<break...break> - the numeric class breaks. There must be <#classes> - 1 values.

CLASS NONE - turns off the classification scheme.

Until the CLASS NONE command is given the classification remains in effect, and will cause all subsequent uses of ARCLINES, LABELMARKERS, POINTMARKERS, and POLYGONSHADES to be classified.

Another method of classifying data is to use lookup tables. These are INFO files that perform a similar function as the CLASS MANUAL command. A manual procedure from ARC for the creation of a lookup tables is:

1) Use PULLITEM to extract the attribute table item that holds the data values;
2) Use ADDITEM to add a numeric field called SYMBOL;
3) Enter INFO, SELECT the table and PURGE the old data;
4) ADD records to the lookup table;
    a) Specify the breakpoint value for numeric data, or the alphanumeric for text data;
    b) Specify the symbol number;
5) SORT on the data field (not the SYMBOL field).

It is possible to automate the creation and specification of lookup tables within ARCPLOT. SETMAN.aml allows the manual update of symbol values within a lookup table, as well as creation a new table. The AML handles both nominal and numeric data. The syntax is:

SETMAN <lookup_table>
<lookup_table> will be created if it does not exist, existing tables will be selected for update.

### Natural Breakpoints

Because the Jenks' Optimal classification scheme is generally considered to provide the best classification of numeric data (MacEachren 1992), it should be included in ARC/INFO (and with any mapping program). Unfortunately it is not, but ARC/INFO's does provide a macro language, commands to export data, and a method for calling system programs. These can be combined to allow the automated generation of the breakpoints for Jenks' optimal classification. The AML, JENKS.aml, exports the data to be classified, calls the C program, jenks (which must be in the executable search path), that calculates the breakpoints, and constructs a lookup table. The cartographer must still specify symbol matches though. Please note that the routine generates a lookup table that contains more information than is required for ARCPLOT's use (although ARCPLOT can still use it); this additional information is included for use by the AML's presented in later chapters. The additional information is: an initial line that is one smaller than the smallest data value (this is included only in non-nominal lookup tables, including the output from Jenk's); a column that contains the number of values in the associated coverage in the first record, and the number of values in each category in the following records; and a column that contains the coverage name in the first record, an 'n' (for nominal data), an 'o' (for ordinal data) or an 'r' (for interval/ratio data), the coverage type in the third record (point, arc, etc), and for route and section lookup tables, the route name in the forth record. This additional information allows these AMLs to select the coverage it is based on and, for legends, provide category totals and appropriate symbolization.

The C program jenks.c can be compiled on a workstation with an ANSI C compiler (including Data General's DG/UX 5.4) by using the command line:

```
cc -ansi -o jenks jenks.c -lm
```

The command syntax for using JENKS.aml is:

```
JENKS <cover> <feature_class> <data_item> <out_lookup> <classes>
    <cover> - the coverage that the lookup table will be created for;
    <feature_class> - the feature type (point, line, poly, etc) of <cover>;
    <data_item> - an interval or ratio level data field in the attribute table of <cover>;
    <out_lookup> - the name of the lookup table to be generated;
    <classes> - the number of data classes in the output lookup table.
```

### Eyton's Equiprobability Ellipse Bivariate Classification

The uniqueness of Eyton's ellipse as a bivariate mapping scheme requires that a column be added to the attribute table of the data to be mapped. As with the calculation of Jenks' optimal classifications this is best done with a combination of macro and C program. EYTON.aml requires two ratio data items--the classification system is based on the parametric statistic, Pearson's r. It also requires that the system program, eyton, be in the executable search path. Like jenks.c, the eyton.c program can be compiled on a workstation with an ANSI C compiler by using the command line:

```
cc -ansi -o eyton eyton.c -lm
```

The command line syntax for using EYTON.aml is:

EYTON <cover> <feature_class> <data_item_1> <data_item_2> <out_lookup>
{classes} {chi_square_value}
<cover> - the coverage that the lookup table will be created for;
<feature_class> - the feature type (point, line, poly, etc.) of <cover>;
<data_item_1> <data_item_2> - interval or ratio data fields in the attribute table of < cover>;
<out_lookup> - the output lookup table;
{classes} - the number of classes on each axis, valid codes are 2 (default) or 3;
{chi_square_value} - a value for the selecting the number of points in the central ellipse, defaults to 1.386, which is 50% of the observations.

## Symbol Value Update

The lookup tables generated by these commands may require the modification of the symbol numbers. There are several ways this can be accomplished: use of '&SYS ARC INFO' to enter INFO from ARCPLOT; manually declaring a cursor and using it to update the lookup table; using the AML given above for manually changing values; or use of the AML, SETAUTO.aml. This AML updates a lookup table by replacing the SYMBOL values with an ordered set of numbers. It requires a starting value, a step value, and can optionally be given two other values. Nonlinear progressions can be specified by including a 'scale' value, and decreasing numbers can be obtained by specifying a 'subtraction_value.' These non-linear progressions should be used when a SYMBOL value is used for specifying something other than a predefined symbol, such as symbol value. In this case value should range from black to very light grey with a greater change in the black/dark-gray values (value differences are easier to discriminate for lighter values). For five classes, the use of 0 50 0.8 and, if necessary, 69 as the subtraction value, will set up an appropriate value progression.

SETAUTO <lookup_table> <start_value> <step_value> {scale} {subtraction_value}
<lookup_table> - the lookup table to be updated;
<start_value> - the beginning value of the progression;
<step_value> - the value used to change the beginning value;
{scale} - an exponent that is applied to <step_value>--defaults to 1;
{subtraction_value} - value that a symbol value will be subtracted from in order to generate decreasing progressions

## Unclassed Maps

Unclassified maps (those that are continuously symbolized) can be created in ARC/INFO, but the symbolization schemes available can be limiting and the potential improvement in the ability to depict data (as suggested by Monmonier, 1976) may not justify the effort for presentation of data. For exploration and analysis however, unclassified maps are an approach that may be worthwhile, although software/hardware limitations can impinge on truly unclassified displays. Eight bit plane graphic displays can only display 256 colors at once, thereby preventing unclassed display of data sets with more than 256 values. This can be partially circumvented by using the Jenks' optimal classification system to create a lookup table with 256 classes. For data sets with less than 256 data values or for systems with 'true color' (24 bit plane) displays, use the manual lookup table definition AML (SETMAN.aml) to define a nominal lookup table. A limitation of 'true color' displays is that 24 bit graphics systems can only generate 256 shades of gray. So, in general, for 'unclassed' maps, create a lookup table with 256 categories (or fewer if there are not 256 data values); once a lookup table is created, symbol values must be assign on the basis of the lookup table's item value. SETUNCL.aml accomplishes this. The command syntax is:

SETUNCL <lookup_table> {scale_factor} {subtraction_value}
<lookup_table> - the table with symbol values to be updated;
{scale_factor} - an exponent for generating non-linear scaling, defaults to 1;
{subtraction_value} - allows creation of descending scales.

## Page Layout

Once the information to be displayed is classified, the data must be displayed. This is the process of page layout, which can be broken into: layout within parts and layout within the whole. Layout within parts of a display involves the selected use of visual variables to highlight a part or parts of a data set or display. This is the establishment of graphic hierarchies and involves two techniques: visual isolation, and visual levels (MacEachren 1992, 27) (see Figures 1.1 on page 2 and 1.7 on page 14).

Visual isolation is the use of visual variables to give the appearance of the separateness of part of a display. The visual variables that can accomplish this include location, size, focus, value, hue, saturation, texture and orientation. Location is the most obvious control--display items that are not close together will not generally be associated (see Figure 2.8). Size can influence isolation because any feature that is drawn smaller than the area allocated to it, will appear separate from the surroundings (this is done in cartograms). Focus influences isolation because unfocused, fuzzy symbols blend into the surrounding, reducing isolation. Value, hue, saturation, texture and orientation influence isolation by enabling one item to be displayed in a manner that is noticeably different than its surroundings (dark blue surround by yellow, for example) and thus isolating it.

Like visual isolation, visual levels are used to separate an item from its surroundings, but whereas visual isolation seeks to accomplish that in the x-y space of the page, visual levels give the appearance of separation in the third dimension. The visual variable that control visual levels are: size, value, saturation, texture, focus and hue. These visual variables all play a part in depth perception. Size controls visual levels because things that appear larger, appear to be closer. Value controls visual levels by control of contrast from the display background--objects with high contrast will appear above the background (see Figure 2.9 on page 32). Saturation controls contrast because intense colors appear to be closer than less-intense colors.

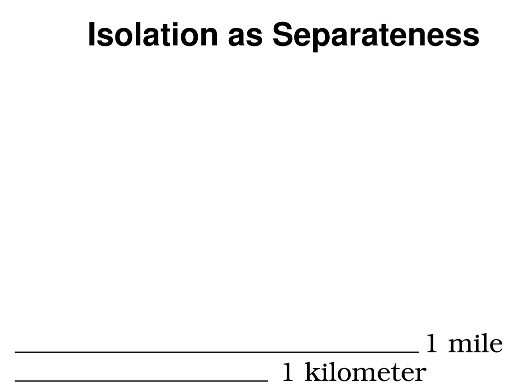



**Figure 2.8** Visual isolation. The letters in each word 'go together,' as do the lines and their labels, but not the lines and the display title.

**Figure 2.9** Visual levels. Darker value usually is a foreground (at a higher level) than lighter values, particularly on white backgrounds.

Texture influences visual levels because coarse textures appear to be closer to the observer. Focus influences visual levels because sharply focused objects appear to be closer than fuzzy objects. Hue influences visual levels because, at a constant value and saturation, yellow is more noticeable than other colors, but the influence of hue can be hard to predict and control.

In designing a page, the layout within the whole page involves the use of space (and for animation, time) to present different parts of a message. The type of display has an influence on what can, and should, be said on an image. Paper displays can be of any size, but two are of more importance: 8.5 by 11 inch paper, and full size electrostatic plotters. 8.5 by 11 inch paper is probably the most common format for graphic publication. Its small size restricts the amount of information than can be shown on one page, but generally high resolution and the ability of the reader to study the graphic at length permit a great deal of information communication. The portrait orientation, which is normally used with text, limits the left-right extent of a display, so graphics that have a large left-right extent should be displayed as landscape, when the graphic cannot be redesigned to take advantage of the more common reading orientation. Generally the primary part of the display should be shown with a maximum left-right extent. Titles can then be placed above, and additional information such as legends and locator diagrams can be placed below the primary part of the display.

Large size paper products, such as posters, can be created with less emphasis on the orientation of the page, and more emphasis on the size and shape of the area of interest. Orientation of the page should follow the larger axis of the data set. Additional information should then be placed around the main part of the graphic, in order to take advantage of 'dead space.'

Slides and overheads limit the amount of data that can be displayed because of the distance that they force on the viewer (see Figure 2.10 on page 34). The farther a viewer is from the display, the smaller the display will appear, limiting the visibility of detail. This implies that the page layout decisions for slides and overheads require greater generalization in both space and data. Only the most important information needed to support a point should be included on the display. As such, additional information such as legends and locator diagrams should not be included. This allows the maximization of the available space for data display, and like posters, the orientation of the display should be aligned with the larger axis of the data area. This loss of data on the graphic is offset by the explanations of the presentor of the slide, overhead or video.

Video involves both high resolution computer monitors, and lower resolution National Television Standards Committee displays. Video displays must, like slides, take into account the distance from the viewer, as well as lower resolutions and display flicker. Minimize the amount of ancillary information, in order to maximize the amount of space available for the primary data. Like slides, this loss of visual information must be offset by audio explanations.

### Titles and Type

The text that is used to label a map or graphic should be selected carefully. This involves the determination of what should be labeled, how it should be labeled, where the labeling should be placed, and what type of lettering to use. Which items must be labeled should be identified during the planning phases of the map design, particularly in identification of the communication goal--label those things that are necessary to accomplish the goal. Additional labeling may be necessary for locational reference, etc., but this ancillary labelling should be kept to a minimum in order to emphasize the new information.

As with the amount of features that get labelled, the amount of text in each label should be kept to a minimum. Often leading 'a,' 'an' and 'the's can be dropped from labels. Phrases such as 'the city of,' 'plot of' or 'Legend' can also be dropped without loss of information and result in an improvement in communication by getting rid of the clutter.

Labels should be placed on or near the object they label. Locations for titles can include prominent positions such as top center, but may also include making use of otherwise empty space (such as using the Gulf of Mexico's space in a map of Florida). Labels of point features should, when possible, be put at the upper left of the feature. Labels of linear features should be placed on straight sections of the feature, or if necessary fit along a smooth curve, and be oriented for maximum readability (top up, whenever possible). Area features should be labeled inside their boundaries, when possible. Although it is not entirely unavoidable, text should not be broken up by linear features or area boundaries; it may be preferable to have a break in the line, although this can be difficult to accomplish in ARC/INFO.

There are two general classes of fonts that are used for cartographic displays, roman and sans-serif, both of which can be *italicized*, <u>underlined</u>, etc. Roman fonts, such as the font used for this text, have 'serifs'--the small extension at the end of the strokes that make up the character. These fonts usually are more legible than fonts without (sans) serifs, particularly for small text. Sans-serif fonts are simpler, but are generally best used only for titles and other large text. For either type of font that may be used to label a body of water, it is cartographic tradition to use an italic font. This may connotate a sense of flowing.

The size of text influences legibility, particularly when the speed of a presentation is controlled by the presenter, rather than the reader. For paper displays that the reader will be able to study at length, text as small as 5 points (0.07 inches) can be used. Differences in text sizes should be at least 35% (0.07\5 pt, 0.095\7 pt, 0.13\9 pt, etc.) (Keats, 1982). For paper displays which are

**Poster Title**

**Poster Title**

**Figure 2.10** The apparent size of 16 point text from 8 feet away (top) is the same as 6 point lettering from 1.5 feet (bottom) (MacEachren 1992, 71).

not available in depth study (such as flip charts and posters) the following suggestion for slides and overhead are more appropriate, although detail information could be included in posters.

For slides and overheads, not only must there be less text, but the text should be larger (see Figure 2.10 on page 34). As a general rule text should be at least 18 point with larger differences than paper (50%) (MacEachren 1992, 71). Video displays also need to make use of large points sizes, because of lower resolution and with some systems, flicker.

### Insets and Legends

There are two main types of insets: locator diagrams and additional graphics. Locator diagrams show the location of a smaller area, in relation to a larger and presumably more well known area (see Figure 2.11). When multiple graphics are generated for the same region, only one locator diagram should be needed, and it may be advisable simply to make the locator diagram a separate graphic. For displays which need integrated location diagrams, they should be visually isolated--the diagram should be secondary to the main display. Therefore avoid bright colors, if not colors entirely, but ensure the area of the main display is readily discernible.

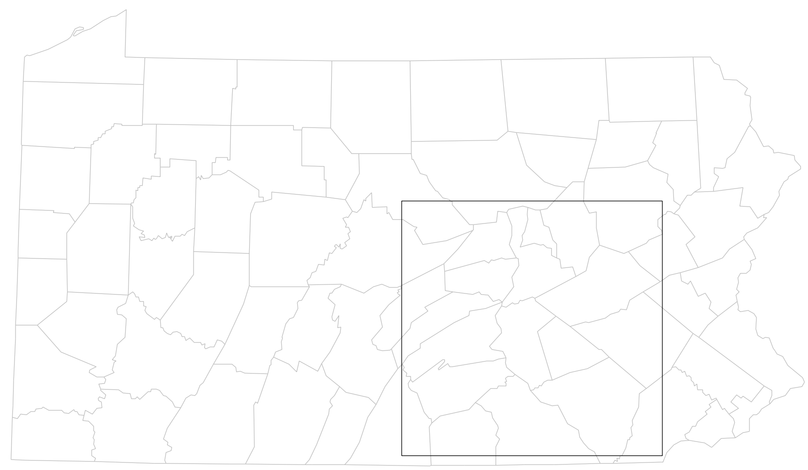


**Figure 2.11** A simple locator diagram.

Other graphics include legends, north arrows, scales and, for bi- or multivariate displays, monovariate maps. Legends must convey the information necessary to extract information from the main display. North arrows assist in orientation when north is not at the top, which it does not have to be. For maps to be presented to an international audience, include a north arrow, because south may be expected at the top (as is the Chinese custom). For maps where a large and hopefully well-known area is displayed (the United States or a subset of them, for example), a north arrow is not needed. Map scales, like north arrows, are not necessary for maps of well known areas and generally are not needed, unless the map may be used for measurement of extent. As suggested by Olson (1980), monovariate maps may be helpful in assisting readers who are not familiar with bi- or multivariate maps. These single variable maps should be designed to be both isolated and at a lower level in the visual hierarchy than the main display. Monochrome displays may be the most appropriate technique for all but nominal data.

Because of the requirements of different media, the amount of information that should be carried in a legend varies, as does the need for a legend. Because of the generalization that goes into slides and video displays, the need for legends should be minimal, particularly when the presenter is available to answer questions. A slide for a legend may be a waste, because it cannot be referred back to, and handing out a legend before a presentation may only serve to focus attention on the legend, not on the data that is being presented.

Paper displays, particularly those where the author is not available to answer questions, require the most thoughtful use of additional graphics. Explanations of data must be written out on the graphic. For bi- and multivariate displays, the mapping technique should be explained, and monovariate maps should be included. Legends should present the character of the data (data level, continuous or discontinuous, classification system used). This can, in part, be accomplished by defining discreet variable symbols separately with a label on each symbol, and by defining continuous variables as a continuum with breakpoints, rather than midpoints, labeled (see the legends in Figures 5.1 on page 67 and 5.2 on page 69, for example).

An approach to legend design that may be most helpful to inexperienced map readers is natural legends (see Figures 6.2 on page 86 and 9.1 on page 99). Natural legend design is particularly helpful with abstract symbolization, such as isolines, which can be difficult to interpret by a novice (especially when only given a contour interval). By providing a legend that shows isolines on a surface and spot values, the concept of a continuous surface and how data values are shown can be communicated. Isoline data, particularly when several data items are shown together, can be displayed as a single-variable, fishnet inset; this will convey the concept of a continuous surface and allow analysis of individual data sets, in addition to the presentation of the interplay of multiple data sets in the main display.

## ARC/INFO Hints

There are two lines that should be added to the $HOME/app-defaults/ Mwm file. These lines allow ARC to automatically place popup windows:

```
*clientAutoPlace: FALSE
*interactivePlacement: FALSE
```

There are a few general tools available in ARCPLOT that may assist in the design or production of maps and graphics. When using 'DISPLAY 9999,' ARCPLOT defaults to a black background; this is a problem with the 'true color' display of value--black objects will be visible on paper, but invisible on the display, and vice-versa for white. The remedy for this is to include the following line in either a .cshrc or .login file:

```
setenv CANVASCOLOR WHITE
```

When using nineteen inch monitors, it is possible to specify a display canvas size that has (as far as ARC is concerned) the same dimensions as an 8.5 by 11 sheet of paper. The dimension for portrait layout are 691 by 930. The dimensions for landscape layout are 896 by 727.

All AMLs can be executed in two ways: specifying an AML path and running with the &RUN directive; or by placing the AMLs in an atool directory and running them as a standard command. The $ARCHOME/atool directory can be used, or the AMLs can be placed in another directory and linked into $ARCHOME/atool subdirectories. The AMLs can also be placed in any other directory, with subdirectories for each module (arc and arcplot), and use the &ATOOL directive to specify the directory path (see the &ATOOL page in the *AML User's Guide*).

The AMLs discussed in this document, for the most part, follow a naming convention. For the first character:

P = point or node displays.
L = line, route or section displays.
C = choropleth area displays.
G = graduated symbol area displays.
D = dot density area displays.
R = raster (grid) area displays.
S = surface displays.

For each additional character (one for each variable displayed):

H = hue.
V = value.
I = intensity.
O = orientation
S = shape.
T = texture.
Z = size.
B = box (used with points only).
C = circle (used with points and graduated symbols only).
G = cartoGram (used with graduated symbols only).
D = density (used with dot density only--all are DD).
F = fishnet (used with surfaces only).
AO = angle (for single variable displays only: cao, pao)
ISO = isoline (used with surfaces only).

Other codes include:

P = pie (graduated circles with subdivisions).
L = legend (gcl, ggl, gpl, sfl; al - single variable; ol - orientation)
BL = bivariate legend.
CC= complementary colors.
DH= dual hue colors.
EE = equiprobability ellipse (and eel - legends).
IL = choropleth area intersecting lines with PAT data (and cill - legends).
TL = choropleth area intersecting lines with lookup tables.
RGB = red, green and blue (and rgbl - legends).

# Chapter Three
# Point Symbolization in ARC/INFO

For point symbolization, the visual variables of hue, orientation and shape can be changed by the use of an INFO lookup table that relates nominal distinctions to different symbols; the visual variables of size and value can be used to symbolize ordinal and interval/ratio data either with a lookup table or with data from the file attribute table. For lookup tables use, one of the methods discussed in Chapter Two should be used to generate a lookup table. The AMLs discussed in this chapter generally require some of the additional information that the lookup table generating AMLs in Chapter Two provide.

## Monovariate Symbolization

The macros in this chapter demonstrate the use of AML to accomplish map design. The AMLs become progressively more complex in order to set the stage for bi- and multivariate mapping, which, for all intensive purposes, must be done with AMLs (most of which also require the use of cursors).

Each of the examples makes use of the markers that are provided with ARC/INFO: a default set of markers in the markerset file, PLOTTER.MRK; and several other available markersets: BW.MRK, COLOR.MRK, MINERAL.MRK, MUNICIPAL.MRK, OILGAS.MRK, USGS.MRK, and WATER.MRK. Most of these are displayed in the *Map Display and Query* guide's Appendix B, and the ability to create new symbols sets is discussed in Chapter Three of *Map Display and Query*.

### Nominal Data--Hue

Hue is best used for nominal data, and the COLOR.MRK markerset provides fifteen symbols of the same shape, but each with a different color. All of the other markersets, except BW.MRK, have the same symbol in black, red, green and blue. An appropriate markerset can be selected with the MARKERSET command, and used with a lookup table that relates changes in data values to changes in hue. POINTMARKERS can then be used in conjunction with the lookup table to plot the points. This is the method used in PH.aml; see Figure 3.1a.

PH <cover> <data_item> <lookup_table> <markerset>
<cover> - the coverage to be displayed;
<data_item> - an item in the point attribute table of <cover>;
<lookup_table> - an info table that relates <data_item> values to markers;
<markerset> - a markerset file.

### Nominal Data--Orientation

Orientation is best used for nominal data, and the MINERAL.MRK markerset provides some ready-to-use symbols that vary in orientation (such as markers 143 through 149). PH.aml (above) uses command line arguments to select this markerset and draw these symbols; see Figure 3.1b.

### Nominal Data--Shape

Shape should only be used for nominal data. All of the markersets, except COLOR.MRK, have symbols that change in shape. The macro PS.aml uses cursors to extract the coverage name and type (point or node) from the lookup table (which

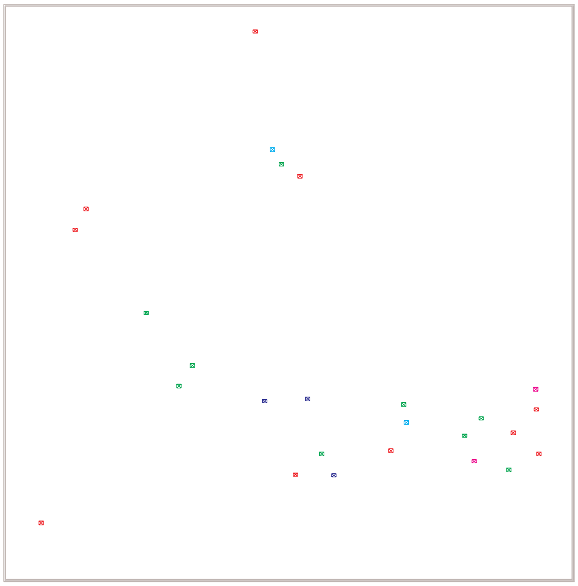


**Figure 3.1a: PH.aml**
Nominal data displayed with hue using the color.mrk markerset.

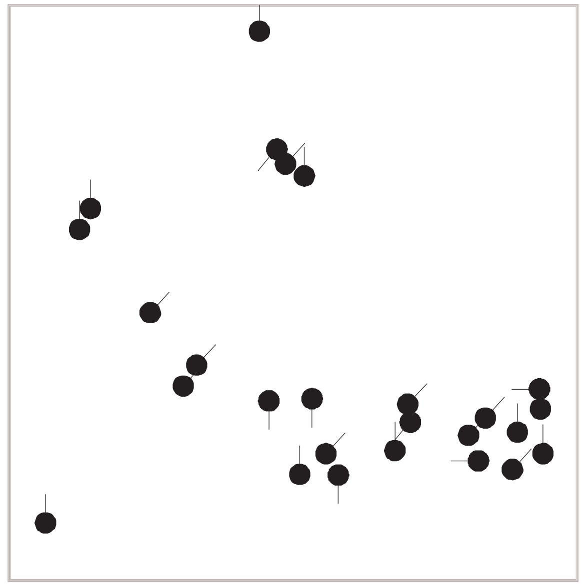


**Figure 3.1b: PH.aml**
Nominal data displayed with orientation by using symbols from mineral.mrk.

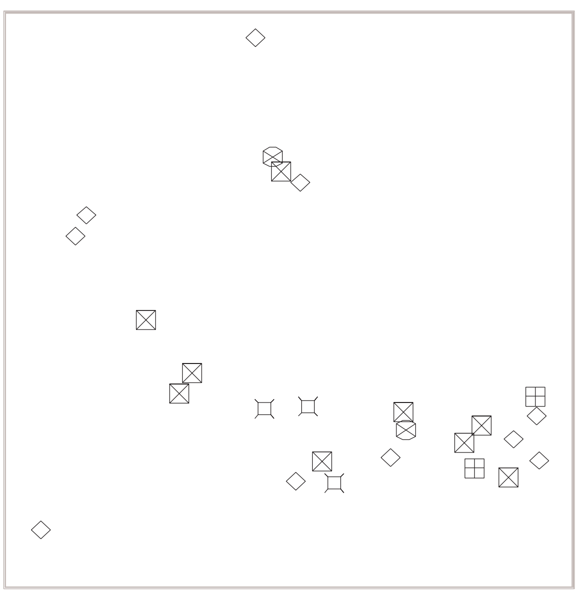


**Figure 3.1c: PS.aml**
Nominal data displayed with shape.

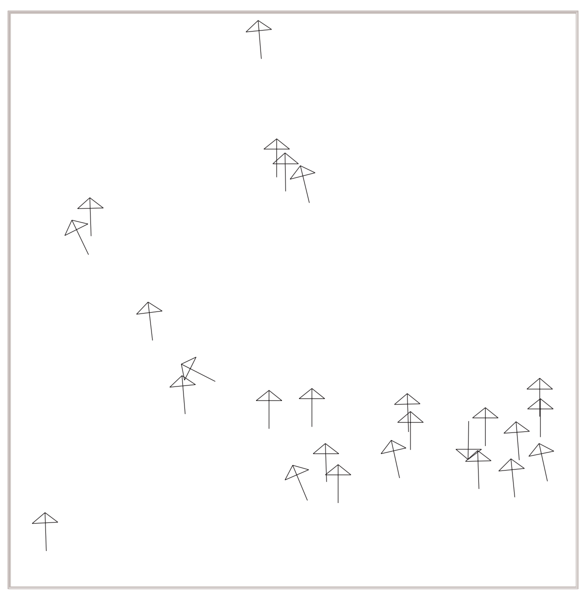


**Figure 3.1d: PAO.aml**
Ratio data displayed with symbol orientation.

must be in the format generated by the lookup table AMLs discussed in Chapter Two), and then draws the points with the lookup table; see Figure 3.1c.

PS <lookup_table> {markerset}
<lookup_table> - a lookup table generated by SETMAN;
{markerset} - a markerset file--the current markerset is the default.

### Ordinal to Ratio Data--Orientation

Orientation is best used for nominal data, but ARC/INFO allows the specification of orientation that can be used with higher data levels. For point symbols, the MARKERANGLE command allows specification of the drawing angle of the current symbol. The macro, PAO.aml, uses cursors to determine the minimum and maximum data values to be displayed, and then uses these values to adjust the drawing angle of the selected marker; see Figure 3.1d.

PAO <cover> <orientation_item> {markerset} {markersymbol} {markersize}
{max_angle} {point|node}
<cover> - the coverage to be displayed;
<orientation_item> - a numeric data field in the attribute table of <cover>;
{markerset} {markersymbol} - specify a marker--the current marker is the default;
{markersize} - drawing size of the marker--the default is 0.15 inches;
{max_angle} - angle of the maximum data value--the default is 179 degrees;
{point|node} - symbolize point (the default) or node feature.

### Ordinal to Ratio Data--Value

Color value is best used for ordinal data. For point symbols, the MARKERCOLOR command allows specification of the drawing color of the current symbol. ARC/INFO has a Munsell-like color specification: the HLS color model; the parameters for the HLS model are <hue> <lightness> <saturation>. Hue is an integer from 0 to 360 (red = 0, green = 120, blue = 240). Lightness is an integer from 0 to 100 (black = 0, white = 100). Saturation is an integer from 0 to 100 (gray = 0, fully saturated = 100). Specification of changes in value are accomplished by: setting hue to any valid number; adjusting lightness to control the grayness (for output on white paper, generally use a percentage less than 90); saturation must be set to 0. The macro, PV.aml, uses MARKERCOLOR to control value; see Figure 3.2a.

PV <value_lookup> {markerset} {markersymbol} {markersize} {hue} {intensity}
<value_lookup> - a lookup table generated by SETMAN or JENKS:
{markerset} {markersymbol} - specify a marker--the current marker is the default;
{hue} {intensity} - specify a color that will be value shaded:.

### Ordinal to Ratio Data--Size

Size can be used in point symbols for ordinal and interval/ratio data, and within ARC/INFO changes in size can be achieved in two ways: for circles and boxes, ARC/INFO provides the SPOTSIZE/POINTSPOT command pair to generate circle or box point symbols; for other symbol shapes, size is changed by changing the drawing size of point symbols. The MARKERSIZE command allows the specification of the size of the current point symbol. This macro, PZ.aml, uses markersize to display a ratio data set; see Figure 3.2b.

**Figure 3.2a: PV.aml**
Classed ratio (ordinal) data displayed with color value.

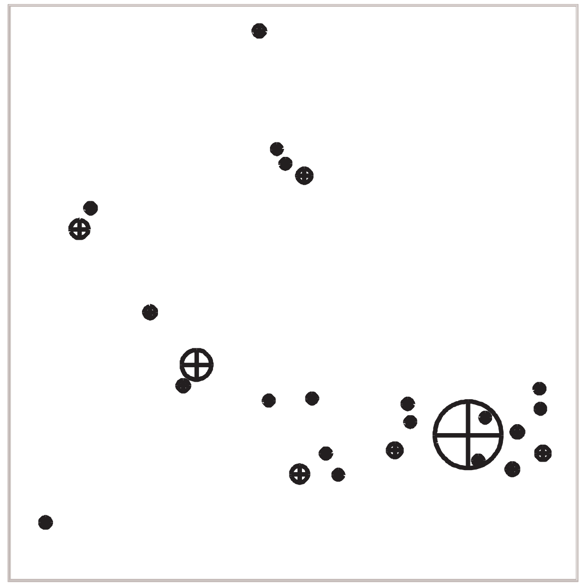


**Figure 3.2b: PZ.aml**
Ratio data displayed with symbol size.

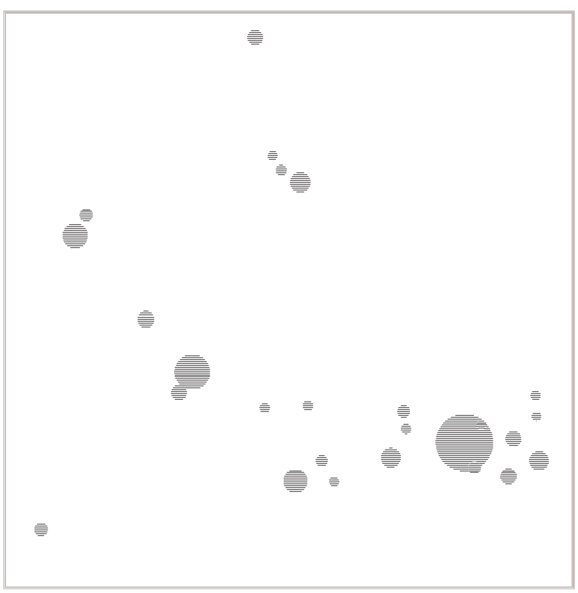


**Figure 3.2c: PC.aml**
Ratio data displayed with graduated circles.

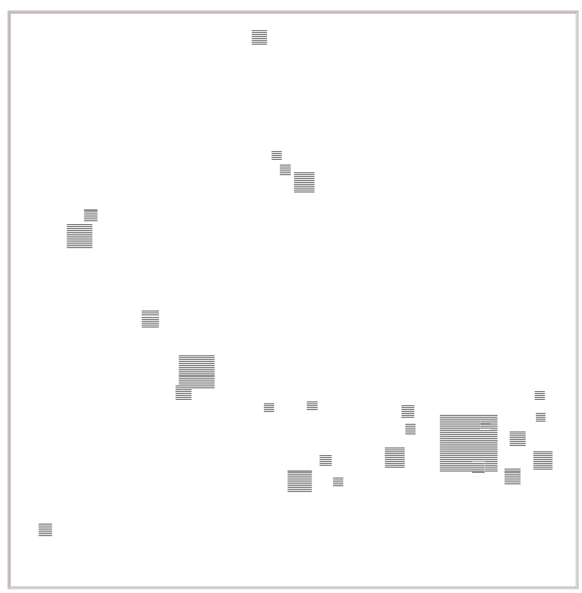


**Figure 3.2d: PB.aml**
Ratio data displayed with graduated boxes.

PZ <cover> <size_item> {size_exponent} {size_factor} {minimum_size}
{markerset} {markersymbol} {point|node}
<cover> - the coverage to be displayed;
<size_item> - a numeric data field in the attribute table of <cover>;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

All of the size symbolization AMLs use the formula:

$$\text{size} = \text{size\_factor}\,(\text{normalized\_data\_value}^{\text{size\_exponent}}) + \text{minimum\_size}$$

This is just a formula for a line in x (normalized_data) and y (symbol size) space (with size_exponent set to one), with the ability to create non-linear progressions (with size_exponent not equal to one). The slope of the line is determined by size_factor, and the y-intercept of the line is minimum_size. The data is normalized prior to calculating the size in order to set the minimum data value to minimum_size; data values are normalized by:

$$\text{normalized\_data\_value} = (\text{actual} - \text{minimum}) \,/\, (\text{maximum} - \text{minimum})$$

### Ordinal to Ratio Data--Graduated Circle Size

Symbol size is good for either ordinal or interval/ratio data. For the generation of graduated circle (or box) symbols, ARC/INFO provides two commands that allow rapid generation of these maps: SPOTSIZE and POINTSPOT. SPOTSIZE must be given before POINTSPOT can be used. SPOTSIZE allows the creation of point symbols that can be linearly or exponentially scaled; of the two, exponential scaling is generally preferred, although the command line syntax is more complicated. Once SPOTSIZE has been given, POINTSPOT can be used to create graduated symbol maps, with either circle or box symbols. These two macros generate circle (PC.aml) and box (PB.aml) symbols (see Figures 3.2c and d).

PC <cover> <size_item> {minimum_size} {maximum_size} {point|node}

PB <cover> <size_item> {minimum_size} {maximum_size} {point|node}
<cover> - the coverage to be displayed;
<size_item> - an interval or ratio data field in the attribute table of <cover>;
{minimum_size} - size of the smallest symbol--defaults to 0.05 inches;
{maximum_size} - size of the largest symbol--defaults to 0.5 inches;
{point|node} - the type of <cover> to be displayed--defaults to point.

## Bivariate, Monochrome Symbolization

Most of the AML's presented in this section make use of shape to indicate a nominal distinction. An appropriate symbol set should be selected (or created) and one of the bivariate methods should be selected on the basis of the second variable's type.

### Two Nominal Data Sets--Shape and Orientation

For nominal data and meta-data, shape and orientation can be combined to create bivariate point symbolization. Shape should generally be used for the primary data, and orientation for the less important data, or meta-data (varying

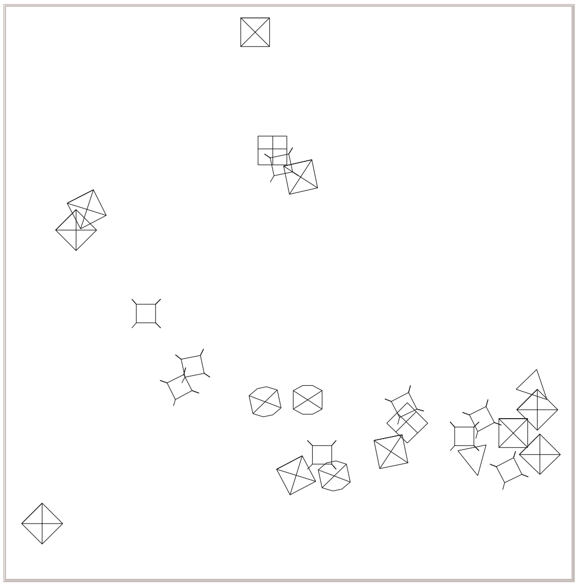


**Figure 3.3a: PSO.aml**
Two nominal data sets displayed with shape (primary data) and orientation (secondary data).

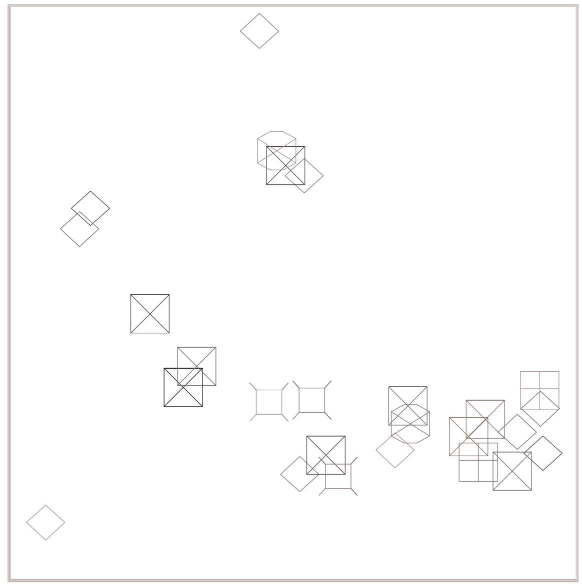


**Figure 3.3b: PSV.aml**
A nominal data set displayed with shape and an ordinal data set displayed with value. This is better for meta-data than a second data set.

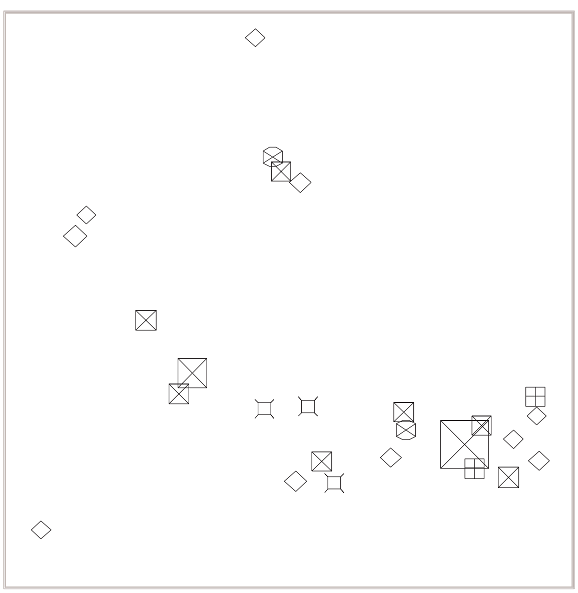


**Figure 3.3c: PSZ.aml**
A nominal data set displayed with shape and a ratio data set displayed with size. This is more effective than value for a second data set.

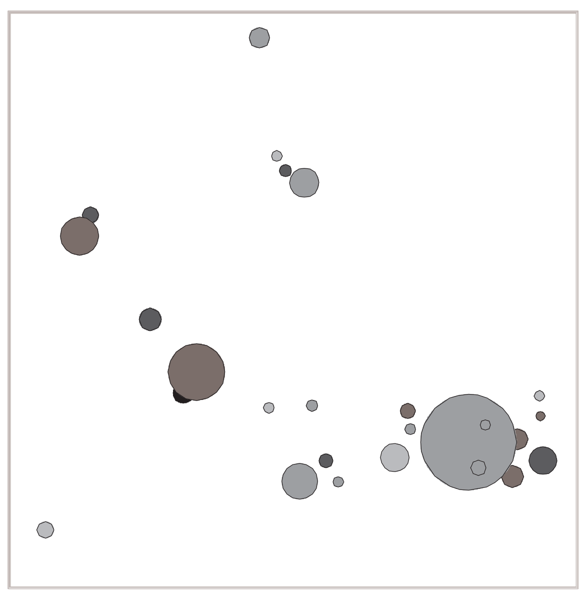


**Figure 3.3d: PCV.aml**
Interval/Ratio data displayed with graduated circles that are value shaded on the basis on an ordinal data set.

orientations of the same shape seem to 'go together' more than the same orientation of varying shapes). PSO.aml, uses lookup tables for both shape and orientation to generate a bivariate display; see Figure 3.3a. Note that this AML (as with all that follow) require that any necessary lookup tables be set up in the format generated by SETMAN.aml and JENKS.aml prior to running this routine.

PSO <shape_lookup> <orientation_lookup> {markerset} {markersize}
<shape_lookup> - specifies marker symbol numbers;
<orientation_lookup> - specifies angles in degrees;
{markerset} - a markerset for shapes--defaults to the current markerset;
{markersize} - a size for the markers--defaults to 0.15 inches.

### Nominal Data, and Ordinal Data--Shape and Value

As with the next AML, this routine uses nominal primary data and ordinal secondary data, but the visual hierarchy established by value makes it more appropriate for display of meta-data than size. With value, uncertain values of the primary data can be faded into the background; see Figure 3.3b and PSV.aml.

PSV <shape_lookup> <value_lookup> {markerset} {markersize} {hue} {intensity}
<shape_lookup> - specifies marker symbol numbers;
<value_lookup> - specifies HLS lightness data (0 to 100);
{markerset} - a markerset for shapes--defaults to the current markerset;
{markersize} - a size for the markers--defaults to 0.15 inches;
{hue} {intensity} - specify a color that will be value shaded.

### Nominal Data, and Ratio Data--Shape and Size

For nominal primary data, shape can be used in conjunction with size to create bivariate point symbols. This seems to work better for two data sets rather than data and meta-data, which should be shown with shape and value; see Figure 3.3c and PSZ.aml (and the size discussion on page 39).

PSZ <shape_lookup> <size_item> {markerset} {size_exponent} {size_factor} {minimum_size}
<shape_lookup> - specifies marker symbol numbers;
<size_item> - an interval or ratio data item in the coverage named by <shape_lookup>;
{markerset} - a markerset for shapes--defaults to the current markerset;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

### Ratio Data, and Ordinal Data--Size and Value

Symbol size and value are good for either ordinal or interval/ratio data. The AML PCV.aml combines the two; see Figure 3.3d. Because symbol size is calculated directly from the Point Attribute Table, only one lookup table is needed, but rather than containing point symbol numbers, it should contain color value (0--100) numbers. This system can be used to represent a ratio data value with size, and a meta-data estimate of accuracy with value.

PCV <size_item> <value_lookup> {minimum_size} {maximum_size} {hue} {intensity}
<size_item> - a numeric item in the coverage referenced by <value_lookup>;
<value_lookup> - specifies HLS lightness data;
{minimum_size} - graduated circle size for the smallest data value--default = 0.05 inches;
{maximum_size} - graduated circle size for the largest data value--default = 0.5 inches;
{hue} {intensity} - specify a color that will be value shaded.

## Bivariate, Color Symbolization

For color (hue or hue/intensity) based bivariate mapping, a point symbol should be selected from the available marker sets (PLOTTER, COLOR, MINERAL, MUNICIPAL, OILGAS, TEMPLATE, USGS, or WATER); these markersets are displayed in Appendix B of the Map Display and Query guide. The color specification of that point symbol will then be changed to show variations in data and meta-data. For size and hue based maps, only a sets of hues must be chosen; symbol size and shape is calculated by ARC/INFO.

### Two Nominal Data Sets--Shape and Hue

This macro (PSH.aml) uses two lookup tables for symbolizing two nominal data sets. Unlike shape and orientation, neither shape or hue is, in general, the dominant visual variable. Visual hierarchies can be established by selecting intense color and similar shapes--this will make color hue the more prominent of the two nominal visual variables. See Figure 3.4a.

PSH <shape_lookup> <hue_lookup> {markerset} {markersize} {value} {saturation}
<shape_lookup> - specifies marker symbol numbers;
<hue_lookup> - specifies HLS hue data (values of 0 to 360);
{markerset} - a markerset for shapes--defaults to the current markerset;
{markersize} - a size for the markers--defaults to 0.15 inches:
{value} {saturation} --defaults of 50 and 100 (maximum intensity).

### Two Nominal Data Sets--Dual Hue Ranges

Color hue is best used for nominal data, although with well selected colors, hue can be used with ordinal data. Use of the spectral encoding bivariate mapping scheme requires careful selection of colors. For the specification of colors for dual-hue range mapping, the CMY (Cyan, Yellow, Magenta) color scheme is most useful. A color chart for this color specification system is in *Appendix J* of the *Map Display and Query* manual. For a four by four color matrix, color specifications like the following can be used:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cyan: | 100 | Cyan: | 100 | Cyan: | 100 | Cyan: | 100 |
| Magenta: | 0 | Magenta: | 33 | Magenta: | 67 | Magenta: | 100 |
| Cyan: | 67 | Cyan: | 67 | Cyan: | 67 | Cyan: | 67 |
| Magenta: | 0 | Magenta: | 33 | Magenta: | 67 | Magenta: | 100 |
| Cyan: | 33 | Cyan: | 33 | Cyan: | 33 | Cyan: | 33 |
| Magenta: | 0 | Magenta: | 33 | Magenta: | 67 | Magenta: | 100 |
| Cyan: | 0 | Cyan: | 0 | Cyan: | 0 | Cyan: | 0 |
| Magenta: | 0 | Magenta: | 33 | Magenta: | 67 | Magenta: | 100 |

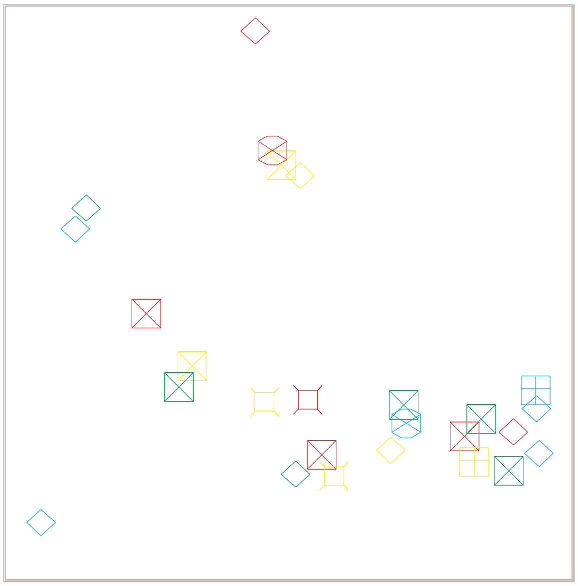


**Figure 3.4a: PSH.aml**
Two nominal data sets displayed with shape and hue.

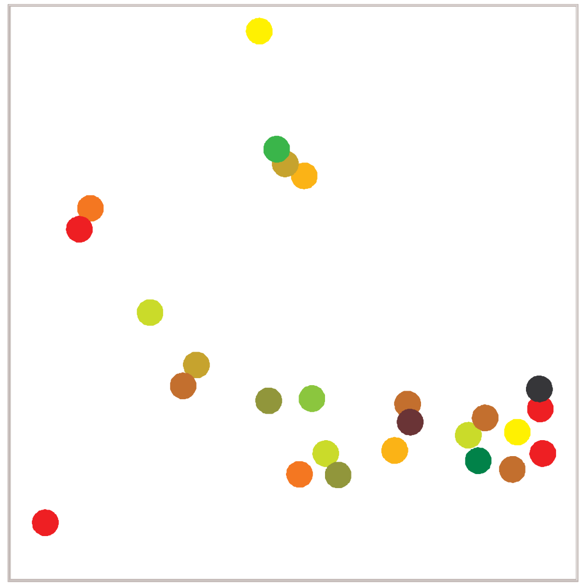


**Figure 3.4b: PDH.aml**
Two nominal data sets displayed with the dual-hue ranges. Note the lack of discernible order in color changes--this necessitates a legend

**Figure 3.4d: PBL.aml**
The AML that generates these legends automatically places the labelling text and the total number of occurrences in each column or row.

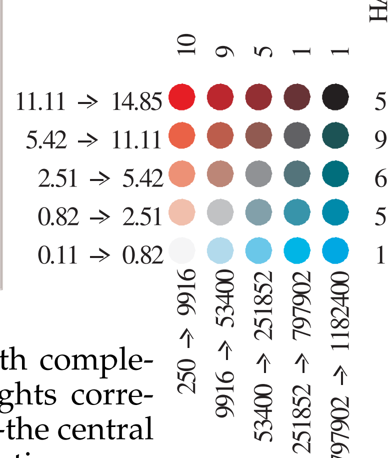




**Figure 3.4c: PCC.aml**
Two ordinal data sets displayed with complementary colors. This technique highlights correlation better than dual-hue range maps--the central gray diagonal could indicate a linear relation.

This pattern can be generated with SETAUTO.aml with the command line 0 33. Yellow should be held constant for each of the sixteen positions--generally, 100 should be good. Even with this control of changes in color, dual-hue range maps should generally be used only for nominal data--the complementary-color system tends to present ordinal data better. See Figure 3.4b and PDH.aml.

PDH <yellow-green_lookup> <yellow-red_lookup> {markerset} {markersymbol} {markersize}
<yellow-green_lookup> - specifies changes in cyan (values of 0 to 100);
<yellow-red_lookup> - specifies changes in magenta (values of 0 to 100);
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - silicifies the size of the marker--defaults to 0.15 inches.

### Two Ordinal Data Sets--Complementary Colors

This AML is a variation on the previous macro. The change is in the colors used to symbolize data; complementary colors are hues that are on opposite sides of the Munsell or Tektronix color spaces and mix to form grey. This mixing allows highlighting of data that is not highly correlated, because these areas will appear in color, whereas the central axis of correlated data will appear in grey. This method allows the representation of both positive and negative correlations; negative correlations should be represented by reversing the values in one of the lookup tables--this changes the direction of the slope of the central axis. Note that white should be avoided because the entire symbol will disappear on a white sheet of paper; use a range from 5 to 100 for percent area inked. See Figure 3.4c and PCC.aml.

PCC <cyan_lookup> <red_lookup> {markerset} {markersymbol} {markersize}
<cyan_lookup> - specifies changes in cyan (values from 0 to 100);
<red_lookup> - specifies changes in red (values from 0 to 100);
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - specifies the size of the marker--defaults to 0.15 inches.

### Point Legend Creation

Although the usage of this AML (PBL.aml) is lengthy, it allows one macro to generate a bivariate legend for three different types of point symbolization schemes: dual hue, complementary colors, and hue and intensity. See Figure 3.4d for both dual-hue and complementary-color legends and Figure 3.5b for a hue and intensity legend.

PBL <data1_lookup> <data2_lookup> <start_x> <start_y> <step_x> <step_y> <d|c|h>
{markerset} {markersymbol} {markersize} {textset} {font} {point} {decimal_precision}
<data1_lookup> - the first lookup given in one of the bivariate AMLs;
<data2_lookup> - the second lookup given in one of the bivariate AMLs;
<start_x> <start_y> - the lower left corner of the legend matrix, in PAGEUNITS;
<step_x> <step_y> - the separation of symbols on the x and y axes, in PAGEUNITS;
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - specifies the size of the marker--defaults to 0.15 inches;
{textset} {font} {point} - specify a textset for legend labels--defaults to a roman, 10 point;
{decimal_precision} - number of decimal places shown for ratio data labels--defaults to 2.

### Nominal Data, and Ordinal Data--Hue and Intensity

Color hue is best used for nominal data; color intensity, on the other hand, is best used for ordinal data (and generally only for meta-data, not a second data

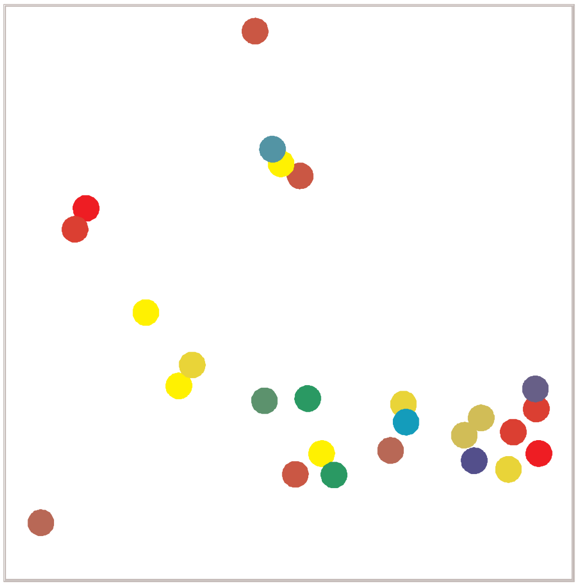


**Figure 3.5a: PHI.aml**

A nominal data set displayed with hue and an ordinal data set used to display meta-data. Intense (bright) colors tend to be more noticeable and are used to present more certain values.

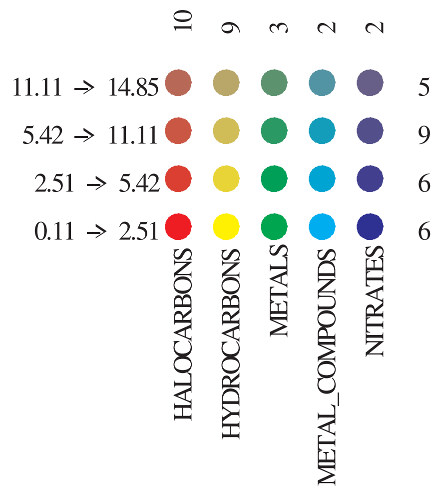



**Figure 3.5b: PBL.aml**

A hue and intensity legend. Note the lack of a diagonal, as in the complementary color system. This makes hue and intensity better suited to display of meta-data than two correlated data sets.

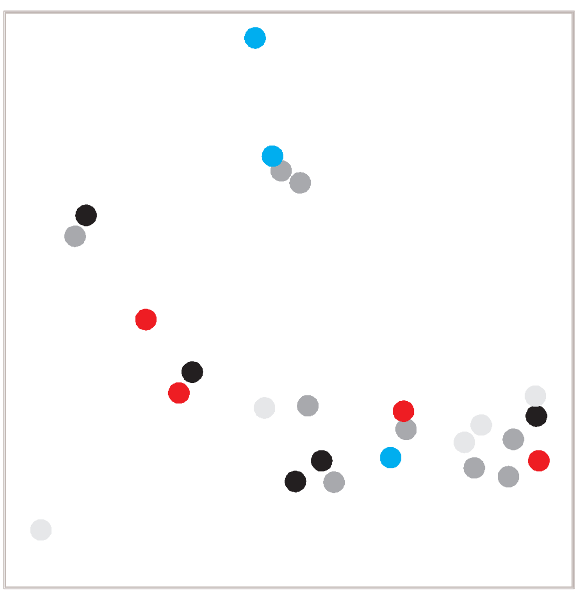


**Figure 3.5c: PEE.aml**

Two ratio data sets displayed with the Eyton's equiprobability ellipse system. This should be used to highlight correlated data.

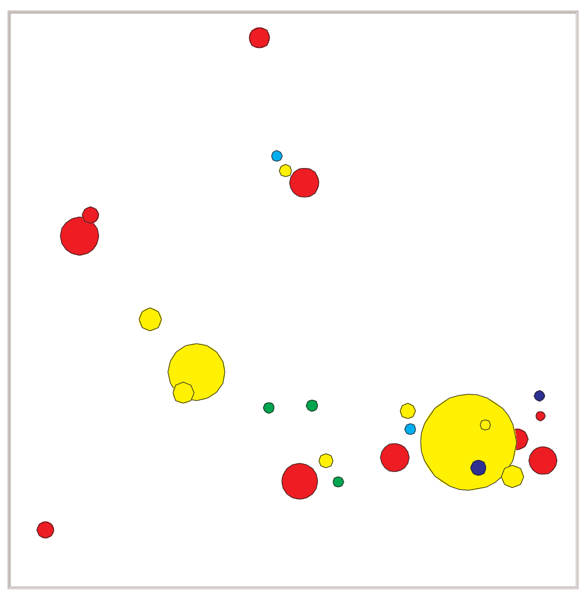


**Figure 3.5d: PCH.aml**

Interval/Ratio data displayed with graduated circles that are hue shaded, displaying nominal data.

variable). This color scheme represents meta-data better than the dual-hue and complementary-color bivariate systems, because the data variable is clearly displayed in a constant hue, unlike the other color bivariate methods. See Figure 3.5a and PHI.aml.

PHI <hue_lookup> <intensity_lookup> {markerset} {markersymbol} {markersize}
<hue_lookup> - specifies changes in HLS hue (from 0 to 360)
<intensity_lookup> - specifies changes in HLS saturation (from 0 to 100)
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - specifies the size of the marker--defaults to 0.15 inches.

### Two Ratio Data Sets--Equiprobability Ellipse

Eyton's ellipse is a variation on the complementary color system. The colors that are used are the same, but the linear correlation between the two variables is used to determine a central category, which specifically highlights correlation. This AML requires that the EYTON.aml, presented in chapter two, be run first. See Figures 3.5c and 5.7b on page 79 for a legend display, and PEE.aml.

PEE <ellipse_lookup> {markerset} {markersymbol} {markersize}
<ellipse_lookup> - a lookup table generated by EYTON.aml;
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - specifies the size of the marker--defaults to 0.15 inches.

### Ratio Data, and Nominal Data--Size and Hue

This macro is the color equivalent of the one that generated Figure 3.3d. Unlike that AML though, this should be used for one nominal data variable and one ratio data variable. See Figure 3.5d and PCH.aml.

PCH <size_item> <hue_lookup> {minimum_size} {maximum_size} {value} {intensity}
<size_item> - a numeric data item of the coverage reference by <hue_lookup>;
<hue_lookup> - specifies changes in HLS hue (from 0 to 360);
{minimum_size} - graduated circle size for the smallest data value--default = 0.05 inches;
{maximum_size} - graduated circle size for the largest data value--default = 0.5 inches;
{value} {intensity} --default to 50 and 100 (maximum intensity).

## Multivariate Symbolization

Multivariate point symbolization can be accomplished by several means, each of which is suited to various combinations of data levels. Because of the increased complexity involved in multivariate mapping, care must be taken to insure legends are well designed and convey the methods that should be used in interpreting map symbols.

### Three Ordinal to Ratio Data Sets--Red, Green and Blue Symbolization

The 'false color' images that are often generated with satellite derived data use red, green and blue to symbolize data values from three spectral bands. This technique is applied here to allow display of three data values. See Figures 3.6a and 5.8b on page 82 for a legend. Note that the AML (PRGB.aml) performs a linear-stretch on the items in the Point Attribute Table. This AML also allows color specification as cyan, magenta, and yellow. This can be helpful because this color scheme tends to allow numbers in the low end of the data range to be distinguished more readily than the RGB scheme.

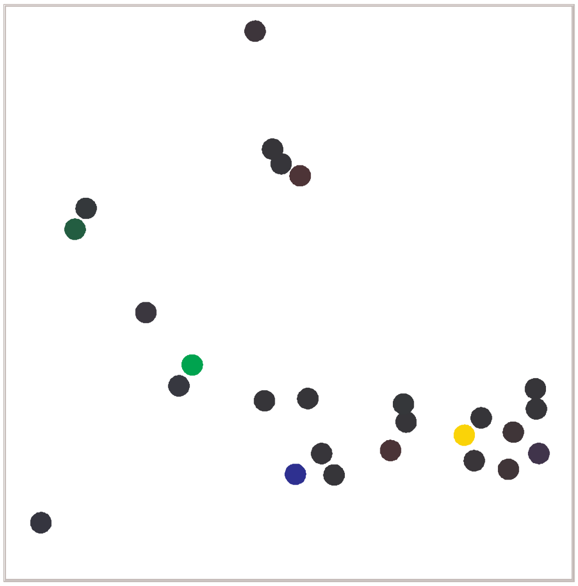


**Figure 3.6a: PRGB.aml**
Three ratio data sets displayed by linearly stretching the data sets from 0 to 255 and using data set one for red, two for green and three for blue.

**Figure 3.6b: PSZV.aml**
A nominal data set displayed with shape, a ratio data set displayed with size, and an ordinal data set (size meta-data) displayed with color value.

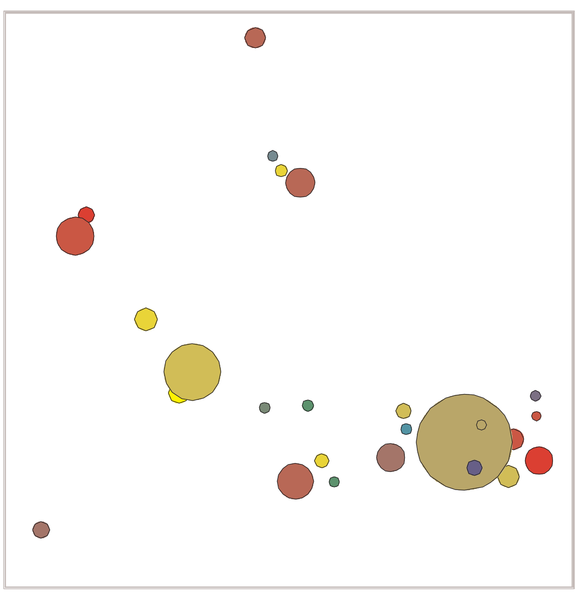


**Figure 3.6c: PCHI.aml**
A nominal data set displayed with color hue, ratio data displayed with size, and an ordinal data set (size meta-data) displayed with color intensity.

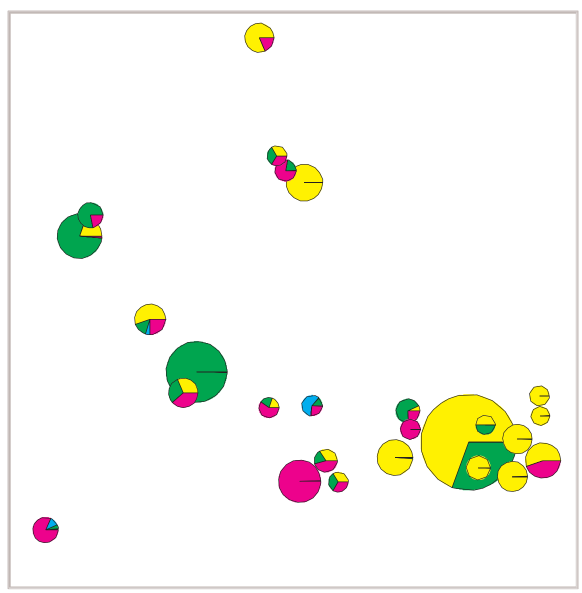


**Figure 3.6d: PP.aml**
Ratio data displayed with size; each pie slice is a nominal difference within the ratio whole--pie slice values are percentages of the whole.

PRGB <cover> <red_item> <green_item> <blue_item>
{markerset} {markersymbol} {markersize} {point|node} {r|c}
<cover> - the coverage to be displayed;
<red_item> <green_item> <blue_item> - numeric items in the attribute table of <cover>;
{markerset} {markersymbol} - specify which symbol will be drawn with--defaults to current;
{markersize} - specifies the size of the marker--defaults to 0.15 inches;
{point|node} - symbolize point (the default) or node features;
{r|c} - display with the RGB (default) or CMY color system.

### Nominal Data, Ratio Data and Ordinal Data--Shape, Size and Value

Shape can be used to display nominal data, and size can be used to display ratio data. This AML (PSZV.aml) adds to this bivariate representation by allowing color value to be used to display an ordinal data set. Color value can be used to display uncertainty on monochrome output devices, such as laser printers. The calculation of symbol size is discussed on page 39; see Figure 3.6b.

PSZV <shape_lookup> <size_item> <value_lookup> {markerset}
{size_exponent} {size_factor} {minimum_size} {hue} {intensity}
<shape_lookup> - specifies marker symbol numbers;
<size_item> - a numeric data item in the coverage referred to by <shape_lookup>;
<value_lookup> - specifies HLS lightness data (from 0 to 100);
{markerset} - a markerset for shapes--defaults to the current markerset;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches;
{hue} {intensity} - define a color that will be value shaded.

### Ratio Data, Nominal Data, and Ordinal Data--Size, Hue and Intensity

Like the previous AML, size, hue and intensity should be used to display a nominal data variable, a ratio data variable, and an ordinal data variable (meta-data for the ratio data variable would be appropriate). This AML however uses color hue and intensity, and therefore requires full color displays. See Figure 3.6c and PHIZ.aml.

PHIZ <hue_lookup> <intensity_lookup> <size_item>
{markerset} {markersymbol} {size_exponent} {size_factor} {minimum_size}
<hue_lookup> - specifies HLS hue data (from 0 to 360);
<intensity_lookup> - specifies HLS saturation data (from 0 to 100);
<size_item> - a numeric data item in the coverage referred to by the lookup tables:
{markerset} {markersymbol} - specify a marker--the current marker is the default;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

### Ratio Data--Point Pie Graphs

This AML generates pie symbols for point data by use of the POINTSPOT command. The type of data that should be used in the creation of this type of map is a size value that represents a sum of several other values; these other values should have nominal distinctions. POINTSPOT uses the sum value to calculate the size of the circle, and it calculates the pie slice size that will be drawn as a function of the ratio a sub-value to the whole. Colors for each slice are calculated by the AML. See Figures 3.6d and 6.3d on page 88 for a legend, and PP.aml.

PP <cover> <point|node> <size_item> <minimum_size> <maximum_size>
<number_of_subitems> <item_1...item_n>

<cover> - the coverage to be displayed;
<point|node> - symbolize points or nodes;
<size_item> - a numeric item in <cover> that specifies the size of the circle;
<minimum_size> <maximum_size> - the smallest and largest circle sizes;
<number_of_subitems> - the number of pie slice data items--the value of n, next;
<item_1...item_n> - names of numeric items in <cover> that specify the size of pie slices.

# Chapter Four
# Line Symbolization in ARC/INFO

For line symbolization, the visual variables of hue, shape, and texture can be controlled by the use of lookup tables, which relate nominal difference to symbolization. Value and size can be used either by means of lookup tables, or by referencing Arc Attribute Table values. Lookup tables should be generated with one of the lookup table AMLs presented in Chapter Two.

## Monovariate Symbolization

As with point symbolization, the AMLs discussed in this chapter start out relatively simply, and the output could be accomplished almost as readily by hand. Like the monovariate point AMLs, these AMLs are primarily background for the bi- and multivariate AMLs of later sections.

ARC/INFO provides a default set of lines in the lineset file, PLOTTER.LIN; there are several other available linesets: 50.LIN, BW.LIN, CALCOMP2.LIN, CARTO.LIN, COLOR.LIN, HP.LIN, and OILGAS.LIN. These are displayed in the *Map Display and Query* guide's *Appendix A*, and the ability to create new symbols sets is discussed in chapter three of *Map Display and Query*.

### Nominal Data--Hue

Hue is best used for nominal data, and the COLOR.LIN lineset provides fifteen symbols of the same shape, but each with a different color. All of the other linesets, except BW.LIN, have the same symbol in black, red, green and blue. This AML (LH.aml) uses the COLOR.LIN lineset for generating a display for arcs--see the next section for display of routes, and Figure 4.1a.

```
LH <cover> <data_item> <lookup_table> <lineset>
        <cover> - the arc coverage to be displayed;
        <data_item> - the data item in <cover> referenced by <lookup_table>;
        <lookup_table> - an info table that relates <data_item> to <lineset>;
        <lineset> - a lineset.
```

### Nominal Data--Shape or Texture

Shape should be used for nominal data, and a line's shape can be changed in several ways in ARC/INFO. The linesets PLOTTER.LIN, TEMPLATE.LIN and OILGAS.LIN have symbols that change in shape. LINESET can be used to select one of these, and LINESYMBOL can be used to change the shape of the current line type, as well as change the line color and width (although the selection is limited). This allows the selecting of several different shapes; the LINETYPE command allows the generation of nine other types of shapes.

Texture is best used for nominal data, and the CARTO.LIN lineset provides many ready-to-use symbols that vary in texture (such as lines 106, 110, 114 and 118). These symbols can be selected by using LINESET to select this symbol set, and LINESYMBOL to select the individual symbols. A line's texture can also be directly varied; LINEINTERVAL and LINETEMPLATE can be used to control texture. LINEINTERVAL determines the space between successive parts of a line symbol; it defaults to 0 (no space). LINETEMPLATE requires that a lineinterval be

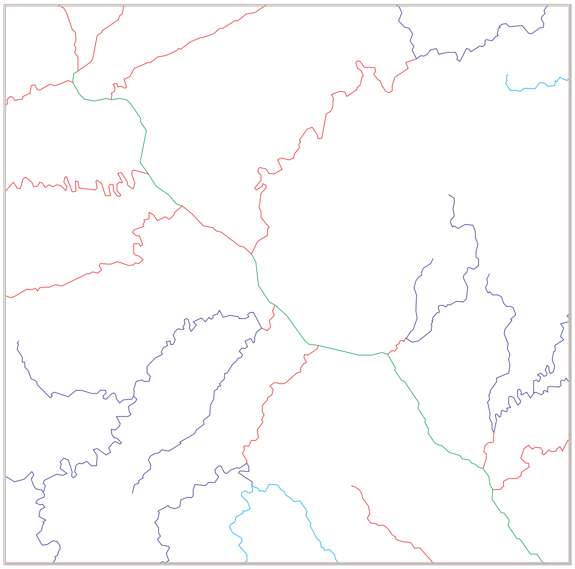


**Figure 4.1a: LH.aml**
Nominal data in an arc coverage distinguished by hue with the color.lin lineset.

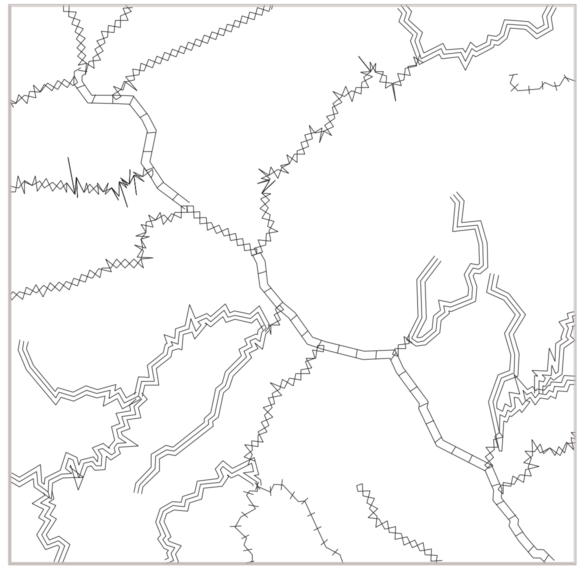


**Figure 4.1b: LS.aml**
A nominal data set in a route system distinguished by shape using symbols from plotter.lin.

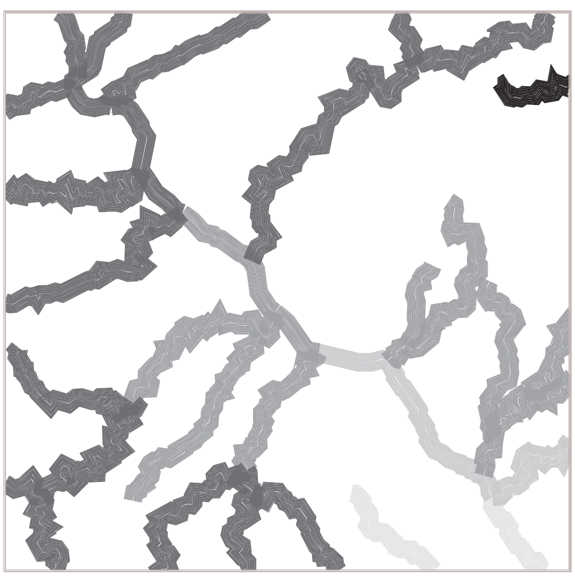


**Figure 4.1c: LV.aml**
Classed ratio (ordinal) data displayed with color value by using a lookup table.

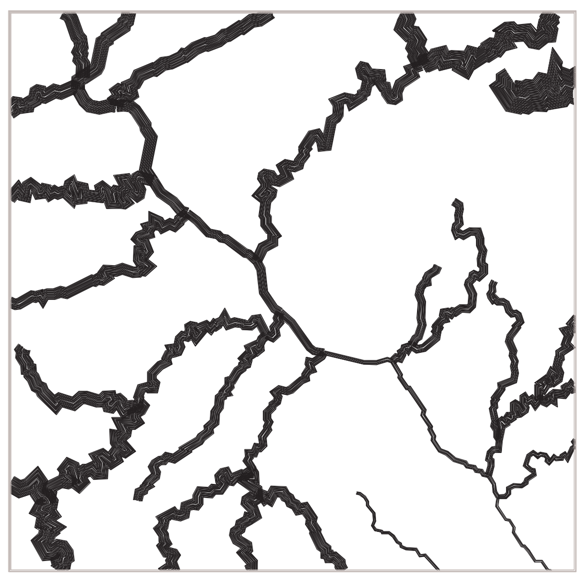


**Figure 4.1d: LZ.aml**
Ratio data displayed with size; sizes are calculated by scaling the data in the coverage Arc Attribute Table.

set. The LINETEMPLATE command then allows control over both the length of the mark and the length of the space between marks. The <template> digits are used to scale the lineinterval, with the first digit used for a line length, the second for a space size, the third for a line, etc. For example, if the lineinterval is set to 0.25 inches and linetemplate is set to 2122, a line pattern of: a line 0.5 inches long, followed by a space 0.25 inches long, followed by a line 0.5 inches long, followed by a 0.5 inches long space is used to draw new lines. LS.aml uses a preset lookup table to select line symbols for routes (see Figure 4.1b).

LS <cover> <route> <data_item> <lookup_table> <lineset>
<cover> - the route coverage to be displayed;
<data_item> - the data item in <cover> referenced by <lookup_table>;
<lookup_table> - an info table that relates <data_item> to <lineset>;
<lineset> - a lineset.

### Ordinal Data--Value

Color value is best used for ordinal data, and the application of lookup tables to a ratio data set, provides this data level. This AML (LV.aml) requires an ordinal or ratio lookup table that specifies HLS color value (0 to 80) in its symbol column. So the AML can control line value, any LINETYPE except hardware can be used. The AML uses LINETYPE WIDE; this allows control of the width of a solid line, which should be wide enough that dithering will not cause the line to appear broken (a line width of at least 0.02 inches on a 300 dpi printer). See Figure 4.1c.

LV <value_lookup> {size} {hue} {intensity}
<value_lookup> - a lookup table generated by SETMAN or JENKS;
{size} - a line width--the default is 0.1 inches:
{hue} {intensity} - specify a color that will be value shaded.

### Ratio Data--Size

Size can be used for ordinal data, and the PLOTTER.LIN, CARTO.LIN and OILGAS.LIN linesets provide symbols that vary in size. An appropriate lineset can be selected with the LINESET command, and used with a lookup table that relates changes in data values to changes in size. For interval/ratio level data, the LINESIZE command allows direct control of a line's width. Prior to using the LINESIZE command however, the LINETYPE command must be given with a type not equal to HARDWARE. A LINETYPE such as WIDE can be used. Once this is done, the line's width can be controlled to represent data. This AML uses the interval/ratio method to control width (see Figure 4.1d and LZ.aml).

LZ <cover> <size_item> {size_exponent} {size_factor} {minimum_size} {linetype}
<cover> - the arc coverage to be displayed;
<size_item> - an interval/ratio data item in the Arc Attribute Table of <cover>;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

All of the size symbolization AMLs use the formula:

$$\text{size} = \text{size\_factor}\,(\text{normalized\_data\_value}^{\text{size\_exponent}}) + \text{minimum\_size}$$

This is just a formula for a line in x (normalized_data) and y (symbol size) space (with size_exponent set to one), with the ability to create non-linear progressions (with size_exponent not equal to one). The slope of the line is determined by size_factor, and the y-intercept of the line is minimum_size. The data is normalized prior to calculating the size in order to set the minimum data value to minimum_size; data values are normalized by:

normalized_data_value = (actual - minimum) / (maximum - minimum)

## Bivariate, Monochrome Symbolization

The visual variables of shape, texture, value and size can be combined to allow monochrome bivariate mapping of linear features. This can be used to present two data sets, or data and meta-data.

### Two Nominal Data Sets--Shape and Texture

This AML (LST.aml) uses two lookup tables to determine symbolization. Because the selection of predefined shape/texture symbols is limited, LINETYPE can be combined with LINETEMPLATE. This is assisted by assigning numeric codes in the shape lookup table:

1000 = WIDE
1001 = DIAMOND
1002 = DOTS
1003 = HASH
1004 = SCALLOP
1005 = SCRUB
1006 = SLANT
1007 = ZIGZAG

These codes do not interfere with the Agfa Compugraphic glyph numbers, which this AML will use for LINETYPE MARKER. The second (texture) lookup table requires an integer value that specifies texture (see the texture discussion in Nominal Data--Shape or Texture, above). See Figure 4.2a.

LST <shape_lookup> <texture_lookup> {interval} {size}
<shape_lookup> - specifies either Agfa Compugraphic glyphs or the above codes;
<texture_lookup> - specifies an integer code for the LINETEMPLATE;
{interval} - a ratio size in PAGEUNITS--the default is 0.1 inches:
{size} - specifies a line size--the default is 0.1 inches.

### Nominal Data, and Ordinal Data--Shape and Value

Shape should be used to specify nominal differences and value should be used to specify ordinal differences. Generally, value should be used for meta-data and size (the next AML) should be used for a second data set. This AML (LSV.aml) requires a lookup table that relates nominal distinctions to line symbols in a lineset, and a second lookup table that specifies HLS values. See Figure 4.2b.

LSV <shape_lookup> <value_lookup> {lineset} {linesize} {hue} {intensity}
<shape_lookup> - specifies line symbol numbers;
<value_lookup> - specifies HLS lightness data (from 0 to 80);
{lineset} - specifies a lineset for shapes--defaults to the current lineset;
{linesize} - specifies a line width--defaults to 0.1 inches;
{hue} {intensity} - specify a color that will be value shaded.

**Figure 4.2a: LST.aml**
Two nominal data sets displayed with shape (primary data) and texture.

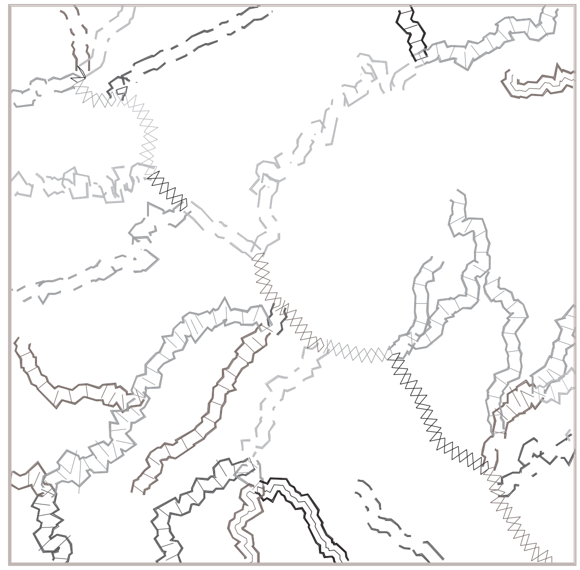


**Figure 4.2b: LSV.aml**
A nominal data set displayed with shape and an ordinal data set displayed with value. This is more effective with meta-data than a second data set.

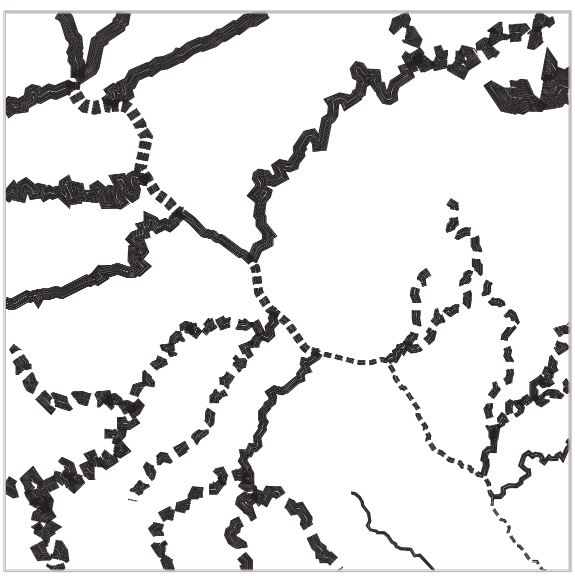


**Figure 4.2c: LSZ.aml**
A nominal data set displayed with texture and a ratio data set displayed with size. Size is more effective than value in conveying a second data set.

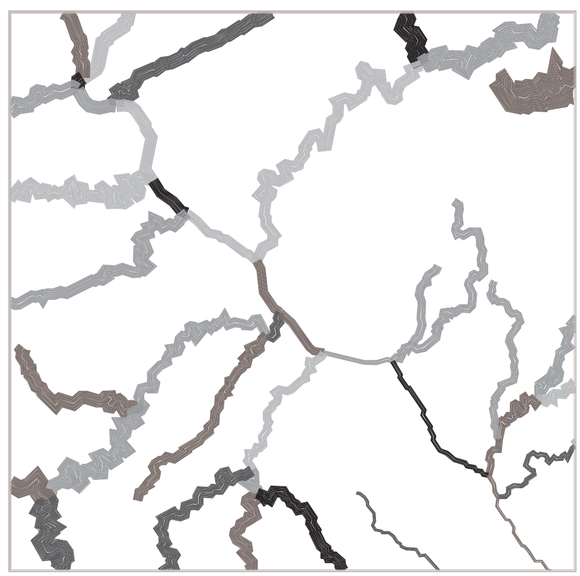


**Figure 4.2d: LVZ.aml**
Ratio data displayed with size, which is value shaded on the basis of meta-data.

### Nominal Data, and Ratio Data--Shape and Size

The AML, LSZ.aml, should be used for two data sets--one of nominal information (shape or texture) and one of ordinal to ratio information (size). Like the shape and value AML, this AML requires a lookup table that relates nominal categories to line symbols, but requires the specification of a ratio data item in the line coverage's attribute table. See the discussion of size specification on page 50 and Figure 4.2d.

LSZ <shape_lookup> <size_item> {lineset} {scale_factor} {size_factor} {minimum_size}
<shape_lookup> - specifies line symbol numbers;
<size_item> - an ratio data item in the attribute table referenced by <shape_lookup>;
{lineset} - specifies a lineset for shapes--defaults to the current lineset;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

### Ratio Data, and Ordinal Data--Size and Value

This AML uses size to indicate a ratio data set, and value to represent a second data set. Because of the establishment of visual levels by both size and value, this combination is good for conveying data and meta-data. The AML requires a numeric attribute item for size, and a lookup table for HLS value specification. See Figure 4.2d and LVZ.aml.

LVZ <value_lookup> <size_item> {size_exponent} {size_factor} {minimum_size} {hue} {intensity}
<value_lookup> - specifies HLS lightness data (from 0 to 80);
<size_item> - a ratio data item in the attribute table referenced by <value_lookup>;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches;
{hue} {intensity} - specify a color that will be value shaded.

## Bivariate, Color Symbolization

Color hue and intensity add two important visual variables to the representational techniques possible in ARC/INFO. Hue allows better specification of nominal variables than texture (particularly in monochrome, bivariate mapping) and can also be used for ordinal data. Intensity is useful for display of uncertainty meta-data.

### Two Nominal Data Sets--Texture or Shape, and Hue

This AML (LSH.aml) takes advantage of the ability of shape to convey nominal data, and it uses color hue to specify a second nominal category. The AML requires a lookup table that specifies line symbols for the shape (or texture) variable, and a second AML that specifies HLS hue values (0 to 300). See Figure 4.3a.

LSH <shape_lookup> <hue_lookup> {lineset} {linesize} {value} {intensity}
<shape_lookup> - specifies line symbol numbers;
<hue_lookup> - specifies HLS hue data;
{lineset} - select a lineset for shapes--defaults to the current lineset;
{linesize} - specifies a line width--defaults to 0.1 inches;
{value} {intensity} - defaults: 50, 100 (maximum brightness).

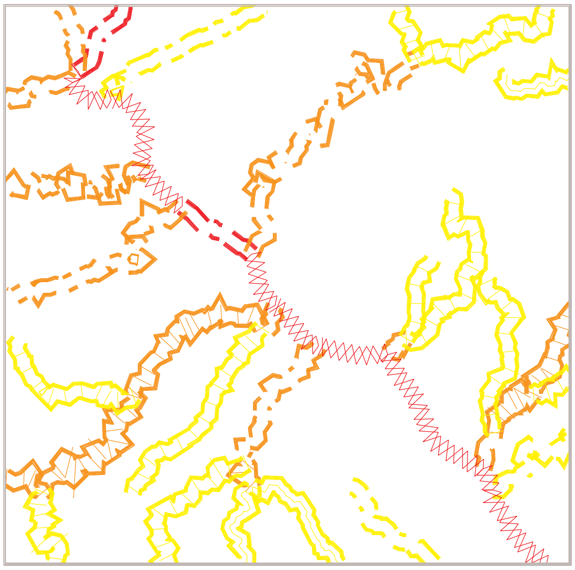


**Figure 4.3a: LSH.aml**
Two nominal data sets displayed with shape and hue.

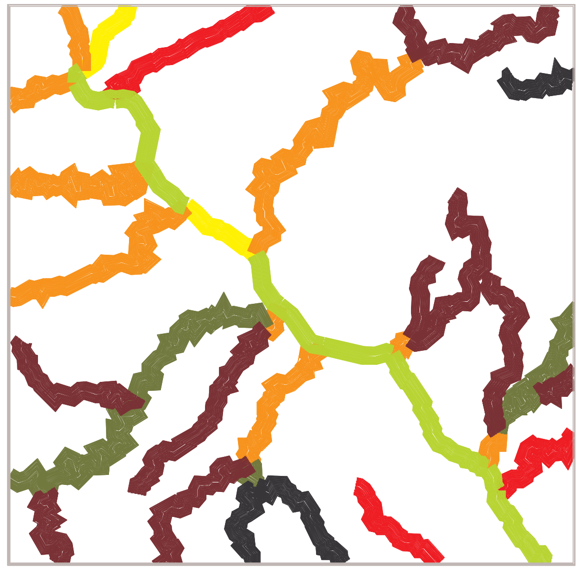


**Figure 4.3b: LDH.aml**
Two nominal data sets displayed with the dual-hue range technique. Note the lack of discernible order in color changes--this necessitates a legend (below).

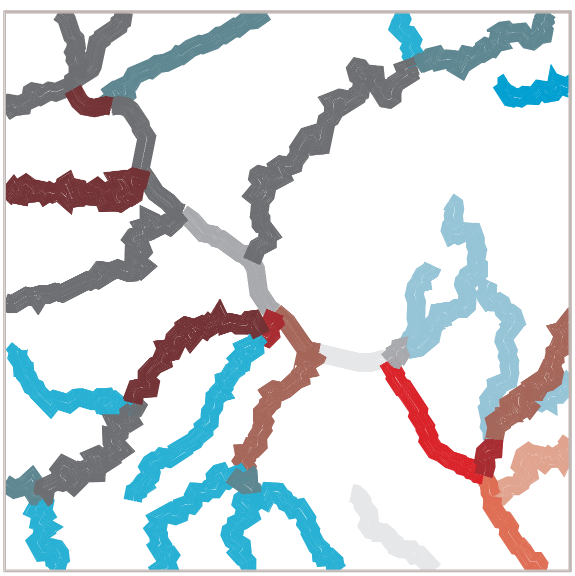


**Figure 4.3c: LCC.aml**
Two ordinal data sets displayed with the complementary color technique. This method highlights correlation better than dual-hue mapping.

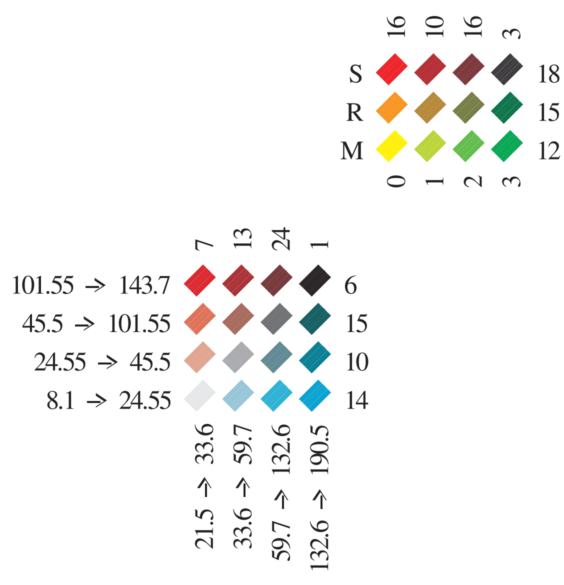



**Figure 4.3d: LBL.aml**
Dual-hue (top) and complementary color legends. Note the central gray diagonal in the complementary color legend.

### Two Nominal Data Sets--Dual Hue Ranges

Another method of displaying two nominal data sets is the dual-hue range technique. This uses two hue ranges (yellow to red and yellow to green) to specify data values. The combination of the two color ranges results in unique colors for each combination of data values. The AML requires two lookup tables, which specify color for the CMY color model (ranges of 0 to 100). See Figure 4.3b and LDH.aml.

LDH <yellow-green_lookup> <yellow-red_lookup> {linesize}
<yellow-green_lookup> - specifies changes in cyan (from 0 to 100);
<yellow-red_lookup> - specifies changes in magenta (from 0 to 100);
{linesize} - specifies a line width--defaults to 0.1 inches.

### Two Ordinal Data Sets--Complementary Colors

The use of color to specify ordinal data requires careful selection of colors. By use of color intensity in two complementary colors (equal amounts of each yield gray), it is possible to specify a color system that represents ordinal information well. This AML (LCC.aml) accomplishes this with two lookup tables for the CMY color system (color numbers of 10 to 100), one of which is used for cyan and the other is used for magenta and yellow, together (this yields red). A central diagonal results, which is gray. With color values based directly on data values, this gray line would represent linear correlation between the two variables. See Figure 4.3c.

LCC <cyan_lookup> <red_lookup> {linesize}
<cyan_lookup> - specifies changes in cyan;
<red_lookup> - specifies changes in red;
{linesize} - specifies a line width--defaults to 0.1 inches.

### Bivariate Color Legends

Legends should present the map reader with as much information as possible. This AML not only generates a line symbol for each color bivariate scheme, but also labels the rows and columns with the appropriate name or data range and provides the total number of items in each row or column. See Figure 4.3d, Figure 4.4d and LBL.aml.

LBL <lookup_1> <lookup_2> <start_x> <start_y> <step_x> <step_y> <d|c|h>
{linesize} {textset} {font} {point} {floating_point_precision}
<lookup_1> - the first lookup specified in LDH, LHI or LCC;
<lookup_2> - the second lookup table from on of the display AMLs;
<start_x> <start_y> - the lower left corner of the legend;
<step_x> <step_y> - the size of each row and column;
<d|c|h> - the symbolization method: Dual hue, Complementary color, or Hue and intensity;
{linesize} - specifies a line width--defaults to 0.1 inches;
{textset} {font} {point} - define the labelling text--defaults to a 10 point roman font;
{floating_point_precision} - the number of decimal places to be shown--the default is 2.

### Two Ratio Data Sets--Eyton's Ellipse for Lines

A development of the complementary color system that specifically highlights correlation is Eyton's equiprobability ellipse. This system calculates the correlation between two ratio variables and specifies a complementary color shade to each data pair. The central category contains the central cluster of correlated

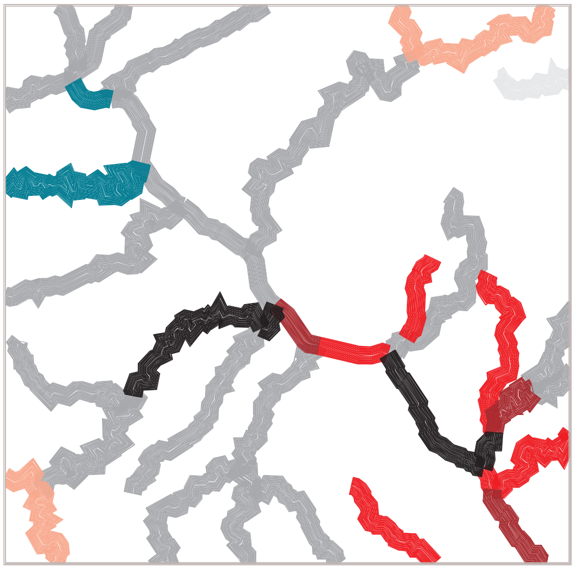


**Figure 4.4a: LEE.aml**
Two ratio data sets displayed with an equiprobability ellipse. This selects the central cluster of correlated data as a fifth category in a 2x2 matrix.

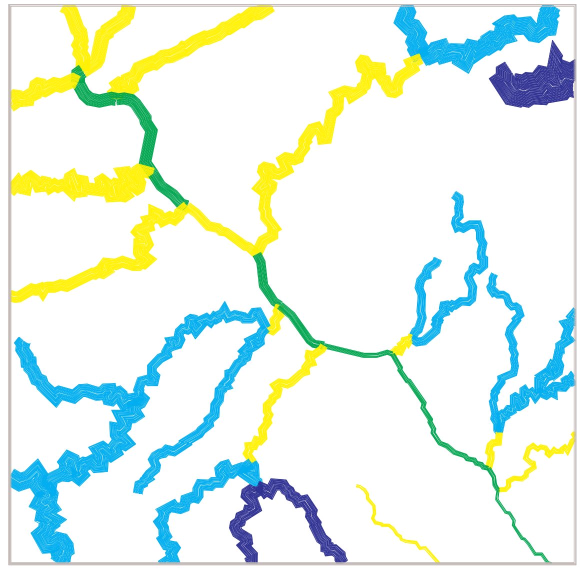


**Figure 4.4b: LHZ.aml**
A nominal data set displayed with hue and a ratio data set displayed with size.

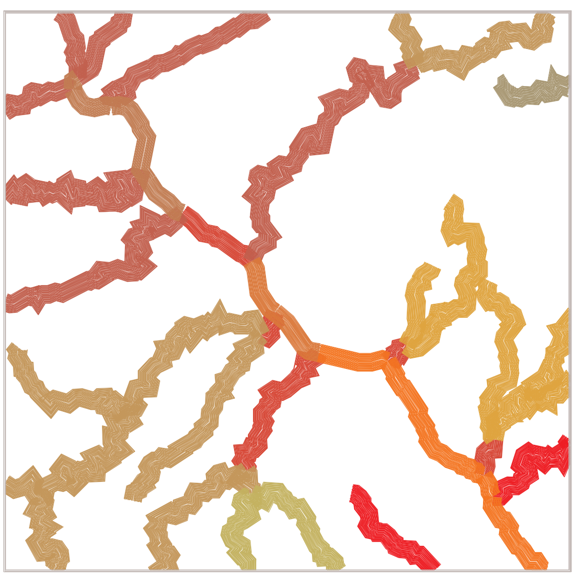


**Figure 4.4c: LHI.aml**
A nominal data set displayed with hue and an ordinal data set displayed with intensity. Intense colors tend to stand out in an image, highlighting data.

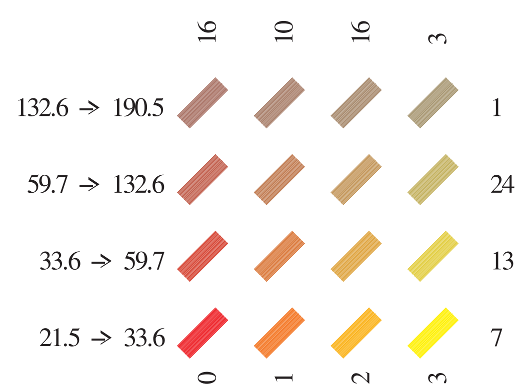



**Figure 4.4d: LBL.aml**
Hue and intensity legend. Note the lack of a diagonal, as in the complementary color system; hue and intensity is better suited to data and meta-data.

data. This system requires that the AML described in chapter two be run to determine the classification. Once this is done, this AML (LEE.aml) displays linear data. See Figure 4.4a.

LEE <ellipse_lookup> {lineset} {linesymbol} {linesize}
<ellipse_lookup> - a lookup table for linear data created by EYTON.aml;
{lineset} {linesymbol} - define a line to use for shading--defaults to the current line;
{linesize} - specifies a line width--defaults to 0.1 inches.

### Ratio Data, and Nominal Data--Size and Hue

For combinations of nominal and ratio data, shape (which can be used in monochrome maps) or hue can be used to specify the nominal data and size can be used to display the ratio information. This AML uses a lookup table that relates nominal data to HLS color hue (0 to 300) and a numeric item in the coverage's attribute table. See Figure 4.4b and LHZ.aml.

LHZ <hue_lookup> <size_item> {size_exponent} {size_factor} {minimum_size} {value} {intensity}
<hue_lookup> - specifies HLS hues;
<size_item> - a numeric data item in the attribute table referred to be <hue_lookup>;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches;
{value} {intensity} - default to 50 and 100--maximum intensity.

### Nominal Data and Ordinal Meta-Data--Hue and Intensity

For display of a nominal data and meta-data, color hue can be used for the nominal information and color intensity can be used for the meta-data. This AML (LHI.aml) requires a lookup table for relating nominal data to HLS hue (0 to 300) and a second table for relating meta-data to HLS saturation (20 to 100). See Figure 4.4c.

LHI <hue_lookup> <intensity_lookup> {linesize}
<hue_lookup> - specifies HLS hue;
<intensity_lookup> - specifies HLS saturation;
{linesize} - specifies a line width--default is 0.1 inches.

## Multivariate Symbolization

Multivariate line symbolization allows the representation of three or more variables about the same line segment on the same map. This is useful for the communication of spatial patterns, but--as with all bi- and multivariate maps--legends should include explanations of the data representation technique.

### Three Ordinal to Ratio Data Sets--Red, Green and Blue Symbolization

For representing three ratio data values, line color can be specified by control of the red, green and blue (or cyan, magenta and yellow) components of the color additive (or subtractive) process. This use of RGB color is similar to that used to display satellite imagery. As such, the AML LRGB.aml requires three numeric data items, which are linearly stretched from 0 to 250 (this prevents white lines on white paper). See Figure 4.5a.

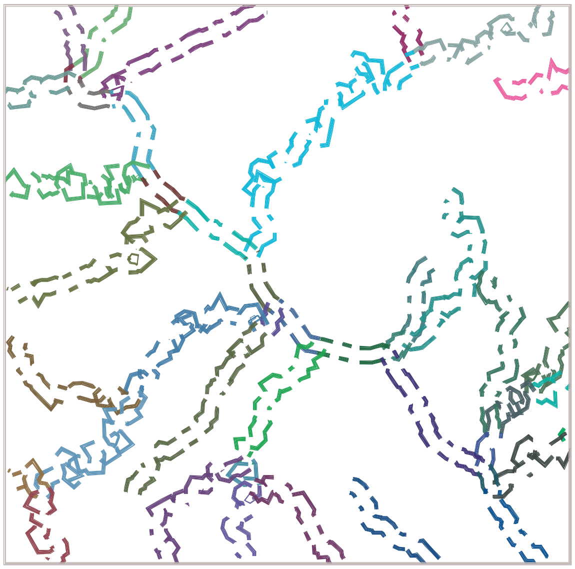


**Figure 4.5a: LRGB.aml**

Three ratio data sets displayed by linearly stretching each data set from 0 to 250 and using data set one for red, data set two for green and data set three for blue.

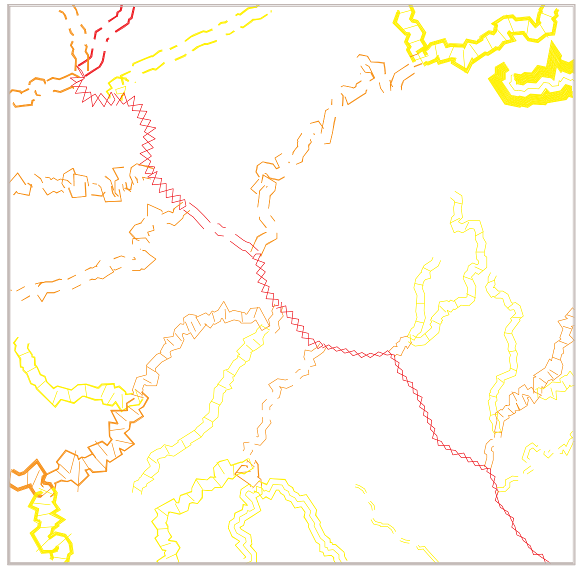


**Figure 4.5b: LSHZ.aml**

A nominal data set displayed with shape, a second nominal data set displayed with color hue, and a ratio data set displayed with size.

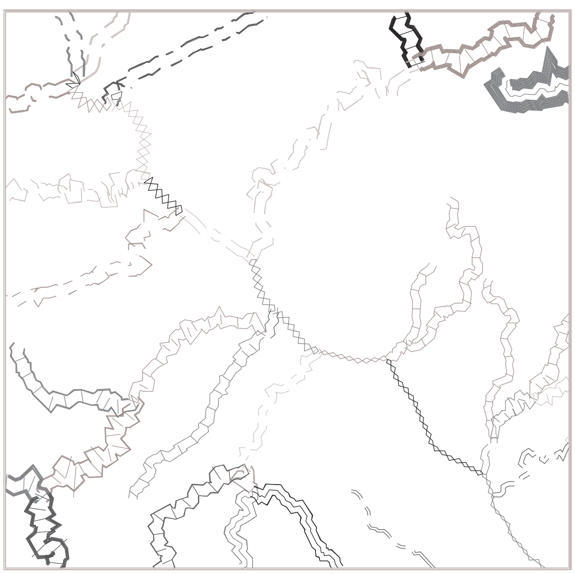


**Figure 4.5c: LSZV.aml**

A nominal data set displayed with shape, a ratio data set displayed with size, and an ordinal data set (possibly ratio meta-data) displayed with value.

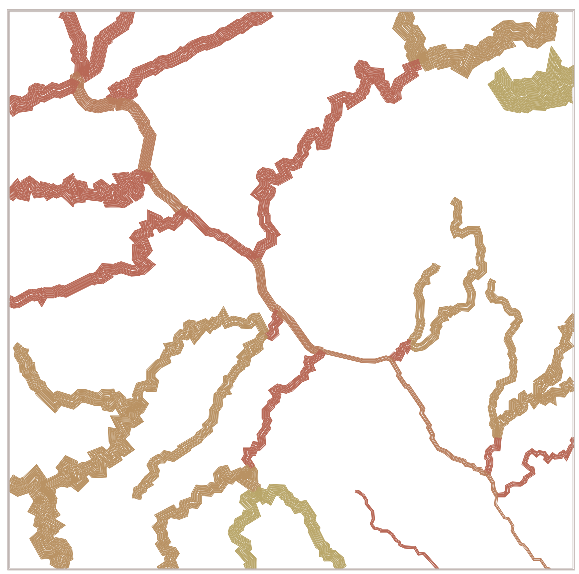


**Figure 4.5d: LHIZ.aml**

A nominal data set displayed with color hue, a ratio data set displayed with size, and an ordinal data set displayed with color intensity.

LRGB <cover> <red_item> <green_item> <blue_item> {linesize} {r|c}
{arc|route|section} {subclass}
<cover> - the coverage to be displayed;
<red_item> <green_item> <blue_item> - three numeric data fields in <cover>;
{linesize} - specifies a line width--default is 0.1 inches;
{r|c} - specifies the RGB or CMY color model;
{arc|route|section} {subclass} - specifies arcs (default) or route/section and subclass.

### Nominal Data, Ratio Data, and Nominal Data--Shape, Size and Hue

By adding a ratio variable to the bivariate representation shown in Figure 4.3a, a trivariate representation results. This allow communication of two nominal variables, and a third, ratio, variable. The AML requires two lookup tables: one that relates nominal data to line symbols; the other relating nominal data to HLS hue (0 to 300). The size information is generated directly from a numeric item in the arc (or route or section) attribute table. See Figure 4.5b, the discussion of size on page 50 and LHSZ.aml.

LSHZ <shape_lookup> <hue_lookup> <size_item> {lineset}
{size_exponent} {size_factor} {minimum_size} {value} {intensity}
<shape_lookup> - specifies line symbol numbers;
<hue_lookup> - specifies HLS hues;
<size_item> - a ratio data field in the attribute table referenced by <shape_lookup>;
{lineset} - specifies a lineset for shapes--defaults to current lineset;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes,
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches;
{value} {intensity} - default to 50 and 100--maximum intensity.

### Nominal Data, Ratio Data, and Ordinal Data--Shape, Size and Value

This AML is addition of the bivariate representation techniques shown in Figure 4.2b, Figure 4.2c and Figure 4.2d. This monochrome, trivariate mapping system allows representation of nominal data and ratio data, along with an ordinal data set, which can be used for representing the ratio data set's meta-data. The AML requires the two lookup tables: one for relating nominal data to line symbols; the other for relating ordinal data to HLS value (0--black to 80--almost white). The size of the line symbol is generated from a numeric data item in an attribute table. See Figure 4.5c and LSZV.aml.

LSZV <shape_lookup> <size_item> <value_lookup> {lineset}
{size_exponent} {size_factor} {minimum_size} {hue} {intensity}
<shape_lookup> - specifies line symbol numbers;
<size_item> - a ratio data field in the attribute table referenced by <shape_lookup>;
<value_lookup> - specifies HLS lightness;
{lineset} - specifies a lineset for shapes--defaults to current lineset;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes;
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches;
{hue} {intensity} - specify a color to be value shaded.

### Ratio Data, Nominal Data, and Ordinal Data--Size, Hue and Intensity

As with the previous AML, this macro is displays a nominal data set (with hue instead of shape) and a ratio data set, along with an ordinal data set, which can be used for representing the ratio data's meta-data. The AML requires two lookup tables: for nominal data, a table of HLS hues (0 to 300); for ordinal data, a table of

HLS saturation (20--off gray to 100--maximum intensity). The line symbol size is calculated from a numeric item in the attribute table specified in the second lookup table. See Figure 4.5d and LHIZ.aml.

LHIZ <hue_lookup> <intensity_lookup> <size_item> {size_exponent} {size_factor} {minimum_size}
<hue_lookup> - specifies HLS hue;
<intensity_lookup> - specifies HLS saturation;
<size_item> - a ratio data field in the attribute table referenced by <hue_lookup>;
{size_exponent} {size_factor} {minimum_size} - define the scaling of sizes;
defaults: size_exponent = 1; size_factor = 0.15; minimum_size = 0.02 inches.

# Chapter Five
# Choropleth Symbolization in ARC/INFO

There are several techniques available in ARC/INFO for the production of choropleth maps (for area data that is continuous, and changes abruptly at polygon borders). These techniques include hue, texture, orientation and shape for nominal data, and size for ordinal to ratio data. As with the other monovariate sections, the AML's start off simply in order to build a base for bi- and multivariate mapping. Also these techniques tend to make use of lookup tables, which should be generated prior to executing the display AML's.

ARC/INFO provides a default set of area fills in the shadeset file, PLOTTER.SHD; there are several other available shadesets: CARTO.SHD, COLOR.SHD and COLORNAMES.SHD. These are displayed in the *Map Display and Query* guide's Appendix C, and the ability to create new symbols sets is discussed in chapter three of *Map Display and Query*.

## Monovariate Symbolization

Monovariate symbolization techniques for area data involve the use of hue, value, orientation and shape to represent data. This can be done when only one variable needs to be shown, or as an inset for bi- or multivariate maps.

### Nominal Data--Hue

Hue is best for representations of nominal data. In ARC/INFO's default shadeset, there are only three colors (red, green and blue) available--shade patterns 2, 3 and 4, and white--shade pattern 1. Rather than using the limited default colors, SHADECOLOR can be used to specify the output color. SHADECOLOR requires that the shadetype be set to a type other than the default HARDWARE, such as COLOR. When this is done, hues can be specified directly; an alternate method is to use one of ARC/INFO's color shadesets: COLOR.SHD and COLORNAMES.SHD. This AML (CH.aml) accomplishes the display of a data set with the shadeset method; see Figure 5.1a.

CH <cover> <data_item> <lookup_table> <shadeset>
    <cover> - a polygon coverage;
    <data_item> - the item in Polygon Attribute Table of <cover> referred to in <lookup_table>;
    <lookup_table> - relates data item to shades in <shadeset>;
    <shadeset> - a shadeset file.

### Monovariate Legends--Filled Polygons

In keeping with the idea of presenting as much information as possible in legends, AL.aml generates monovariate legends for area data that include the number of polygons in each category as well as defining the category symbolization. The AML automatically labels each category, as well as adjusting the display for nominal (discreet symbols) and ordinal (or better) data (continuous symbols). See Figure 5.1b for nominal data, and Figure 5.2b for ratio data.

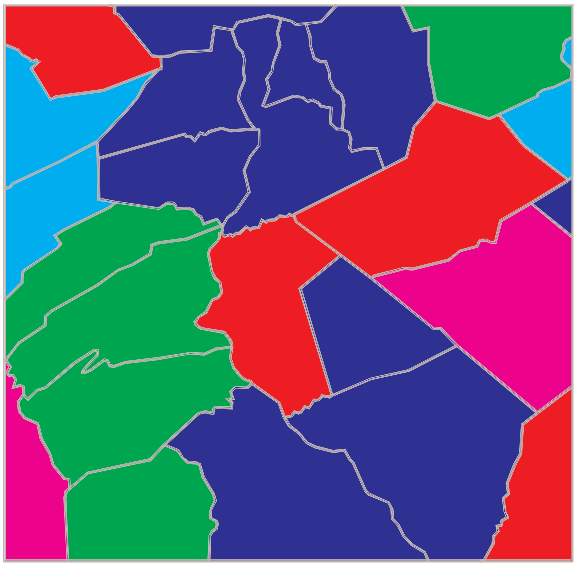


**Figure 5.1a: CH.aml**
Nominal data displayed with hue, using the color.shd shadeset.

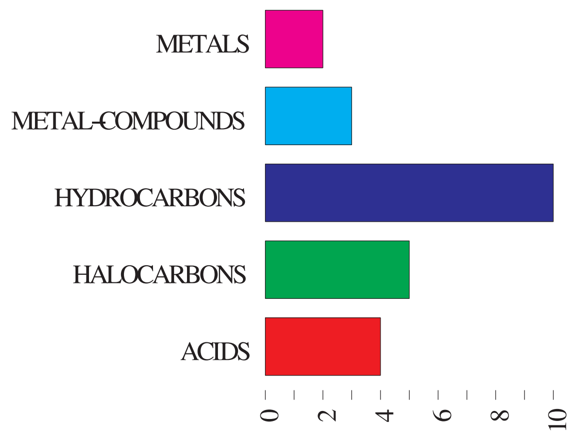



**Figure 5.1b: AL.aml**
A legend for figure a. This legend design gives the names associated with hues as well as the total number of regions in each class.

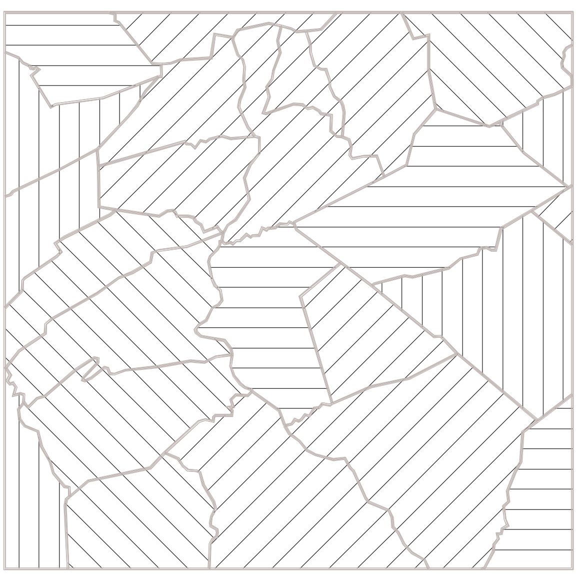


**Figure 5.1c: CH.aml**
A nominal data set displayed with orientation, using symbols from plotter.shd.

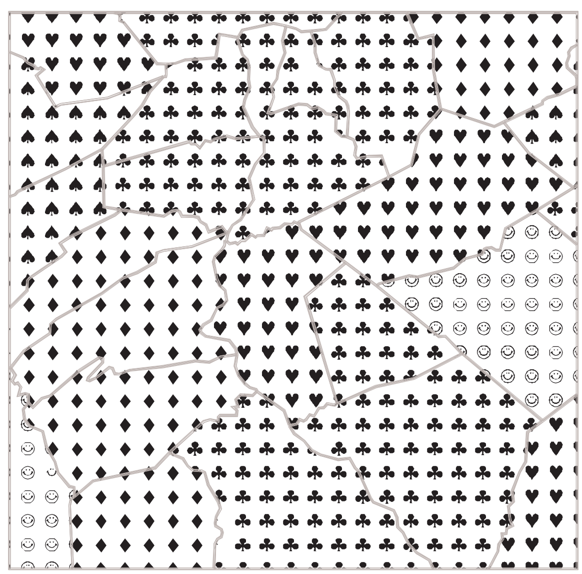


**Figure 5.1d: CS.aml**
A nominal data set displayed with shape by using the Agfa Compugraphic glyphs. This symbolization is best for areas that will not have specific borders.

AL <lookup_table> <start_x> <start_y> <size_x> <size_y> <h|v|t|a|l> {color_parameter_1} {color_parameter_2} {line_size} {textset} {font} {point_size} {floating_point_precision}

<lookup_table> - the lookup table that a legend will be based on;
<start_x> <start_y> - the lower left corner of the legend area;
<size_x> <size_y> - the size of the legend--all locations are in PAGEUNITS;
<h|v|t|a|l> - the symbolization method: Hue, Value, Texture, orientation (Angle), or shadeset symboLs;
{color_parameter_1} {color_paramter_2} - Hue: value and saturation (default: 50 and 100) for Value: hue and saturation (default 0 and 0), no effect for other symbolization;
{linesize} - for Texture and Angle line width--defaults to 0 (minimum displayable width);
{textset} {font} {point_size} - define the labelling text--defaults to a 10 point roman font;
{floating_point_precision} - the number of decimal places to be shown--the default is 2.

**Nominal Data--Orientation**

Orientation is best used for nominal data. There are three ways of achieving changes in area fill orientation: selecting a subset of a shadeset that has a pattern that changes in orientation, the POLYGONSHADES command (see Ordinal to Ratio Data--Orientation), and the SHADEANGLE command. The AML, CH.aml, uses the PLOTTER.SHD shadeset's symbols; see the above discussion of the AML and Figure 5.1c.

**Nominal Data--Shape**

Shape is best used for nominal data. The SHADETYPE command allows the specification of a marker to be used for the area fill. The macro, CS.aml, makes use of the Agfa Compugraphic font set to change area fill shape; see Figure 5.1d. Note the use of a loop to set up area fills.

CS <lookup_table> <shade_size> <shade_separation>

<lookup_table> - contains Agfa Compugraphic glyph numbers (font 94021);
<shade_size> - a size for the glyph symbols in PAGEUNITS;
<shade_separation> - the separation of the glyphs in X and Y in PAGEUNITS.

**Ordinal Data--Value**

Color value is best used for ordinal data. A new command in ARC/INFO release 6.0 allows the rapid generation of ordinal maps based on color value. The command SHADECOLOR allows a 'true color' specification of area fills. Prior to using SHADECOLOR, the shadetype must be set to color. Once this is done, colors can be specified by one of several color models (such as HLS, RGB and CMY). This macro uses lookup table symbols to specify value (lightness) in the HLS color model; see Figure 5.2a and CV.aml.

CV <lookup_table> {hue} {intensity}

<lookup_table> - specifies HLS lightness (values from 0 to 100);
{hue} {intensity} - specify a color that will be value shaded.

**Ordinal to Ratio Data--Orientation**

For interval/ratio data that should not be classed and is continuous with abrupt spatial changes (that is, it should be displayed as a choropleth map), ARC/INFO provides one automated way of generating an unclassed choropleth map. It uses changes in the orientation of area fills to vary the symbolization. The POLYGONSHADES command is used with an area fill pattern that will show

**Figure 5.2a: CV.aml**
An ordinal data set displayed with color value. This data set was classified with the Jenk's optimal classification system.

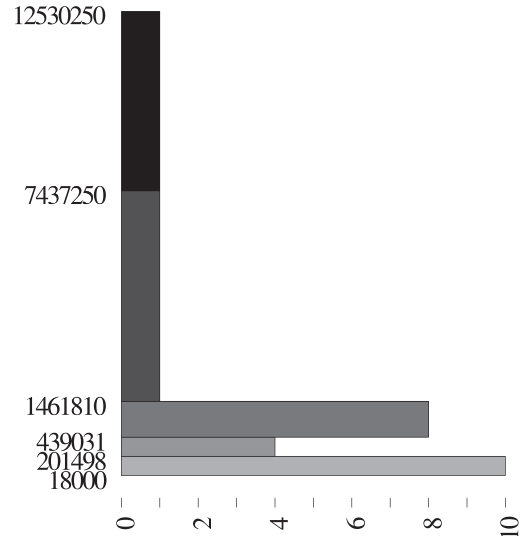



**Figure 5.2b: AL.aml**
A legend for figure a. The continuous data is displayed as an unbroken column.

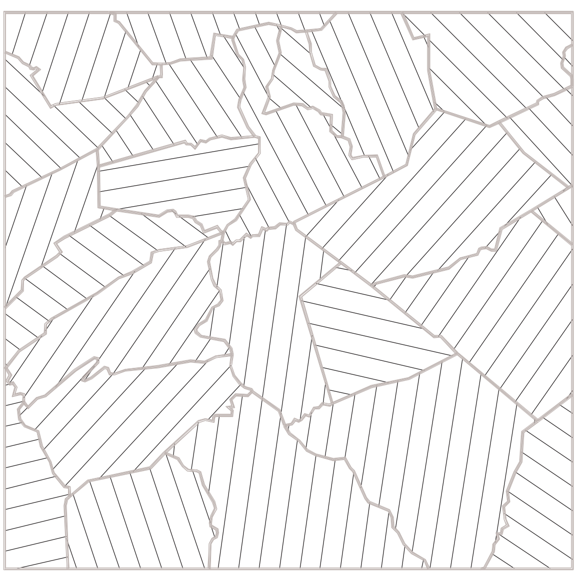


**Figure 5.2c: CAO.aml**
A ratio data set displayed with symbol orientation. The original data was scaled in INFO to range from 0 to 179 degrees.

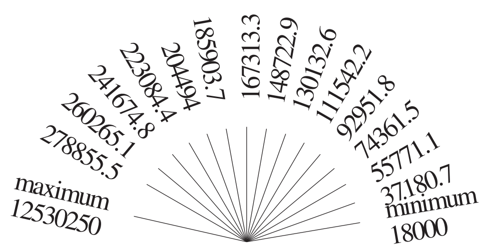



**Figure 5.2d: OL.aml**
A legend for figure c. This AML generates angled lines and labels (the labels are the original data range). It also labels the minimum and maximum.

angular changes (not a solid color fill, for example), such as shade 5 of PLOTTER.SHD, and an angle item (preset to a range from 0 to 180 degrees) in the Polygon Attribute Table. This AML displays this kind of data; see Figure 5.2c and CAO.aml.

CAO <cover> <angle_item> {shadeset} {shadesymbol}
<cover> - the polygon coverage to be displayed;
<angle_item> - a numeric data item in the Polygon Attribute Table of <cover>;
<shadeset> <shadesymbol> - select a shade to be oriented--defaults to plotter.shd 5.

### Monovariate Legends--Orientation

Because orientation can be used in point symbols as well as area symbols, the AML that follows is designed to work with either coverage type. It is also allows specification of a data variable that is not the same as the item that determined the drawing angle--an attribute item can be rescaled in INFO, for example, into another item and the second item can be used to draw area fills. The original item still holds the data values that the angles represent. By assuming a linear transformation between data values and angles, this AML can label the angles with an interpolated data value. See Figure 5.2d and OL.aml.

OL <cover> <point|poly> <data_item> <angle_item> <start_x> <start_y> <size>
{textset} {font} {point_size} {float_precision}
<cover> -the point or polygon coverage with angle data;
<point|poly> - the coverage type;
<data_item> - the original data field;
<angle_item> - the data field used to display rotated symbols;
<start_x> <start_y> - the location of the center point of the legend;
<size> - the length of each radiating line, in PAGEUNITS;
{textset} {font} {point_size} - define the label test--defaults to a 10 point roman font;
{float_precision} - the number of decimal places to be shown--the default is 2.

## Bivariate, Monochrome Symbolization

For area symbolization, the visual variables of texture, orientation and value can be combined to represent to variables. This set of macros is designed for continuous and abruptly changing spatial data of nominal to ratio attribute levels.

### Two Nominal Data Sets--Texture and Orientation

Texture is best for showing nominal data. With the default ARC/INFO shadeset, texture involves the selection of polygon shade patterns such as 5, 9 and 13. A more flexible approach to texture involves using the SHADETYPE and SHADESEPARATION commands. In this method the SHADETYPE must be set to type HATCH, first. Then the shadesize should be set; because of the dithering necessary to create gray shades on paper maps, the shadesize should be set to at least 0.02 inches. The SHADESEPARATION value is then varied to change the texture of the area fill. Because the general ARCPLOT commands require an integer in the SYMBOL column of a lookup table and this how the lookup table generation AMLs are set up, the AML, CTO.aml, scales each of the texture values by dividing by 100. SHADEANGLE is used to control the drawing angle of each shade; the SYMBOL values in the angle lookup table should be values from 0 to 180 degrees. See Figure 5.3a.

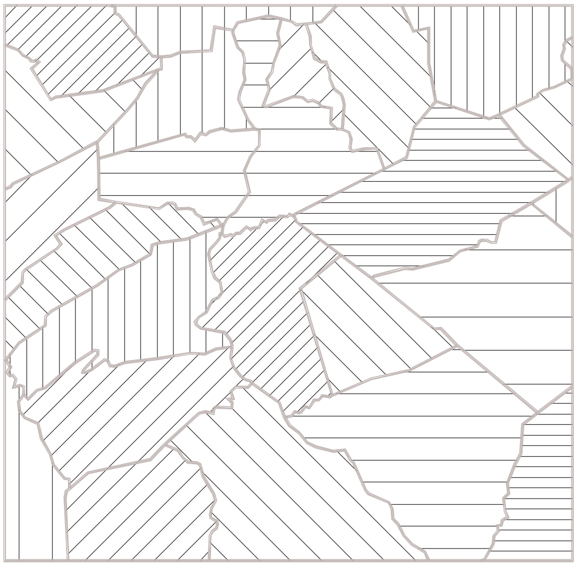


**Figure 5.3a: CTO.aml**
Nominal data displayed with texture and a second nominal data set displayed with orientation.

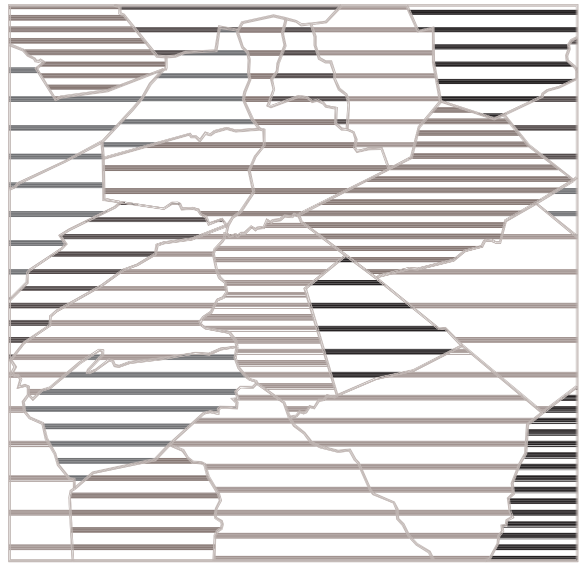


**Figure 5.3b: CTV.aml**
Nominal data displayed with texture and ordinal data displayed with value. This can be used to display a data set with its meta-data.

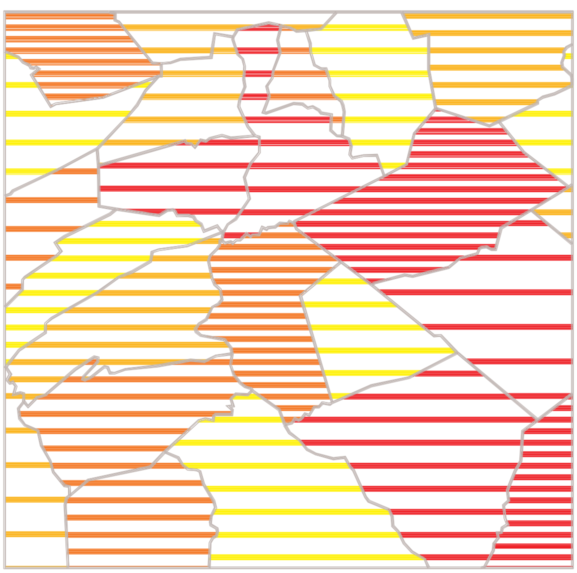


**Figure 5.3c: CTH.aml**
Nominal data displayed with texture and a second nominal data set displayed with hue. This should be used for two separate variables.

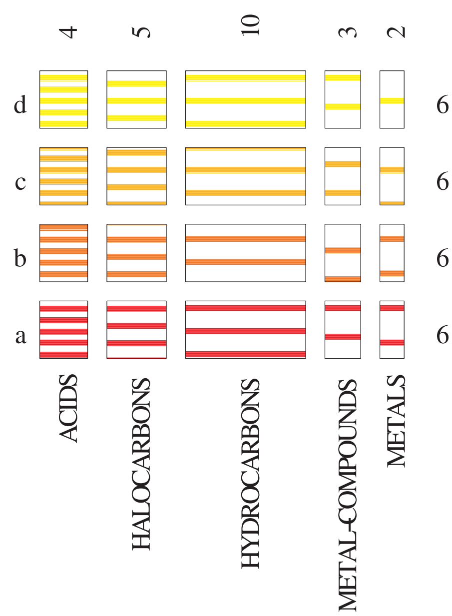



**Figure 5.3d: CBL.aml**
A legend for figure c. This AML will also generate legends for the bivariate mapping techniques shown in figures a and b.

CTO <texture_lookup> <orientation_lookup> {line_size}
<texture_lookup> - specifies line separation in hundredths of PAGEUNITS;
<orientation_lookup> - specifies drawing angles;
{line_size} - specifies the width of the fill lines--defaults to 0 (minimum displayable width).

### Nominal Data, and Ordinal Data--Texture Distinguished by Value

This AML makes use of two lookup tables to display texture and value; the first lookup table's symbol column should contain values for line separation, which will be divided by 100, and the second lookup table's symbol column should contain values for value (0 to 80). This method should be used when presenting data with meta-data, rather than texture, and orientation or hue, which are intended to show spatial correlation. See Figure 5.3b and CTV.aml.

CTV <texture_lookup> <value_lookup> {line_size} {angle} {hue} {intensity}
<texture_lookup> - specifies line separation in hundredths of PAGEUNITS;
<value_lookup> - specifies HLS lightness;
{line_size} - specifies fill line width--defaults to 0.03 inches;
{angle} - specifies a drawing angle--defaults to horizontal;
{hue} {intensity} - specify a color to be value shaded.

### Nominal Data-1 and 2--Texture as Intersecting Lines

One method of bivariate mapping that has been studied (by Carstensen) for communication of two correlated variables is intersecting lines. This involves use of texture as a visual variable by the controlled mixing of horizontal and vertical lines. This AML uses the SYMBOL value in both of the lookup tables to determine the texture of the vertical (the first lookup table) and horizontal (the second lookup table) lines; see Figure 5.4a and CTL.aml (the legend AML is presented in the Bivariate, Color Symbolization section).

CTL <vertical_lookup> <horizontal_lookup> {line_size}
<vertical_lookup> - specifies line separation in hundredths of PAGEUNITS for vertical lines;
<horizontal_lookup> - specifies line separation for horizontal lines;
{line_size} - specifies line width--defaults to 0 (minimum displayable width).

### Two Ordinal to Ratio Data Sets--Texture as Intersecting Lines

A variation of the previous AML allows the display of unclassed, ratio data. This AML (CIL.aml) makes use of the values in two columns of the Polygon Attribute Table to determine the spacing of the vertical and horizontal lines; see Figure 5.4c.

CIL <cover> <vertical_texture_item> <horizontal_texture_item> {maximum_separation} {minimum_separation} {separation_scale} {line_size}
<cover> - the polygon coverage to be displayed;
<vertical_texture_item> - a numeric data item in <cover> for vertical line spacing;
<horizontal_texture_item> - a numeric data item in <cover> for horizontal line spacing;
{maximum_separation} - the maximum width between lines--defaults to 0.3 inches;
{minimum_separation} - the minimum width between lines--defaults to 0.05 inches;
{separation_scale} - a real number to control exponential scaling of data to separation, the default is 1--linear scaling;
{line_size} - specifies a fill line width--defaults to 0 (minimum displayable width).

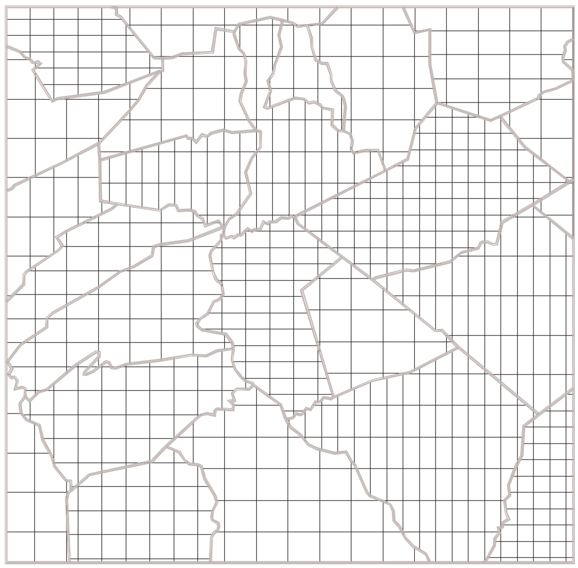


**Figure 5.4a: CTL.aml**
Two nominal data sets displayed with texture. This AML can also be used to display ordinal data.

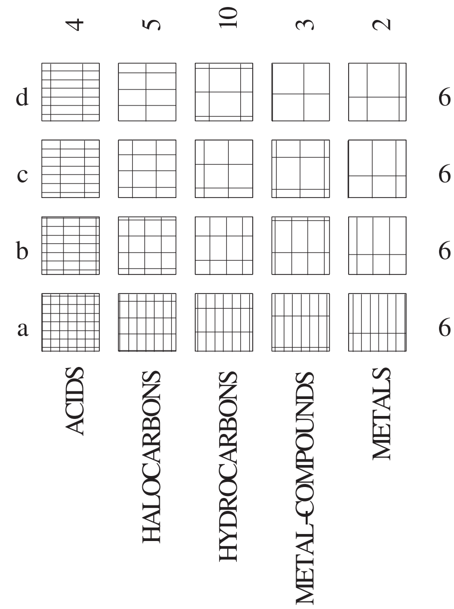



**Figure 5.4b: CBL.aml**
A legend for figure a. The shape of the boxes formed by the intersecting lines designates the values shown.

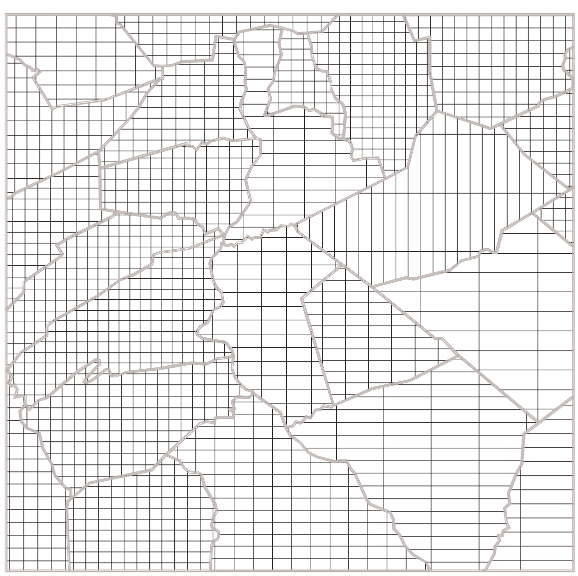


**Figure 5.4c: CIL.aml**
This intersecting lines map is an unclassed representation of two ratio data sets.

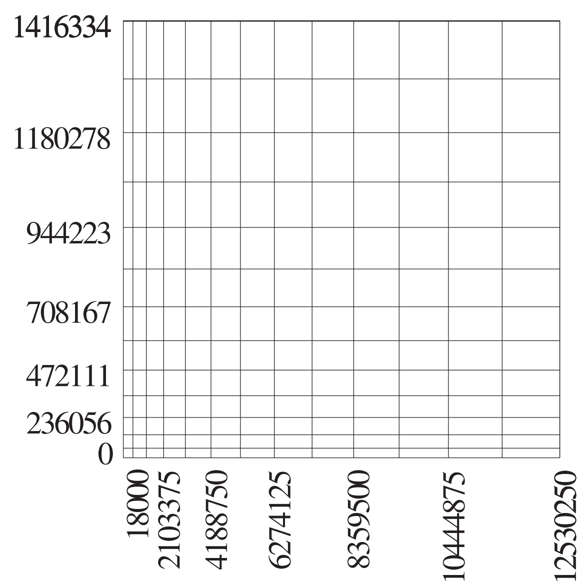



**Figure 5.4d: CILL.aml**
A legend for figure c. As with the classed map in figure a, the shape of the boxes designates the values represented.

### Bivariate Legends--Unclassed Texture

Because the previous AML uses Polygon Attribute Table to determine polygon texture, the AML that generates bivariate legends with lookup tables will not work. This AML only creates an unclassed texture legend, specifically for the maps generated with the previous AML. See Figure 5.4d and CILL.aml.

CILL <cover> <vertical_item> <horizontal_item> <start_x> <start_y> <size_x> <size_y>
{maximum_separation} {minimum_separation} {separation_scale} {line_size}
{textset} {font} {point_size} {floating_point_precision}
<cover> - the polygon coverage displayed with CIL.aml;
<vertical_texture_item> - a numeric data item in <cover> for vertical line spacing;
<horizontal_texture_item> - a numeric data item in <cover> for horizontal line spacing;
<start_x> <start_y> - the lower left corner of the legend's location;
<size_x> <size_y> - the size of the legend diagram;
{maximum_separation} - the maximum width between lines--defaults to 0.3 inches;
{minimum_separation} - the minimum width between lines--defaults to 0.05 inches;
{separation_scale} - a real number to control exponential scaling of data to separation, the default is 1--linear scaling;
{line_size} - specifies a fill line width--defaults to 0 (minimum displayable width);
{textset} {font} {point_size} - define the labeling text--defaults to a 10 point roman font;
{floating_point_precision} - the number of decimal places to be shown--the default is 2.

## Bivariate, Color Symbolization

With the addition of the color visual variables of hue and intensity, there are several more techniques available in ARC/INFO for the production of bivariate choropleth maps (for data that is continuous, and changes abruptly at polygon borders). These techniques include texture distinguished by hue, hue and intensity, dual-hue range and complementary-color maps.

### Two Nominal Data Sets--Texture Distinguished by Hue

This AML is a modification of the AML that generated Figure 5.3a; unlike that macro, this should be used for two nominal data sets, rather than data and meta-data. The AML requires a lookup table that specifies texture (in integers which are divided by 100) and a lookup table that specifies HLS hue. See Figure 5.3c and CTH.aml.

CTH <texture_lookup> <value_lookup> {line_size} {value} {intensity}
<texture_lookup> - specifies line separation in hundredths of PAGEUNITS;
<value_lookup> - specifies HLS lightness (from 0 to 100);
{line_size} - specifies a line width--defaults to 0.03 inches;
{value} {intensity} --defaults of 50 and 100 (maximum intensity}.

### Bivariate Legends--Lookup Table Based Displays

Along with the automation of polygon symbolization, the generation of legends can be automated. This AML (CBL.aml) creates bivariate legends for all of the area symbolization techniques that make use of two lookup tables. It automatically places labels on each of the axes, as well as the total number of occurrences in each column. See Figures 5.3d, 5.4b, 5.5b, 5.5d, 5.6b and 5.6d.

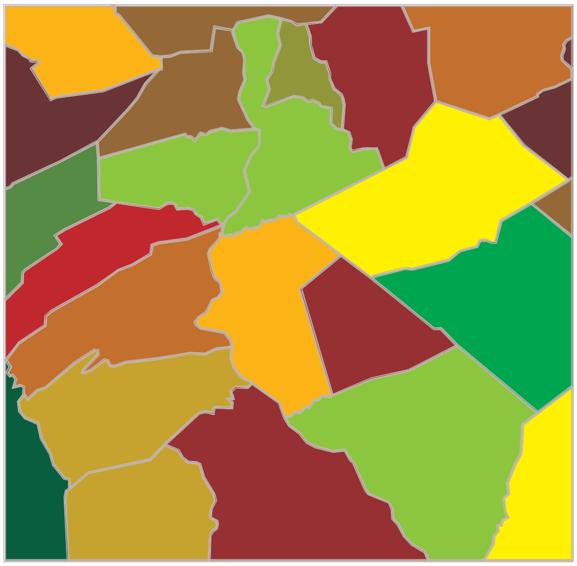


**Figure 5.5a: CDH.aml**
Two nominal data sets displayed with the dual hue range technique. This is best used to show the spatial relationship between two variables.

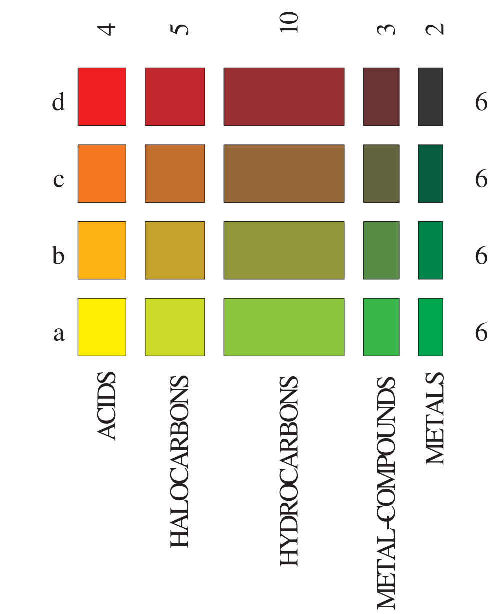



**Figure 5.5b: CBL.aml**
A legend for figure a. Nominal data is displayed with the size of the column or row indicating the total number of items in that column/row.

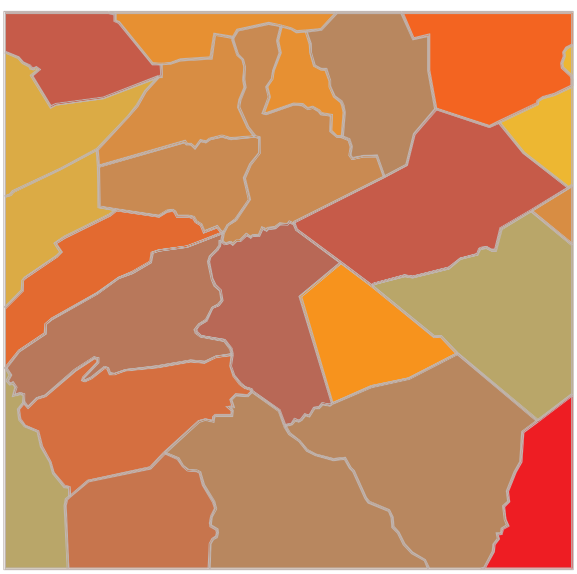


**Figure 5.5c: CHI.aml**
Nominal data displayed with hue and ordinal meta-data displayed with color intensity; intense colors are used to display highly certain values.

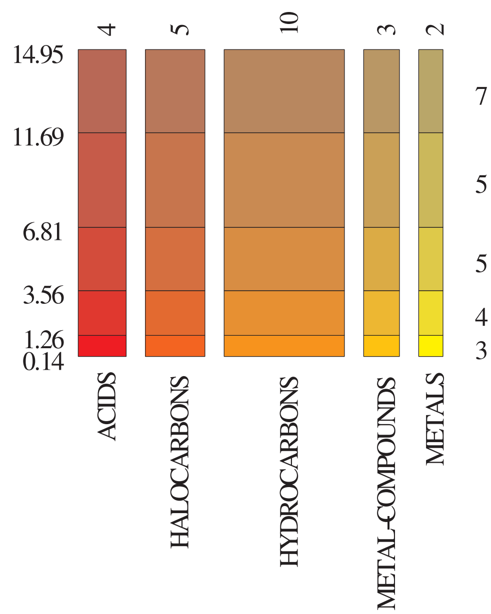



**Figure 5.5d: CBL.aml**
A legend for figure c. The AML separates nominal information and displays continuous (ordinal+) data as an unbroken column.

CBL <lookup_1> <lookup_2> <start_x> <start_y> <size_x> <size_y> <d|c|i|l|v|h|o>
{textset} {font} {point_size} {float_precision} {line_size}
{color_parameter_1} {color_parameter_2}
<lookup_1> <lookup_2> - the two lookup tables used in a bivariate mapping AML;
<start_x> <start_y> - the lower left corner of the legend display;
<size_x> <size_y> - the x and y sizes of the legend display;
<d|c|i|l|v|h|o> - the bivariate method used: Dual hue range, Complementary color, hue and Intensity, intersecting Lines, texture and Value, texture and Hue, and texture and Orientation;
{textset} {font} {point_size} - define the labeling text--defaults to a 10 point roman font;
{float_precision} - the number of decimal places to be shown--the default is 2;
{line_size} - for the texture methods, the fill line width;
{color_[parameter_1} {color_parameter_2} - for the texture with hue or value methods, the additional two color parameters.

### Two Nominal Data Sets--Dual Hue Ranges

To specify a hue, which can be used for mapping nominal or ordinal data, the SHADECOLOR command must be given with parameters that will establish a color. For dual hue range maps the shadecolors must be selected so that each data set uses one of the part-spectral color schemes (yellow to red, and yellow to green). See the table in the Dual Hue Range section in chapter three for a tables of possible color values. The lookup tables should set up to range from 0 to 100; one table will be used in the CMY color system to specify magenta (with yellow at 100, this controls the amount of red), and the other will specify cyan (with yellow at 100, this controls the amount of green). See Figure 5.5a and CDH.aml.

CDH <yellow-green_lookup> <yellow-red_lookup>
<yellow-green_lookup> - specifies CMY cyan (from 0 to 100);
<yellow-red_lookup> - specifies CMY magenta (from 0 to 100);

### Nominal Data, and Ordinal Data--Hue and Intensity

Color hue is best used for nominal data; intensity is best used for ordinal data (and generally only for meta-data, not a second data variable). The AML, CHI.aml, makes use of a lookup table that specifies HLS hue for a nominal data item, and HLS saturation (intensity) for an ordinal data item. See Figure 5.5c.

CHI <hue_lookup> <intensity_lookup>
<hue_lookup> -specifies HLS hue (from 0 to 360) for a nominal data set;
<intensity_lookup> - specifies HLS saturation (from 20 to 100) for an ordinal data set.

### Two Ordinal Data Sets--Complementary Colors

An alternative color scheme to the spectral encoding system given above is the complementary color scheme. This system is best for mapping two ordinal variables, because it highlights negative or positive correlation with a central gray zone (when classes are equal-interval or color specifications relate to data values). For example, to symbolize positive correlations, the following CMY color values generate an appropriate shading:

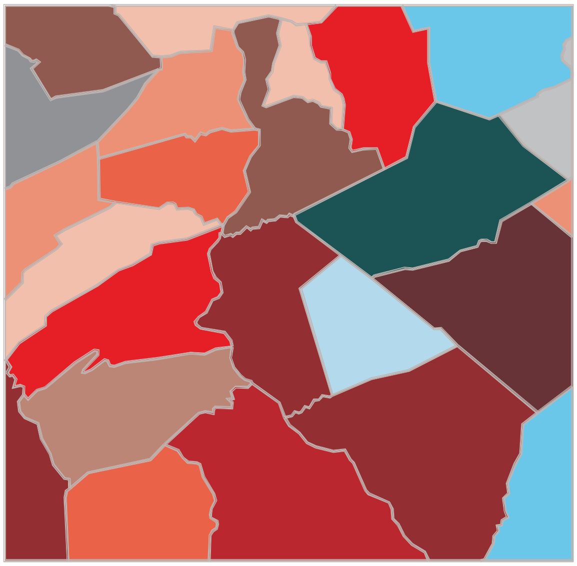


**Figure 5.6a: CCC.aml**
Two ordinal data sets displayed with the complementary color system. The large amount of red in the image can provoke a reaction of alarm.

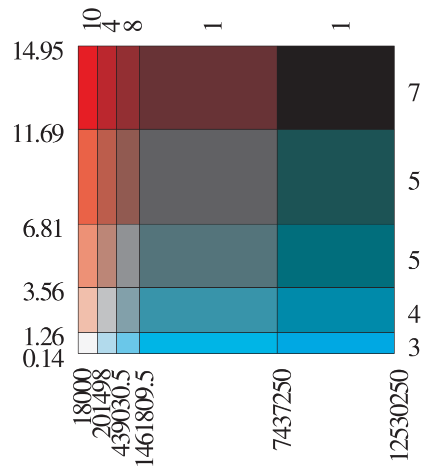



**Figure 5.6b: CBL.aml**
A legend for figure a. The use of two continuous variables is conveyed by using unbroken columns and rows, and placing labels at class breakpoints.

**Figure 5.6c: CCC.aml**
The ordinal data sets from figure c with a reverse progression in the cyan lookup. This causes a change in overall tone of the map.

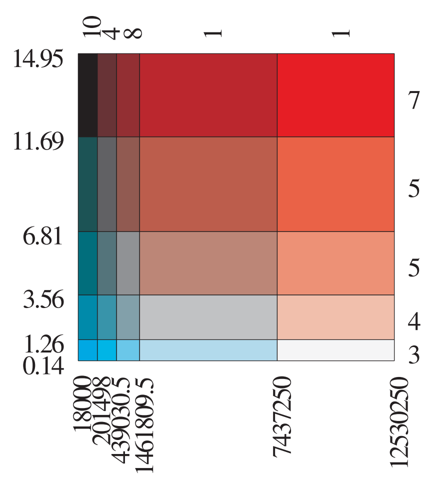



**Figure 5.6d: CBL.aml**
A legend for figure c. Note that with this ordering of cyan, high data values on both axes are displayed in red--the color of danger.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cyan: | 100 | Cyan: | 100 | Cyan: | 100 | Cyan: | 100 |
| Magenta: | 10 | Magenta: | 40 | Magenta: | 70 | Magenta: | 100 |
| Yellow: | 10 | Yellow: | 40 | Yellow: | 70 | Yellow: | 100 |
| Cyan: | 70 | Cyan: | 70 | Cyan: | 70 | Cyan: | 70 |
| Magenta: | 10 | Magenta: | 40 | Magenta: | 70 | Magenta: | 100 |
| Yellow: | 10 | Yellow: | 40 | Yellow: | 70 | Yellow: | 100 |
| Cyan: | 40 | Cyan: | 40 | Cyan: | 40 | Cyan: | 40 |
| Magenta: | 10 | Magenta: | 40 | Magenta: | 70 | Magenta: | 100 |
| Yellow: | 10 | Yellow: | 40 | Yellow: | 70 | Yellow: | 100 |
| Cyan: | 10 | Cyan: | 10 | Cyan: | 10 | Cyan: | 10 |
| Magenta: | 10 | Magenta: | 40 | Magenta: | 70 | Magenta: | 100 |
| Yellow: | 10 | Yellow: | 40 | Yellow: | 70 | Yellow: | 100 |

By reversing the values in either cyan or magenta/yellow, the direction of correlation that is highlighted is reversed. This AML requires two lookup tables, one of which has CMY values for cyan, and the other has values for red (magenta and yellow), like the table above. See Figure 5.6a (positive correlation) and 5.6c (the same data, but with negative correlation highlighting) and CCC.aml.

CCC <cyan_lookup> <red_lookup>
<cyan_lookup> - specifies CMY cyan (from 10 to 100);
<red_lookup> - specifies CMY red (from 10 to 100).

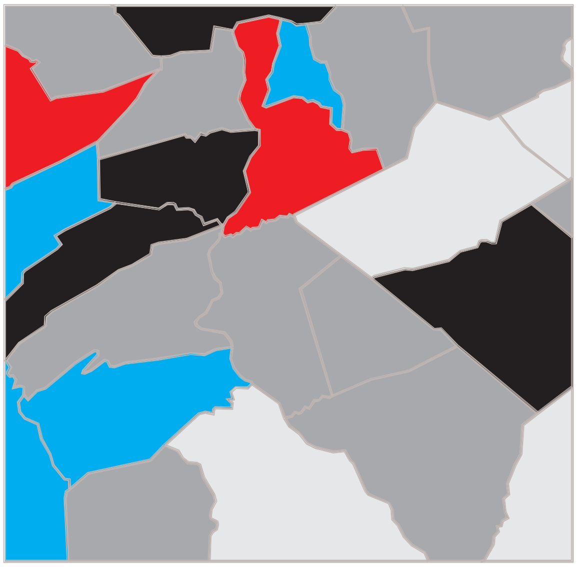
### Two Ratio Data Sets--Eyton's Ellipse

Another technique for displaying correlations between two data sets is the equiprobability ellipse (Eyton 1984). This method uses four categories (a two by two matrix) to indicate the extreme values in the scatter plot of the two data sets, and a fifth category that indicates the central cluster of data. This method is effective at highlighting correlations between two data sets, but may not be effective in communicating uncertainty. See Figure 5.7a and CEE.aml.

CEE <ellipse_lookup>
<ellipse_lookup> - a lookup table for polygons generated by EYTON.aml.

### Bivariate Legends--Eyton's Ellipse

Because of the uniqueness of the this bivariate representation system, a different AML must be used to generate the legend. EEL.aml generates a scatter plot of the two correlated data sets, and shades each marker with the color used to represent that data value. The AML works with all of the equiprobability ellipse displays (points, lines, and polygons). See Figure 5.7b and EEL.aml.



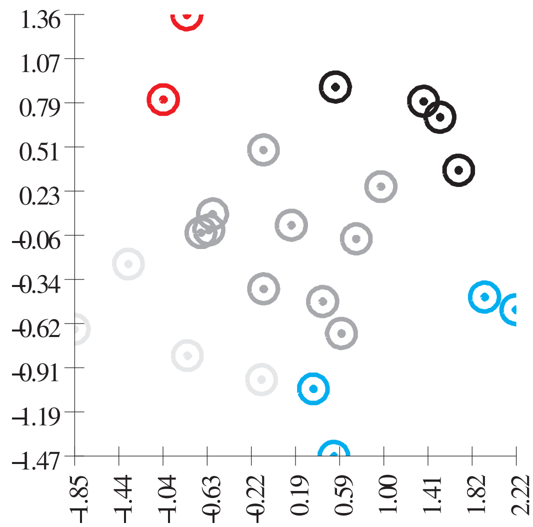
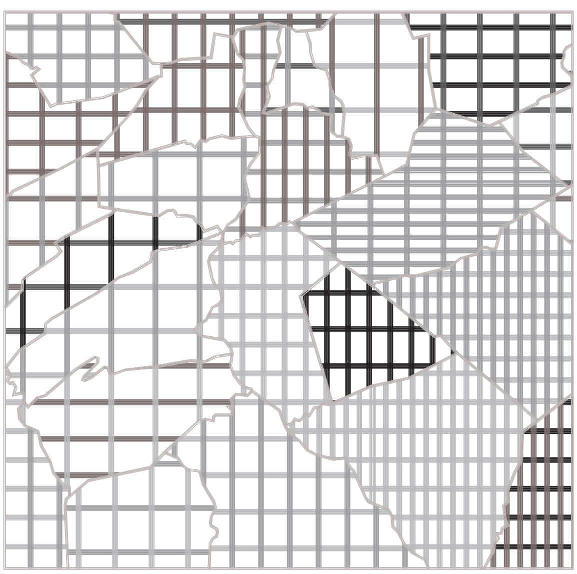



**Figure 5.7a: CEE.aml**

Two ratio data sets displayed using the equiprobability ellipse method. Correlation between the two variables is determined to select a middle class.

**Figure 5.7b: EEL.aml**

A legend for figure a. A scatter plot shows the location of each data pair, and the hues show the color used in the map.



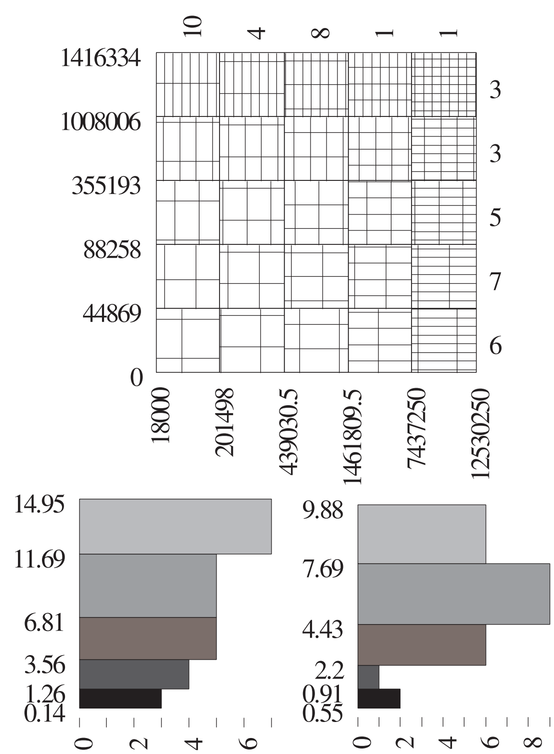



**Figure 5.7c: CTLVV.aml**

Two classed ratio (ordinal) data sets displayed using texture with meta-data for each data set displayed by using color value; a legend is on the right.

EEL <ellipse_lookup> <start_x> <start_y> <size_x> <size_y> {markerset} {markersymbol} {markersize} {textset} {font} {point_size} {float_precision}
<ellipse_lookup> - a lookup table generated by EYTON.aml;
<start_x> <start_y> - the lower left corner of the scatter plot;
<size_x> <size_y> - the size of the scatter plot in PAGEUNITS;
{markerset} {markersymbol} - specify a marker to use in the scatter plot, the default is municipal.mrk, symbol 104;
{markersize} - the size of the scatter plot marker--defaults to 0.15 inches;
{textset} {font} {point_size} - define the labelling text--defaults to a 10 point roman font;
{float_precision} - defines the number of decimal places that will be shown--defaults to 2.

## Multivariate Symbolization

Multivariate choropleth maps can expand on some of the bivariate mapping techniques in order to allow meta-data to be presented with data, in an otherwise bivariate map. Multivariate maps can also be used to present more than two data sets to the map reader.

### Two Nominal or Ordinal Data Sets, Ordinal Data Sets--Texture with Value

For monochrome displays, intersecting lines allow control of texture to convey a data value. By adding control of line value, meta-data can be presented along with the two primary data sets. This AML requires two lookup tables with values that will be divided by 100 to set texture and two lookup tables with HLS value data (0 to 80). See Figure 5.7c and CTLVV.aml. (The legends are generated with CBL.aml and AL.aml).

CTLVV <horizontal_lookup> <horizontal_value_lookup> <vertical_lookup> <vertical_value_lookup> {hue} {intensity} {line_size}
<horizontal_lookup> - specifies line separation in hundredths of PAGEUNITS:
<horizontal_value_lookup> - specifies HLS lightness (from 0 to 80);
<vertical_lookup> - specifies line separation i hundredths of PAGEUNITS;
<vertical_value_lookup> - specifies HLS lightness (from 0 to 80);
{hue} {intensity} - specify a hue to be value shaded;
{line_size} - specifies line width--defaults to 0.03 inches.

### Three Ordinal to Ratio Data Sets--Red, Green and Blue Symbolization

Three ratio data sets can be displayed by scaling the data to the range of the RGB color system (0 to 255) and then using the data values to specify color. This AML requires three numeric data items in the polygon attribute table; a switch can be specified to direct the AML to use the CMY color system, which can represent data values at the low end of each range better than the RGB system. See Figures 5.8a and 5.8c, and CRGB.aml.

CRGB <cover> <red_item> <green_item> <blue_item> {r|c}
<cover> - the polygon coverage to be displayed;
<red_item> <green_item> <blue_item> - three numeric data fields in the Polygon Attribute Table of <cover>;
{r|c} - Rgb or Cmy color space.

### Multivariate Legends--RGB Space

Because the specification of color requires three dimensions, development of a legend is difficult--it must appear to be three dimensional. This AML takes the same three numeric data items used for the previous AML and displays the colors

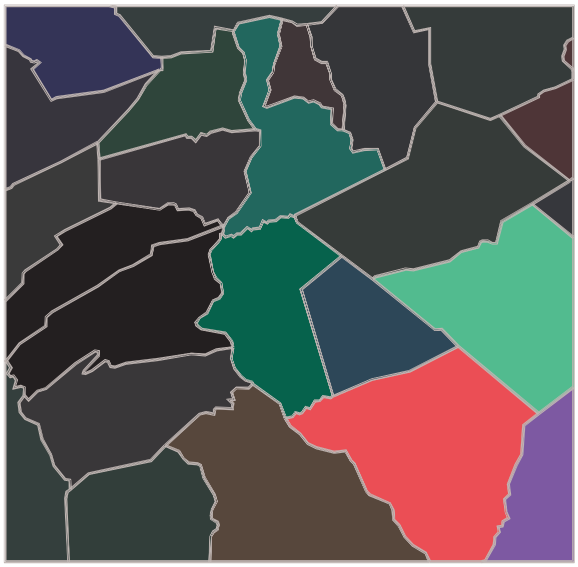


**Figure 5.8a: CRGB.aml**

Three ratio data sets displayed with red, green and blue color ranges. This is the ‘false color’ imagery often used with satellite data.

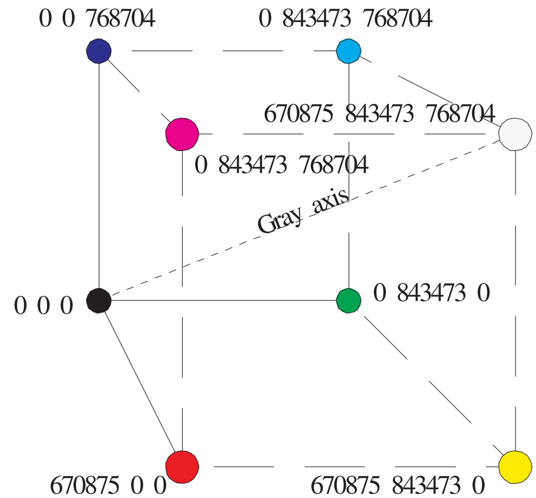



**Figure 5.8b: RGBL.aml**

A legend for figure a. The AML finds the minimum and maximum data values, and displays these along with the colors at the corners of the color cube.

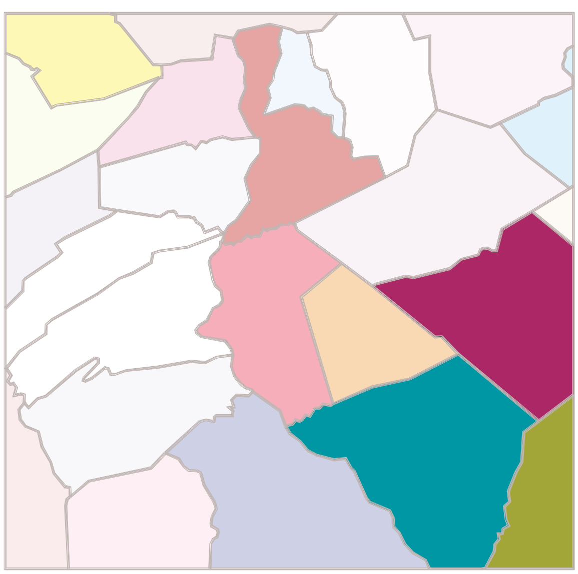


**Figure 5.8c: CRGB.aml**

Three ratio data sets displayed with cyan, magenta and yellow color ranges. This color model shows variability at the low end of the data better than RGB.

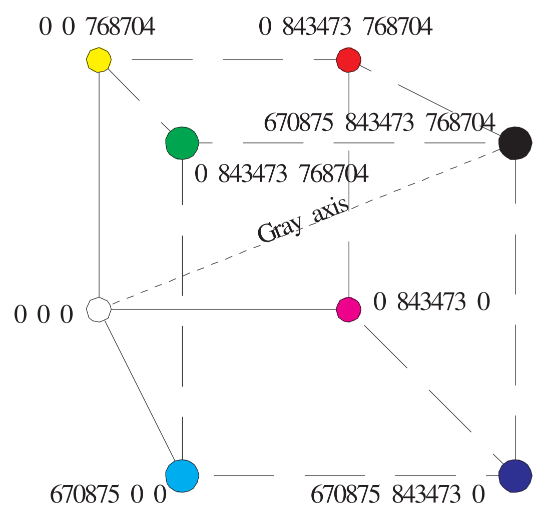



**Figure 5.8d: RGBL.aml**

A legend for figure c. The AML increases the size of the front of the cube and the front four circles, in order to enhance the legend’s ‘3-D’ appearance.

and data values associated with the corners of the color cube. The AML will also work with points, lines and grids. See Figures 5.8b and 5.8d, and RGBL.aml.

RGBL <cover> <point|line|poly|grid> <red_item> <green_item> <blue_item> <start_x> <start_y>
<size_x> <size_y> {r|c} {textset} {font} {point} {floating_point_precision}
<cover> - the coverage that contains the data;
<point|line|poly|grid> - the type of <cover>;
<red_item> <green_item> <blue_item> - numeric fields in the attribute table of <cover>;
<start_x> <start_y> - the lower left corner of the back plane of the legend;
<size_x> <size_y> - the size of the back plane, in PAGEUNITS;
{r|c} - Rgb or Cmy color space;
{textset} {font} {point} - define the labelling text--defaults to a 10 point roman font;
{floating_point_precision} - the number of decimal places to be shown--defaults to 2.

# Chapter Six
# Graduated Symbol Symbolization in ARC/INFO

For the generation of graduated symbol maps (which should be used for discrete, abruptly changing, ordinal to ratio area data), ARC/INFO provides two commands that allow rapid generation of these maps: SPOTSIZE and POLYGONSPOT. SPOTSIZE controls the scaling of data to circle size, and POLYGONSPOT does the plotting of the circles. A second method of graduated symbol mapping is cartograms. These maps scale individual polygons on the basis of a data value. Cartogram creation requires interactive map composition, both for the scaling of polygons and the repositioning of scaled polygons.

## Monovariate Maps

Monovariate symbolization for graduated symbols is based on the manipulation of size to specify data. This makes graduated symbols appropriate for ordinal to ratio data.

### Ordinal to Ratio Data--Graduated Circles

This AML (GC.aml) uses the commands SPOTSIZE and POLYGONSPOT to generate a graduated circle map. Because SPOTSIZE requires the minimum and maximum data values, the AML searches the data set for the extremes, and then displays the data. See Figure 6.1a.

GC <cover> <size_item> {minimum_size} {maximum_size}
<cover> - the polygon coverage to be displayed;
<size_item> - a numeric data field in the attribute table of <cover>;
{minimum_size} - the circle representing the smallest data value--defaults to 0.05 inches;
{maximum_size} - the circle representing the largest data value--defaults to 0.5 inches.

### Ordinal to Ratio Data--Cartograms

There are several varieties of cartograms; this AML simply scales a polygon be a data value, and places the scaled polygon in the center of the original polygon's extents. Because this does not handle unusually shaped polygons well, the AML allows the scaled polygons to be repositioned. Since the representation of data in cartograms is done with the polygon outline, cartograms are best used to display density data (the AML calculates density from the data item and the polygon size). See Figure 6.1b and GG.aml.

GG <cover> <size_item> {scale_factor}
<cover> - the polygon coverage to be displayed;
<size_item> - a numeric data field in the attribute table of <cover>;
{scale_factor} - specifies the exponent used to calculate cartogram size--defaults to 1.

## Monochrome, Bivariate Symbolization

Symbol size is good for either ordinal or interval/ratio data. For the generation of value-shaded, graduated symbol maps (which should be used for discrete, abruptly changing, ordinal to ratio area data and meta-data), size must be used in conjunction with SHADECOLOR. This combination allows the display of either data and meta-data, or two data variables.

**Figure 6.1a: GC.aml**
Discontinuous and abruptly changing, ratio area data displayed with graduated circles.

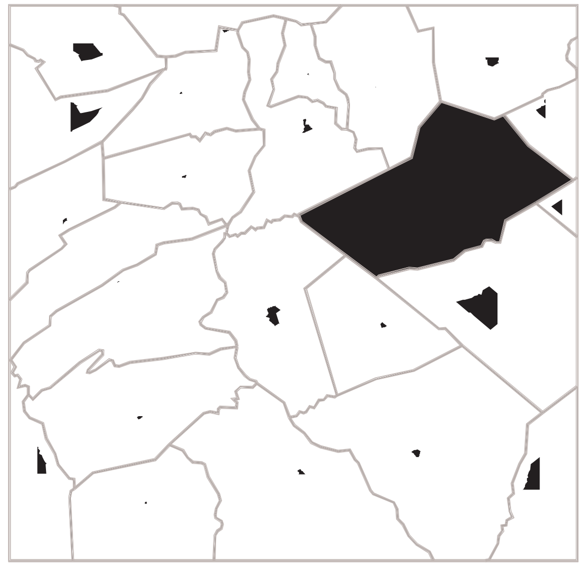


**Figure 6.1b: GG.aml**
Ratio area data displayed with cartograms. This shows the AML's defaults; note the extremely small sizes and the misplaced polygons.

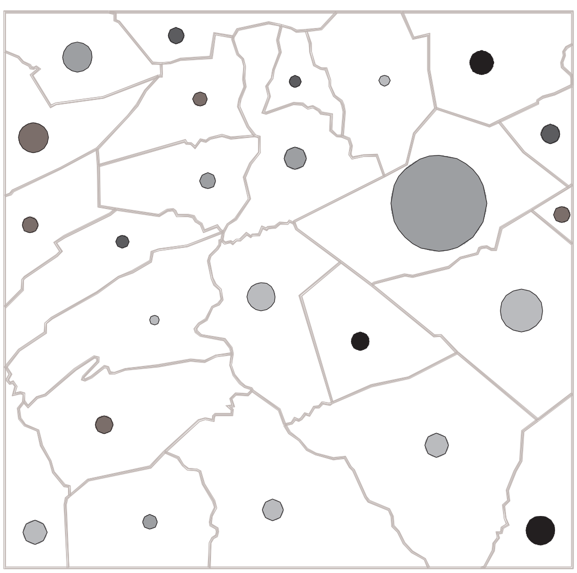


**Figure 6.1c: GCV.aml**
Graduated circles that are value shaded with an ordinal data set. This is best used for meta-data, although a second data set can be displayed as well.

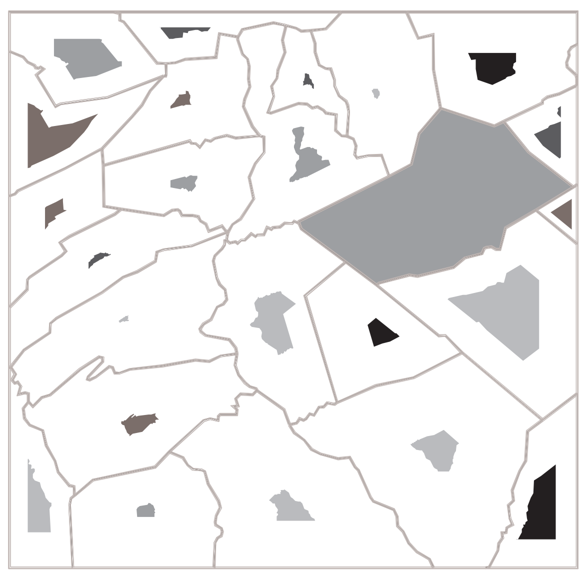


**Figure 6.1d: GGV.aml**
Value shaded cartograms. The size data is the same as figure b, but the scale exponent is 0.5--representing smaller data values with larger polygons.

### Ordinal to Ratio Data, and Ordinal Data--Graduated Circles and Value

This AML (GCV.aml) develops the display of graduated circles to include the value shading of the circle. The value_lookup must specify lightness data in the HLS color model, with value ranging from 0 black to no more than 90 (light gray), in order to avoid white symbols on a white background. See Figure 6.1c.

GCV <size_item> <value_lookup> {minimum_size} {maximum_size} {hue} {saturation}
<size_item> - a numeric data field in the attribute table referred to by <value_lookup>;
<value_lookup> - specifies HLS lightness data;
{minimum_size} {maximum_size} - the minimum and maximum circle sizes, the defaults are 0.05 and 0.5 inches;
{hue} {saturation} - specify a color that will be value shaded.

### Ordinal to Ratio Data, and Ordinal Data--Cartograms Shaded by Value

By using interactive map composition, cartograms (use of scaled areas to represent a data value) can be drawn in ARC/INFO. Value is used to represent a second variable; this can be used for a second variable when color is not an available option, or meta-data. See Figure 6.1d and GGV.aml.

GGV <size_item> <value_lookup> {scale_factor} {hue} {intensity}
<size_item> - a numeric data field in the attribute table referred to by <value_lookup>;
<value_lookup> - specifies HLS lightness data;
{scale_factor} - specifies the exponent used to calculated cartogram size--defaults to 1;
{hue} {intensity} - specify a color that will be value shaded.

## Color, Bivariate Symbolization

Symbol size is good for either ordinal or interval/ratio data. Graduated symbol maps should be used for discrete, abruptly changing, ordinal to ratio area data. Color hue can be used to display a second, nominal data set.

### Ordinal to Ratio Data, and Nominal Data--Graduated Circles and Hue

Unlike the macro used to generate Figure 6.1c, this AML should be used for two data sets--one ratio and one nominal, rather than data and meta-data. See Figure 6.2a and GCH.aml.

GCH <size_item> <hue_lookup> {minimum_size} {maximum_size} {value} {intensity}
<size_item> - a numeric data field in attribute table referred to by <hue_lookup>;
<hue_lookup> - specifies HLS hue data (from 0 to 360);
{minimum_size} - the size of the smallest data value's circle--defaults to 0.05 inches;
{maximum_size} - the size of the largest data value's circle--defaults to 0.5 inches;
{value} {intensity} --default to 50 and 100 (maximum intensity).

### Graduated Circle Legends

This AML (GCL.aml) generates a series of circles that are labeled with the data value represented by that size, and filled with the polygon fill in shadesymbol 1. It works for all of the graduated symbol AMLs (point and area). It should be supplemented by an additional figure if the circles are hue or value shaded (see Figure 6.2b).

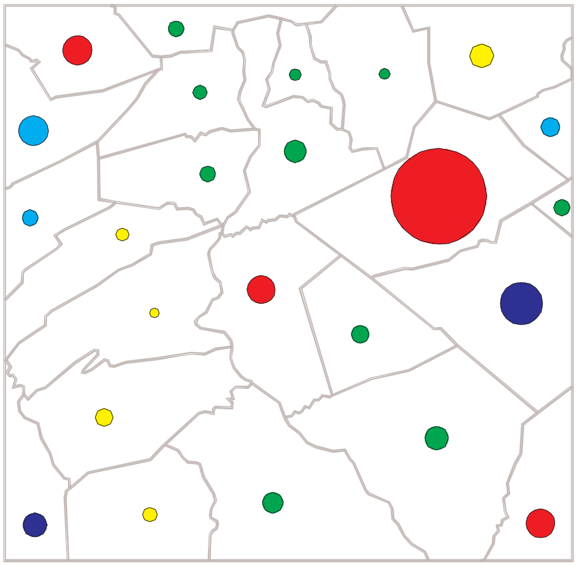


**Figure 6.2a: GCH.aml**
Discontinuous and abruptly changing, ratio area data displayed with graduated circles that are hue shaded (to represent a nominal data set).

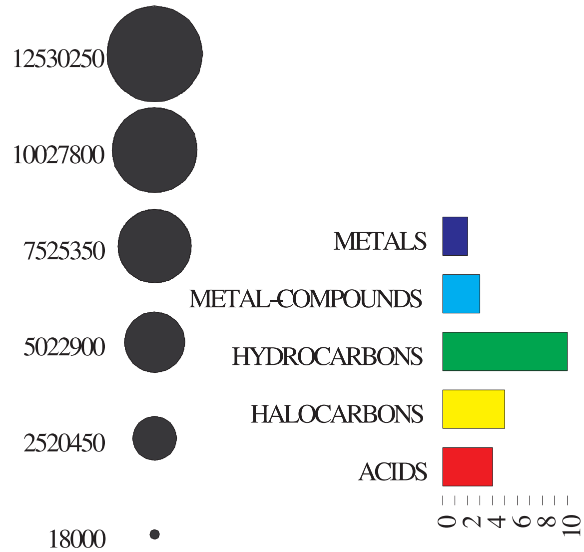



**Figure 6.2b: GCL.aml**
A legend for figure a. The graduated circle legend identifies size data, and the hue data is displayed with AL.aml.

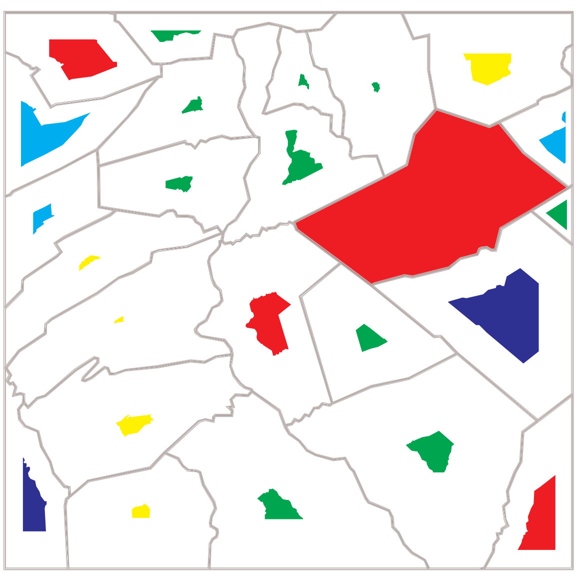


**Figure 6.2c: GGH.aml**
Hue shaded cartograms. These can be used to represent a nominal data set and a ratio (density) data set.

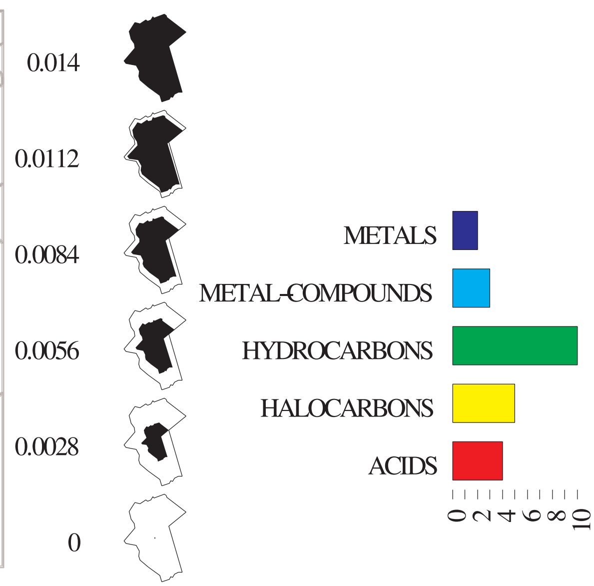



**Figure 6.2d: GGL.aml**
A legend for figure c. Cartogram size data is displayed with reference to an unscaled polygon, indicating that the symbols are based on polygon size.

GCL <cover> <point|node|poly> <size_item> <start_x> <start_y> <number_of_symbols>
{minimum_size} {maximum_size} {textset} {font} {point} {float_precision}
<cover> - the coverage that was displayed with graduated circles;
<point|node|poly> - the type of <cover>;
<size_item> - the numeric data item in the attribute table of <cover>;
<start_x> <start_y> - lower center of the column of circles;
<number_of_symbols> - the number of circle to be drawn for the legend;
{minimum_size} - the size used to display the smallest data value--defaults to 0.05 inches;
{maximum_size} - the size used to display the largest data value--defaults to 0.5 inches;
{textset} {font} {point} - define the labelling text--defaults to a 10 point roman font;
{float_precision} - the number of decimal places to be shown--defaults to 2.

### Ordinal to Ratio Data, and Nominal Data--Cartograms with Hue

This AML should be used to display a nominal data set, along with density data. See Figure 6.2c and GGH.aml.

GGH <size_item> <hue_lookup> {value} {intensity} {size_scale}
<size_item> - a numeric data field in the attribute table referred to by <hue_lookup>;
<hue_lookup> - specifies HLS hue (from 0 to 300);
{value} {intensity} - default to 50 and 100 (maximum intensity);
{size_scale} - specifies the exponent used to calculate cartogram size--defaults to 1.

### Cartogram Legends

This AML, like the graduated circle legend AML, generates a series of polygons that represent data values. Each polygon is placed inside an unscaled outline of the polygon, and labeled. The polygon is drawn with the current polygon shade fill. This legend must also be supplemented by a legend specifying hue, value, or hue and intensity, if more than one variable is shown on the map (see Figure 6.2d).

GGL <cover> <size_item> <start_x> <start_y> <size_x> <size_y> <number_of_symbols
{scale} {polygon} {textset} {font} {point} {precision}
<cover> - the polygon coverage displayed with cartograms;
<size_item> - a numeric data field in the attribute table of <cover>;
<start_x> <start_y> - the lower left corner of the legend area;
<size_x> <size_y> - the drawing size of each polygon;
<number_of_symbols> - the total number of legend symbols;
{scale} - the exponent used to calculate cartogram size--defaults to 1;
{polygon} - the record number of an example polygon--defaults to the middle record;
{textset} {font} {point} - define the labelling text--defaults to a 10 point roman font;
{precision} - the number of decimal places to be shown--defaults to 4.

## Multivariate Symbolization

Two multivariate symbolization methods are presented here. Hue, intensity and size is a development of hue and size. Graduated pie symbols allow the display of a total value and a set of components of that value.

### Ratio Data, Nominal Data, and Ordinal Data--Circles, Hue and Intensity

Like all graduated symbols, size should be used to represent a ratio variable (that is discrete and abruptly changing). Hue can be used to represent a second, nominal variable, and intensity can be used to represent an ordinal data--including ratio meta-data. See Figure 6.3a and GCHI.aml.

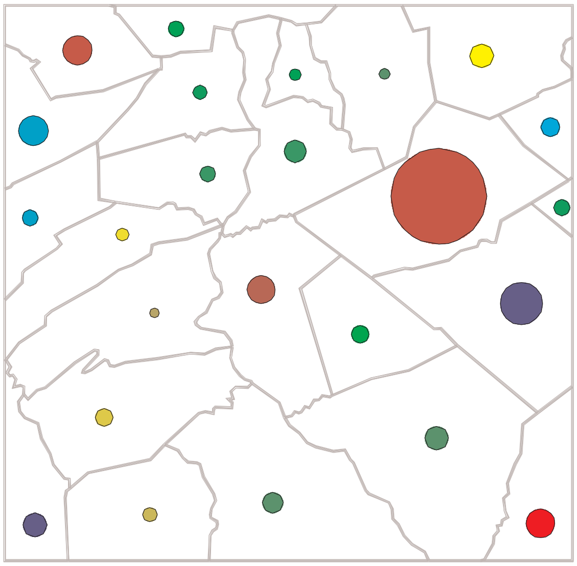


**Figure 6.3a: GCHI.aml**
Graduated circles shaded with color hue (a nominal data set) and intensity (an ordinal data set).

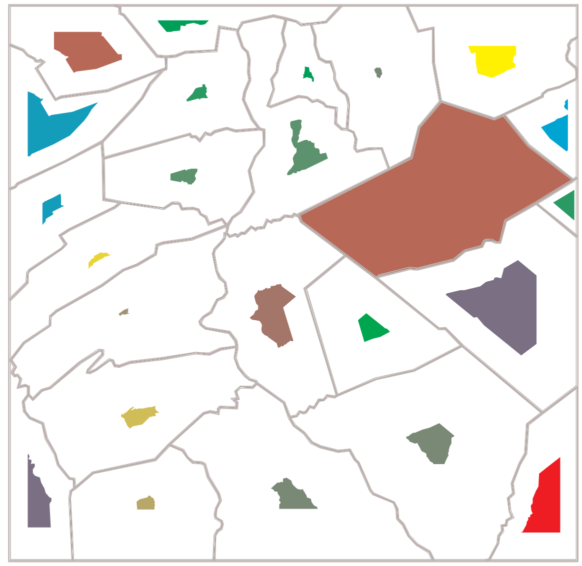


**Figure 6.3b: GGHI.aml**
Cartograms shaded with hue and intensity. Cartograms should be used to represent density data.

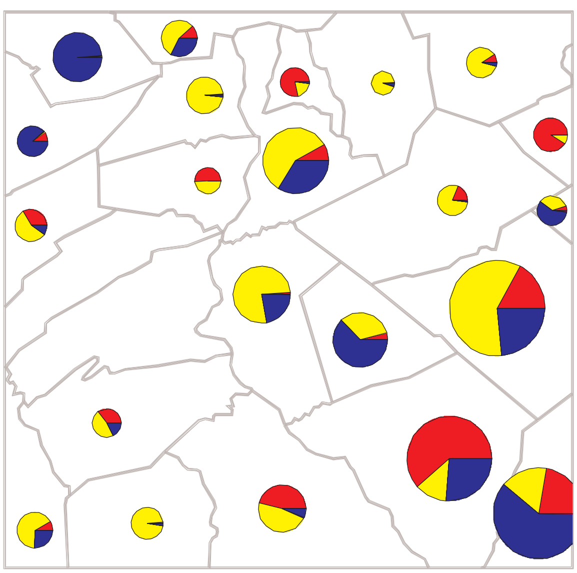


**Figure 6.3c: GP.aml**
Graduated pie symbols. These show a total value (size) and the components of the total (each pie wedge).

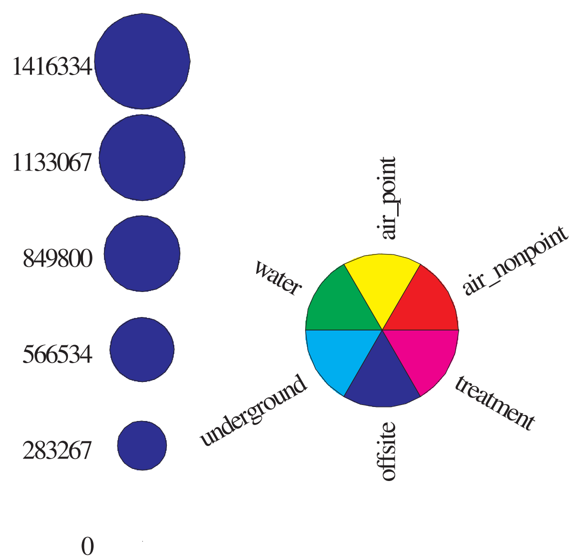



**Figure 6.3d: GPL.aml**
A legend for figure c. The size data can is displayed with the graduated circle legend AML. Pie slices are shaded and labeled by this AML.

GCHI <size_item> <hue_lookup> <intensity_lookup> {minimum_size} {maximum_size}
<size_item> - a numeric data field in the cover referred to by <hue_lookup>;
<hue_lookup> - specifies HLS hue (from 0 to 300);
<intensity_lookup> - specifies HLS saturation (from 20 to 100);
{minimum_size} - the size representing the smallest data value--defaults to 0.05 inches;
{maximum_size} - the size representing the largest data value--defaults to 0.5 inches.

### Ratio Data, Nominal Data, and Ordinal Data--Cartograms, Hue and Intensity

Like the previous AML, this macro should be used to represent a ratio data set, a nominal data set, and an ordinal data set (which can be ratio meta-data). Like the other cartogram AML's this should be used to display density information. See Figure 6.3b and GGHI.aml.

GGHI <size_item> <hue_lookup> <intensity_lookup> {size_scale}
<size_item> - a numeric data filed in the coverage referred to by <hue_lookup>;
<hue_lookup> - specifies HLS hue data (from 0 to 300);
<intensity_lookup> - specifies HLS saturation (from 20 to 100);
{size_scale} - specifies the exponent used to calculate cartogram size--defaults to 1.

### Ratio Data--Polygon Pie Graphs

In addition to the appropriate spatial characteristics, graduated pie maps should also have a data value that represents a total of multiple elements (this total determines circle size). Each element displayed a fraction (pie wedge) of the whole. The AML automatically determines hues for each element. See Figure 6.3c and GP.aml.

GP <cover> <size_item> <minimum_size> <maximum_size> <number_of_subitems>
<item_1...item_n>
<cover> - a polygon coverage with data that is discontinuous and abruptly changing;
<size_item> - a numeric field in the attribute table of <cover>, which is a total of <items>;
<minimum_size> - the size of the smallest data value's circle (0.05 may be good);
<maximum_size> - the size of the largest data value's circle (0.5 may be good);
<number_of_subitems> - the total number of items given in <item_1...item_n>;
<item_1...item_n> - numeric fields in <cover> that represent elements of <size_item>.

### Graduated Pie Legends

Legends for graduated pie can be generated by using the graduated circle legend AML for displaying size data, and using GPL.aml to display the hues used to shade pie slices. This AML uses the same hue selection routine as the graduated pie mapping AMLs, but it only draws one circle with equally sized pie slices. See Figure 6.3d.

GPL <pie_size> <center_x> <center_y> <number_of_data_items> {textset} {font} {point_size}
{data_item_names}
<pie_size> - a numeric value in PAGEUNITS for the circle radius;
<center_x> <center_y> - the center location of the circle;
<number_of_data_items> - the <number_of_items> used in the mapping AML;
{textset} {font} {point_size} - define the labelling text--defaults to a 10 point roman font;
{data_item_names} - the labelling text--if used, there should be <number_of_data_items>
separated by spaces.

# Chapter Seven
# Grid-Cell Symbolization in ARC/INFO

For data that falls in the middle of the Continuous--Discrete range and in the middle of the Abrupt--Smooth range (see Figure 1.1), a grid cell representation should be used. This is accomplished by converting a polygon coverage into a grid data layer; if necessary, smoothing the resulting grid; then displaying the grid. POLYGRID is an ARC command that converts a polygon coverage into the GRID cell format.

### Conversion of Polygons to Grids

POLYGRID is an ARC command that converts polygon coverages into GRID cell format. This is best for nominal data, or other polygon data that needs to retain the shape of the original polygon. For ratio data, the GRID operators: IDW, SPLINE and TREND allow conversion of polygons to grids. These operators take a point file (polygon centers can be used) which also have a 'Z' value and calculate intermediate values, in order to populate the resulting grid. IDW is the inverse distance weighted routine; this is quick, but it may smooth the data by eliminating extreme values. SPLINE retains extremes, but requires more processing time. TREND calculates an overall trend for the input data. POLYGRID.aml links the ARC command to ARCPLOT. An ARC version of the ARCPLOT AML is called POLYGRID.

POLYGRID <in_cover> <data_item> <out_grid> {resolution} {p|I|i|s|t}
<in_cover> - the input polygon coverage;
<data_item> - a numeric data field in the attribute table of <in_cover>;
<out_grid> - the resulting grid;
{resolution} - cell resolution--defaults to one hundredth of the smaller axis;
{p|I|i|s|t} - the method used to generate the grid--defaults to p: Polygon, poInt, Idw, Spline, Trend.

### Lookup Tables for Ratio Grids

Because of the impracticality of calculating optimal breakpoints for ratio grid layers, this AML (GRIDLUT.aml) only generates equal interval, or mean and standard deviation classifications. These are based on the grid statistics table, rather than calculated directly from the data set. Lookup tables for nominal grids can be generated with SETMAN.aml.

GRIDLUT <grid> <lookup_name> <classes> {interval|standard_deviation}
<grid> - the name of the grid layer;
<lookup_name> - a name for the resulting lookup table;
<classes> - the number of classes in the resulting lookup table;
{interval|standard_deviation} - the classification method: equal Interval (the default) or Standard deviation.

## Monovariate Symbolization

Because of the limitations of ARC/INFO, which reflect the nature of the raster data model, single grid layers can only be displayed with hue, value or intensity (not used here). A grid cell is generally represented by a single pixel, which prevents use of area fills based on shape, texture, etc.

**Figure 7.1a: RH.aml**
Grided (raster) nominal data displayed with color hue.

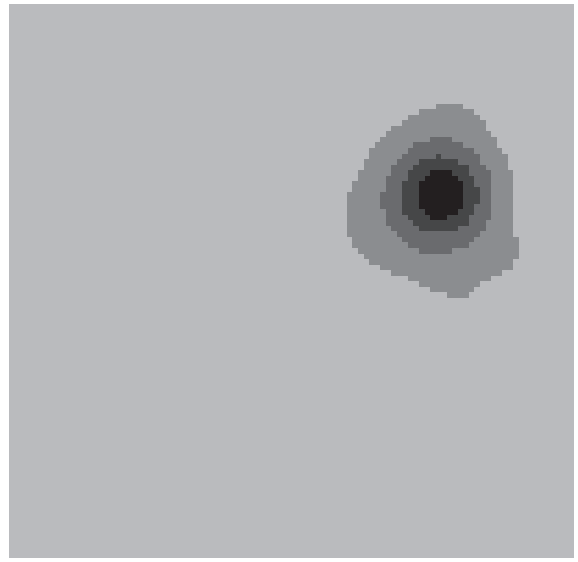


**Figure 7.1b: RV.aml**
Grided ratio data displayed with value. Like the color hue display, legends for this type of data can be displayed with AL.aml.

**Figure 7.1c: RHI.aml**
Nominal data displayed with hue and ordinal data displayed with intensity (compare Figure 5.5a).

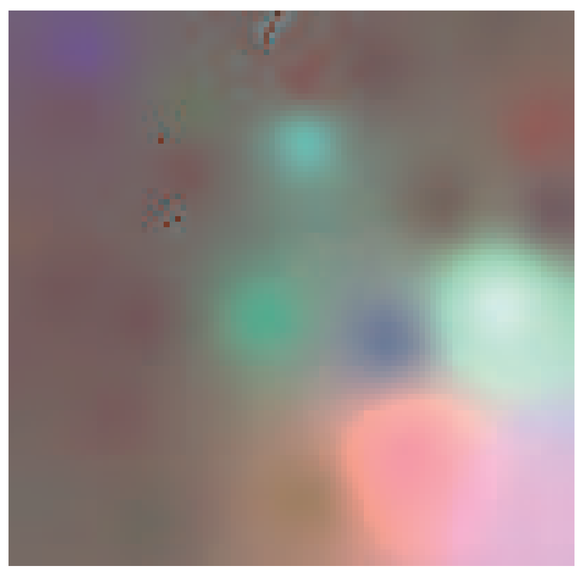


**Figure 7.1d: RRGB.aml**
Three ratio data sets displayed with the RGB (false color) system (compare Figure 5.8a).

### Nominal Data--Hue

Color hue is best used for nominal data. Because of this, RH.aml takes a lookup table generated by SETMAN.aml--the method for creating nominal, grid lookup tables. Legends for hue and value based displays can be generated with AL.aml. See Figure 7.1a.

RH <hue_lookup> {value} {intensity}
<hue_lookup> - specifies HLS hue (from 0 to 300);
{value} {intensity} --default to 50 and 100 (maximum intensity).

### Ordinal to Ratio Data--Value

Because color value is best for ordinal data, this AML requires a lookup table generated by GRIDLUT.aml. For grid layers in which the middle value is important, the lookup table should be generated with the mean and standard deviation method, otherwise equal intervals is probably better. See Figure 7.1b and RV.aml.

RV <value_lookup> {hue} {intensity}
<value_lookup> - specifies HLS lightness (from 0 to 100);
{hue} {intensity} - specify a color to be value shaded--defaults to 0 and 0.

## Bivariate and Multivariate Symbolization

As with single variable displays, bivariate displays are limited to the control of color. This is best accomplished by use of hue and intensity. Three variables can be represented in grid cell format by use of Red, Green and Blue encoding. This is the method used to generate color satellite images, for example.

### Nominal Data, and Ordinal Data--Hue and Intensity

Hue should be used for nominal data (with a lookup table generated by SETMAN.aml). Intensity should be used for ordinal data (with a lookup table generated by GRIDLUT.aml). Legends can be created with CBL.aml. See Figure 7.1c and RHI.aml.

RHI <hue_lookup> <intensity_lookup> {nominal_vat_item} {ordinal_vat_item}
<hue_lookup> - specifies HLS hue (from 0 to 300);
<intensity_lookup> - specifies HLS saturation (from 20 to 100);
{nominal_vat_item} - specifies an item in the hue grid attribute table other than VALUE;
{ordinal_vat_item} - specifies an item in the intensity grid attribute table other than VALUE.

### Three Ordinal to Ratio Data Sets--Red, Green and Blue Symbolization

This AML takes three grids, assigns the first to red, the second to green, and the third to blue. Each grid is linearly stretched from 0 (no color) to 255 (maximum color), and displayed. Legends can be created with RGBL.aml. See Figure 7.1d and RRGB.aml.

RRGB <red_data_grid> <green_data_grid> <blue_data_grid>
<red_data_grid> - a grid that will be displayed in shades of red;
<green_data_grid> - a grid that will be displayed in shades of green;
<blue_data_grid> - a grid that will be displayed in shades of blue.

# Chapter Eight
# Dot Density Symbolization in ARC/INFO

Production of dot density maps, which should be used for discrete, smoothly changing area data, in ARC/INFO requires the use of GRID and TIN for their creation. This is accomplished by converting a polygon coverage to a grid data layer; if necessary, smoothing the resulting grid; then converting the grid into a stepped surface polygon coverage. ARCPLOT then fills the polygons with dots.

### Conversion of Polygons to Dot Density

POLYDOT.aml performs the conversion from polygons based on choropleth boundaries to polygons based on stepped surface boundaries. A lookup table for the input coverage is used to determine the steps. The lookup table can be based on Jenks optimal, or some other classification method. The AML modifies this input lookup table, in order to make it the lookup table for the output coverage. See the discussion of surface generation on page 84.

POLYDOT <lookup_table> <out_cover> {resolution} {i|s|t}
<lookup_table> - specifies a stepped surface classification for <out_cover>;
<out_cover> - the resulting stepped surface coverage;
{resolution} - the grid cell size for creating a surface,
defaults to one hundredth of the smallest axis of the <lookup_table> coverage;
{i|s|t} - the method of creating the surface: Idw, Spline, or Trend.

## Monovariate Symbolization

For single variable dot density maps, the visual variable of texture is used to show changes in data value. By controlling the number of dots placed in an area, data that is discrete, but smoothly changing can be represented.

### Ordinal to Ratio Data--Texture

Texture can be controlled by using SHADESEPARATION with shadetype RANDOMDOTS. Higher values should be represented with a small separation--this will increase the dot density. Because of ARCPLOT's dot placement routine, the shade separation for dots of 0 size must not be greater than 4/100 of an inch. For four classes, SETAUTO.aml can be run with the arguments 0 100 0.5 22. See Figure 8.1a and DD.aml.

DD <lookup_table> {dot_size}
<lookup_table> - specifies dot separation in hundredths of PAGEUNITS;
{dot_size} - the size of individual dots--defaults to 0.02 inches.

### Monovariate Dot Density Legends.

Legends for dot density maps must convey the idea of a continuous range of data values. To accomplish this, DDL.aml generates a column that is subdivided, and labeled, at the surface breakpoints, and filled with dots of the appropriate density. See Figure 8.1b.

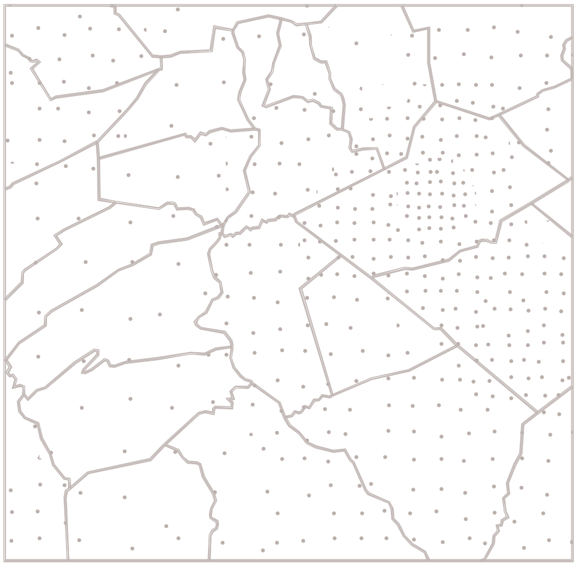


**Figure 8.1a: DD.aml**
Discontinuous and smoothly changing data displayed with dot density. This use of texture is accomplished by converting polygons in stepped surfaces.

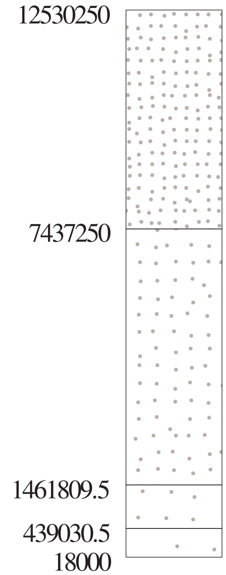



**Figure 8.1b: DDL.aml**
A legend for figure a. The continuous nature of the data is displayed by the unbroken column and labels at transition points.

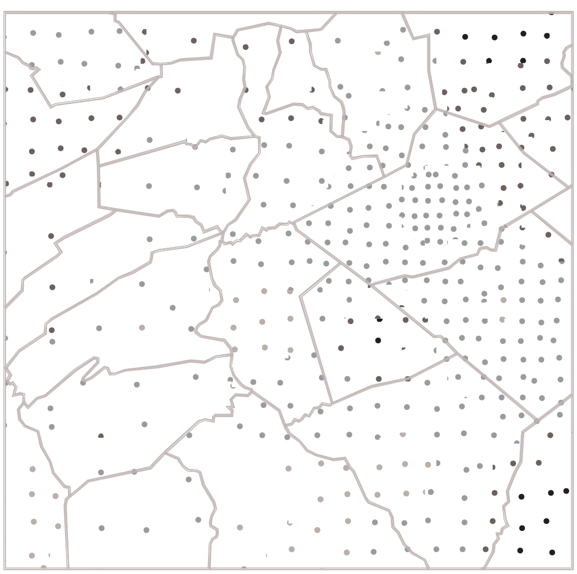


**Figure 8.1c: DDV.aml**
Ratio data displayed with dot density texture and ordinal data displayed with color value. This can be used to show ratio meta-data.

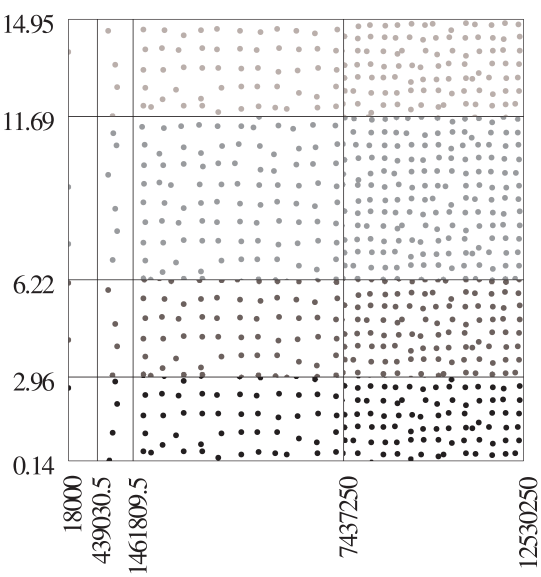



**Figure 8.1d: DDBL.aml**
A legend for figure b. This AML can also be used to generate a legend for dot density and hue maps.

DDL <dot_lookup> <start_x> <start_y> <size_x> <size_y> {dot_size} {textset} {font} {point_size} {float_precisiont}

<dot_lookup> - the lookup table that specifies shade separations;
<start_x> <start_y> - the lower left corner of the legend column;
<size_x> <size_y> - the x and y sizes of the legend;
{dot_size} - the size of dot used in the map--defaults to 0.02 inches;
{textset} {font} {point_size} - define the labelling text--defaults to a 10 point roman font;
{float_precision} - the number of decimal places shown in the legend--defaults to 2.

## Monochrome, Bivariate Symbolization

Monochrome bivariate maps are limited to dot density texture and color value. This can be used to display two data values, or data (texture) and meta-data (value).

### Ratio Data, and Ordinal Data--Texture and Value

Dot density maps that show more than texture must use a dot size larger than that needed for texture alone. This is because specification of halftone shades (such as those used by printers) requires more than one pixel (which can only be on or off). The macro, DDV.aml, adjusts for this by using a larger dot size. See Figure 8.1c.

DDV <dot_lookup> <value_lookup> {dot_size} {hue} {intensity}

<dot_lookup> - specifies dot density in hundredths of PAGEUNITS;
<value_lookup> - specifies HLS lightness (from 10 to 100);
{dot_size} - the size of a dot--defaults to 0.03 inches;
{hue} {intensity} - specify a color to be value shaded.

### Bivariate Dot Density Legends

Like single variable legends, bivariate legends must convey the idea of continuity in dot density, as well as display the character of the second variable. The macro will display hue (nominal) data in separate rows, and value (ordinal) data as continuous. See DDBL.aml and Figure 8.1d.

DDBL <dot_lookup> <lookup_2> <start_x> <start_y> <size_x> <size_y> <h|v> {dot_size} {textset} {font} {text_size} {float_precision} {color_parameter_1} {color_parameter_2}

<dot_lookup> - the lookup table used to specify dot density;
<lookup_2> - the lookup used to specify hue or value;
<start_x> <start_y> - the lower left corner of the legend box;
<size_x> <size_y> - the size of the legend box;
<h|v> - Hue or Value symbolization;
{dot_size} - the size of dot used in the map--defaults to 0.03 inches;
{textset} {font} {text_size} - define the labelling text--defaults to a 10 point roman font;
{float_precision} - the number of decimal places shown--defaults to 2;
{color_parameter_1} - for Hue, value; for Value, hue;
{color_parameter_2} - intensity--defaults are the same as the mapping AMLs.

## Color, Bivariate, and Multivariate Symbolization

Color bivariate maps use texture and hue to display data. Texture should be used for numeric data, and hue should be used for nominal data. More than two variables can be shown in a dot-density map by making use of two or three of the dimensions used to specify color. Hue and intensity allow three variables to be shown and RGB symbolization allow four variable to be shown.

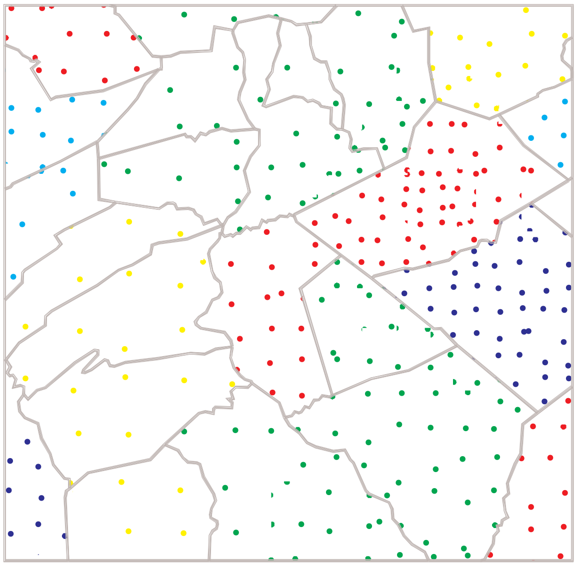


**Figure 8.2a: DDH.aml**
Ratio data displayed with dot density texture and nominal data displayed with color hue.

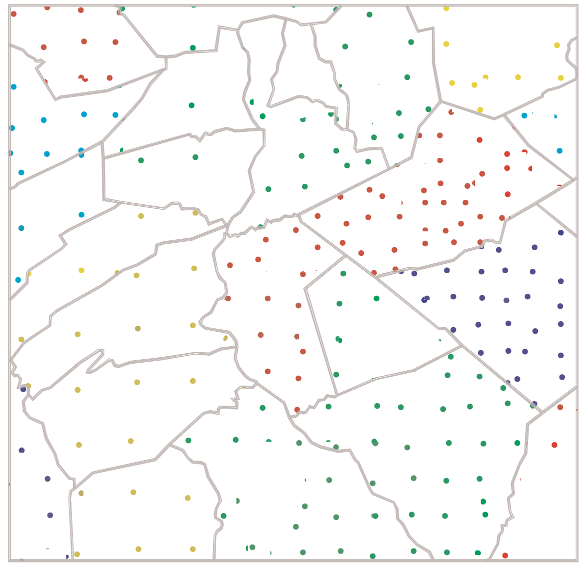


**Figure 8.2b: DDHI.aml**
Ratio data displayed with dot density texture, nominal data displayed with hue and ordinal data (ratio meta-data) displayed with intensity.

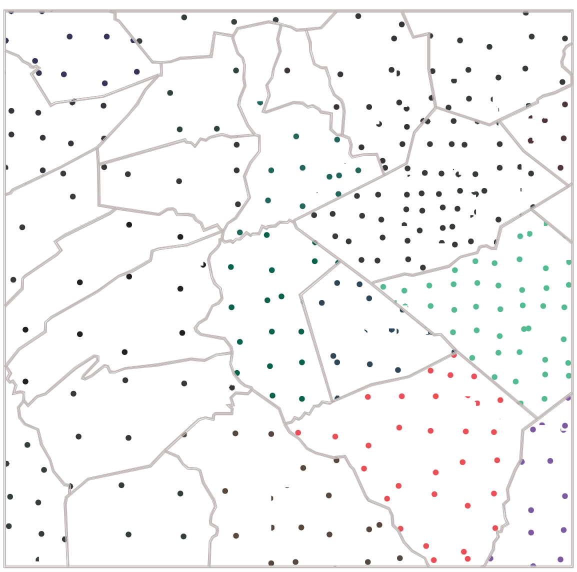


**Figure 8.2c: DDRGB.aml**
Four ratio data sets, one displayed with dot density texture, one with red, one with green and one with blue. Color data values are linearly stretched.

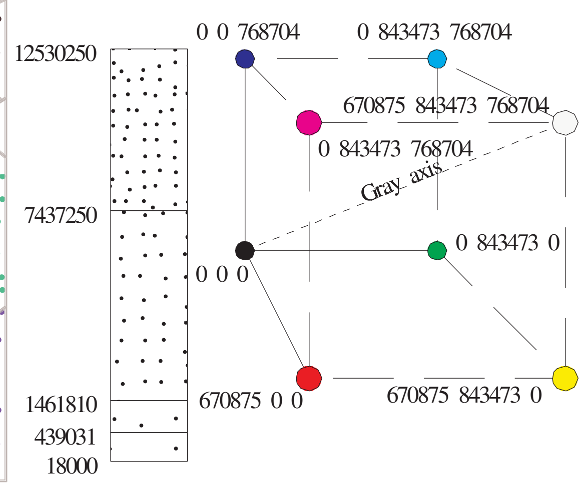



**Figure 8.2d: DDL.aml and RGBL.aml**
A legend for figure c. These two AMLs display all of the data shown, without creating a four dimensional display.

### Ratio Data, and Nominal Data--Texture and Hue

This AML is a variant of the monochrome macro. Like that one, shade separation must not be specified lower than 0.04 inches, or ARCPLOT will hang. See DDH.aml and Figure 8.2a.

DDH <dot_lookup> <hue_lookup> {dot_size} {value} {intensity}
<dot_lookup> - specifies dot density in hundredths of PAGEUNITS;
<hue_lookup> - specifies HLS hue (from 0 to 300);
{value} {intensity} --default to 50 and 100 (maximum intensity}

### Ratio Data, Nominal Data and Ordinal Data--Texture, Hue and Intensity

The use of color intensity can be used to display an ordinal data set, such as meta-data. This can be accomplished by adding intensity to the dot density and hue AML. See DDHI.aml and Figure 8.2b.

DDHI <dot_lookup> <hue_lookup> <intensity_lookup> {dot_size}
<dot_lookup> - specifies shade separation in hundredths of PAGEUNITS;
<hue_lookup> - specifies HLS hue (from 0 to 300);
<intensity_lookup> - specifies HLS saturation (from 20 to 100);
{dot_size} - specifies the size of a dot--defaults to 0.03 inches.

### Three Ordinal to Ratio Data Sets--Red, Green and Blue Symbolization

Three numeric data sets can be displayed by assigning each to one of the color primaries (red, green and blue). In order to ensure that all data values are represented, DDRGB.aml linearly stretches each of the color data sets, so that each value ranges from 0 to 250, or 0 to 100 for the CMY color model (this may show changes at the lower ranges of the data better than RGB). See Figures 8.2c and d.

DDRGB <dot_lookup> <red_item> <green_item> <blue_item> {r|c} {dot_size}
<dot_lookup> - specifies shade separation in hundredths of PAGEUNITS;
<red_item> <green_item> <blue_item> - numeric data fields in the coverage referred to by <dot_lookup>
{r|c} - selects the Rgb (the default) or the Cmy color model;
{dot_size} - specifies the size of a dot--defaults to 0.03 inches.

# Chapter Nine
# Isopleth and Fishnet Symbolization in ARC/INFO

Isopleth maps, or the three-dimensional equivalent, fishnet displays, should be used for data that are continuous and smoothly changing. Both of these can be displayed within ARCPLOT, but first a surface must be generated from a coverage. The result of this conversion, or any grid layer, can be turned into a isoline coverage and displayed as arcs, or displayed as a fishnet. All of the surface display routines can be used for both grids (lattices) and TINs (Triangulated Irregular Networks).

### Polygon to Isoline Conversion

Contours can be used to show continuous, constantly changing data, but if the data are summarized to polygons, isolines must be generated before they can be displayed. POLYCONT.aml accomplishes this. See the discussion of surface calculation on page 84.

POLYCONT <in_cover> <data_item> <out_cover> {contour_interval} {resolution} {i|s|t}

<in_cover> - the coverage to be converted;
<data_item> - a numeric data field in <in_cover> that specifies ‘Z’ values;
<out_cover> - the output line coverage;
{contour interval} - the contour interval of the output coverage--default is 1/10 of the range;
{resolution} - resolution of the surface grid--defaults is 1/100 of the small axis in <in_cover>;
{i|s|t} -surface calculation method: Idw (default), Spline, or Trend surface.

### Polygon to Surface Conversion

The generation of a surface can be accomplished by creating a grid or a TIN. Creation of grids is discussed on page 84. In general, grids may be the better method, unless full use is made of the TIN surface generation capabilities. A simple TIN can be created from ARCPLOT with POLYTIN.aml.

POLYTIN <in_cover> <data_item> <out_tin> {polys|labels}

<in_cover> - the coverage to be converted;
<data_item> - a numeric item in <in_cover> that specifies ‘Z’ values;
<out_tin> - the output surface;
{polys|labels} - whether polygon bounds or label points should be used as the Z reference.

## Monovariate Symbolization

The two primary methods of surface display, isolines and fishnets, display a single variable by control of location. This should only be used for interval or ratio data.

### Ratio Data--Isoline Location

Isoline coverages are arc coverages that can be labeled with a surface height (although elevation is a common use, it is not the only use of isolines and fishnets). SISO.aml draws arcs with the current line type and labels the arcs. Because the current line type is used multivariate maps can be created by changing line hue, value, shape, etc. See Figure 9.1a.


**Figure 9.1a: SISO.aml**
A continuous, smoothly changing ratio surface displayed with isolines. The AML attempts to place contour labels along the lines.

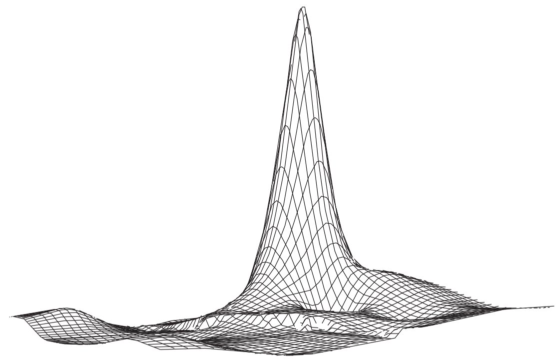


**Figure 9.1b: SF.aml**
Ratio area data displayed with a fishnet. This display shows the default options of SURFACEDEFAULTS and the AML.

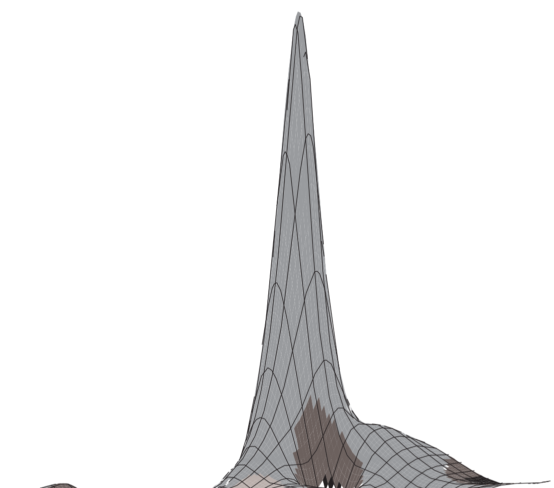


**Figure 9.1c: SFV.aml**
Value shaded surface. This can be used for hypsometric tinting, as well as the display of meta-data.

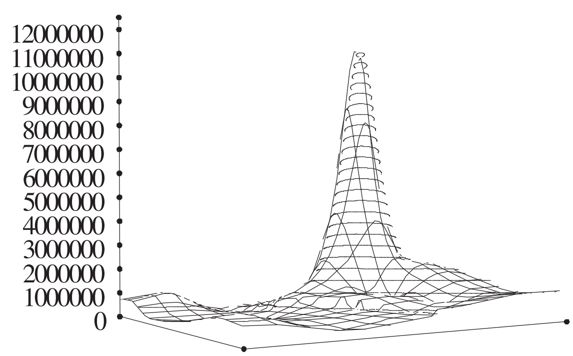



**Figure 9.1d: SFL.aml**
A legend for the surface in figure c. Because of the comparative abstractness of isolines, legends for both fishnets and isolines should show a labeled surface.

SISO <cover> {contour_item} {textset} {font} {point_size} {float_precision}
<cover> - a coverage generated by POLYCONT;
{contour_item} - defaults to the item generated by POLYCONT (contour);
{textset} {font} {point_size} - define labelling text--defaults to a 10 point roman font;
{float_precision} - the number of decimal places shown in contour labels--defaults to 0.

### Ratio Data--Fishnet Height

The macro, SF.aml, simply automates the default setup of a surface display, calculates a better 'Z' expansion value than ARCPLOT, and displays the surface. The expansion value relates the data values used to describe X and Y locations with the value used to describe the height at a location. For example, the default view for a surface could appear absolutely flat, if the X and Y coordinates have a much larger range than the Z values. See Figure 9.1b.

SF <surface> {zfactor} {lattice|tin} {resolution}
<surface> - a grid layer or a TIN;
{zfactor} - 'Z' expansion factor--defaults to range_of_z / the larger of x_range and y_range;
{lattice|tin} - the type of <surface>--defaults to lattice (grid);
{resolution} - the density of fishnet lines--defaults to the SURFACEDEFAULTS value.

### Isoline and Fishnet Legends

Isoline and fishnet legends can be generated by specifying a small area for the display of a fishnet that has isolines drawn on it, and labeled. This is most beneficial for isoline maps, which can be greatly enhanced by the inclusion of the less abstract fishnet as the legend. See SFL.aml and Figure 9.1d.

SFL <surface> <x_start> <y_start> <x_size> <y_size> <contour_interval> {zscale} {lattice|tin} {textset} {font} {point} {precision}
<surface> - a grid layer or a TIN--not an isoline (arc) coverage;
<x_start> <y_start> - the lower left corner of the legend;
<x_size> <y_size> - the size of the legend drawing area;
<contour_interval> - the separation distance in Z of isolines;
{zscale} - a 'Z' expansion factor;
{lattice|tin} - the type of <surface>--defaults to lattice;
{textset} {font} {point} - define the labelling text--defaults to a 10 point roman font;
{precision} - the number of decimal places to be shown in the legend--defaults to 2.

## Bivariate Symbolization

Monochrome bivariate isoline maps can be generated by changing the line shape, texture, size, value, etc of the current line and using the monovariate line AML to draw a second isoline coverage. Monochrome, bivariate fishnet maps can be generated by value shading a surface. Like monochrome, color bivariate isolines can be created by changing the current line's color and drawing a second isoline coverage. Color, bivariate fishnets can be generated by hue shading a surface.

### Ratio Data and Ordinal Data--Surface Shaded with Value

Both polygon coverages and grid layers can be used to value shade a surface. In either case, the data should be ordinal. The ARCPLOT draping routine seems to work better with grids than with polygons. See SFV.aml and Figure 9.1c.

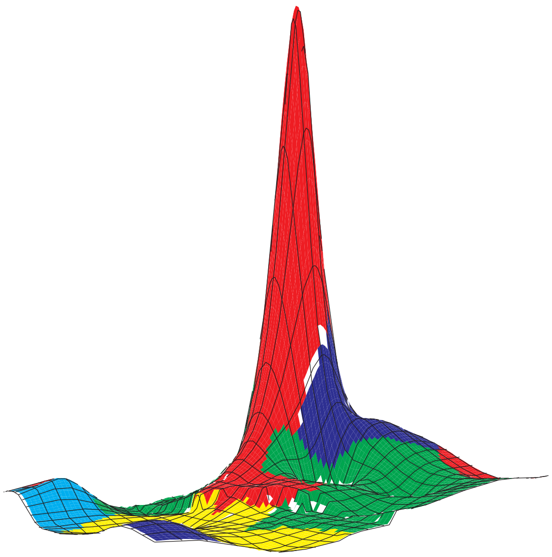


**Figure 9.2a: SFH.aml**
Continuous, smoothly changing ratio data displayed as a surface. The surface is hue shaded with a nominal data set (the AML supports polygons and grids).

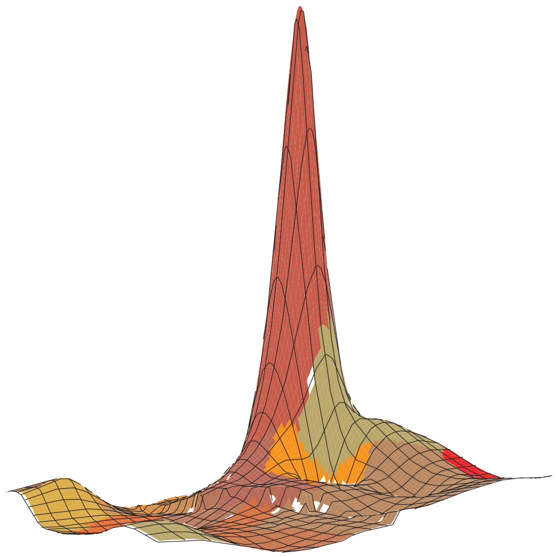


**Figure 9.2b: SFHI.aml**
Hue and intensity shaded fishnet surface. This AML can use either a polygon coverage, or two grids; this image is polygon based (compare Figure 5.5a).

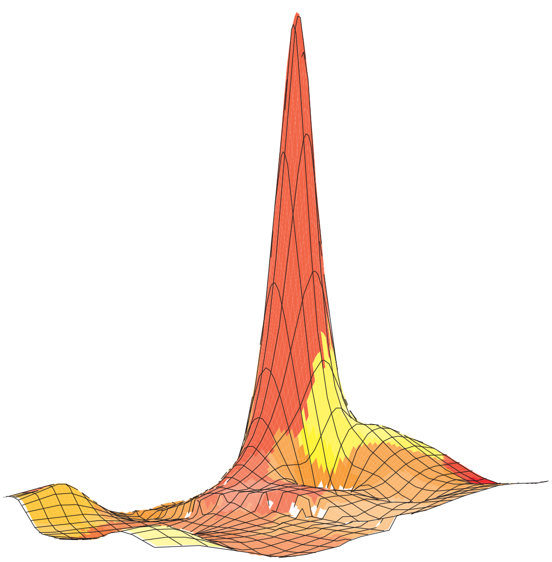


**Figure 9.2c: SFHI.aml**
Hue and intensity shaded fishnet. This is the same AML as figure b, but this figure is based on two grids (compare Figure 7.1c).

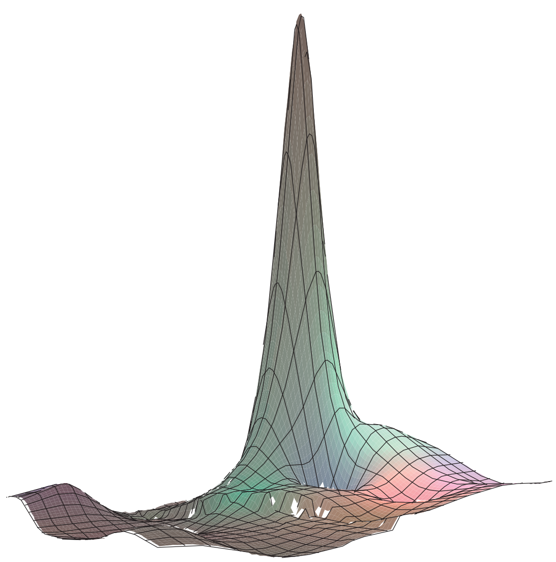


**Figure 9.2d: SFRGB.aml**
A red, green and blue shaded fishnet. This AML only works with grids; polygon data must be converted prior to display with this AML.

SFV <surface> <value_lookup> {zfactor} {lattice|tin} {resolution} {hue} {intensity}
<surface> - a grid layer of a TIN;
<value_lookup> - specifies HLS lightness (0 to 100) for a polygon coverage or a grid;
{zfactor} - 'Z' expansion factor--defaults to z-range / the larger of x-range and y-range;
{lattice|tin} - the type of <surface>--defaults to lattice;
{resolution} - the density of fishnet lines--defaults to the smaller of x-range and y-range / 25;
{hue} {intensity} - define a color to be value shaded.

### Ratio Data and Nominal Data--Surface Shaded with Hue

A prime example of the appropriate use of hue shaded surfaces is the display of land use/land cover data over a digital terrain model. Although this is a common use, hue can be used for any nominal data (this AML supports both choropleth and grid cell data). The surface can be any continuous, smoothly changing ratio data set. See SFH.aml and Figure 9.2a.

SFH <surface> <hue_lookup> {zfactor} {lattice|tin} {resolution} {value} {intensity}
<surface> - a grid layer or TIN;
<hue_lookup> - specifies HLS hue (from 0 to 300);
{zfactor} - 'Z' expansion factor--defaults to z-range / the larger of x-range and y-range;
{lattice|tin} - the type of <surface>--defaults to lattice;
{resolution} - the density of fishnet lines--defaults to the smaller of x-range and y-range / 25;
{value} {intensity} - defaults to 50 and 100 (maximum intensity);.

## Multivariate Symbolization

Multivariate isoline maps can be created by display of multiple isoline coverages, with each coverage drawn with a different current line setting. These difference should reflect either the differences in the coverages (nominal differences by differences in hue or shape, for example), or differences in importance (the main data can be drawn last and in a large, black line--with less important data drawn in smaller lines or grays, for example). For fishnets, multivariate maps can be created by use of hue and intensity, or by RGB encoding.

### Ratio Data, Nominal Data and Ordinal Data--Surface with Hue and Intensity

Surfaces can be shaded with hue and intensity to show nominal distinctions with meta-data, or another ordinal data set. This AML can display hue and intensity from a polygon coverage, or from two grid layers. See SFHI.aml and Figures 9.2b and c.

SFHI <surface> <hue_lookup> <intensity_lookup> {zfactor} {lattice|tin} {resolution}
{nominal_vat_item} {ordinal_vat_item}
<surface> - a grid layer or TIN;
<hue_lookup> - specifies HLS hue (from 0 to 300);
<intensity_lookup> - specifies HLS saturation (from 20 to 100);
{zfactor} - 'Z' expansion factor--defaults to z-range / the larger of x-range and y-range;
{lattice|tin} - the type of <surface>--defaults to lattice;
{resolution} - the density of fishnet lines--defaults to the smaller of x-range and y-range / 25;
{nominal_vat_item} - specifies an item in the hue grid attribute table other than VALUE;
{ordinal_vat_item} - specifies an item in the intensity grid attribute table other than VALUE.

### Four Ratio Data Sets--Surface Shaded with Red, Green and Blue

Unlike the other surface draping AMLs, SFRGB.aml only drapes grid layers, which are linearly stretched. If polygons need to be shown with this method, use the polygon to grid conversion AML discussed on page 84. See SFRGB.aml and Figure 9.2d.

SFRGB <surface> <red_grid> <green_grid> <blue_grid> {zfactor} {lattice|tin} {mesh_resolution}
<surface> - a grid layer or TIN;
<red_grid> <green_grid> <blue_grid> - three numeric grid layers;
{zfactor} - 'Z' expansion factor--defaults to z-range / the larger of x-range and y-range;
{lattice|tin} - the type of <surface>--defaults to lattice;
{mesh_resolution} - the density of fishnet lines--defaults to the smaller of x-range and y-range divided by 25 (half of the default resolution);

# Chapter Ten
# Thematic Mapping in ARC/INFO

Because of the great flexibility of ARC/INFO, a wide variety of maps can be generated within the system. This guide's focus has been on the design of environmental maps. These maps include maps of point data (Toxic Release Inventory sites, for example), linear data (STORET information), and area data (demographic/Census data). With the GRID and TIN modules, ARC/INFO has the ability to show area data with grid cell, dot density, isoline and fishnets, as well as the choropleth and graduated circles methods that are part of ARCPLOT.

## Summary of Presented Methods

The methods presented in this guide have been based on the theoretical construct, 'visual variables.' Use of this construct allows the specification of rules for matching data to symbolization types. Each of these rules is indicated in the subsection headings in Chapters Three through Nine. For point and line data, select the number of variables that need to be represented with one symbol, then match the appropriate data levels to the symbolization method by following the headings (Ratio Data--Size, etc.). For area data, the character of the data must be considered. Data that are continuous and changes abruptly at boundaries should be displayed with choropleth maps (for example, sales tax rates). Data that are continuous and smoothly changing should be displayed with isoline of fishnet maps (for example, elevation). Data that are discrete and smoothly changing should be displayed with dot density maps (for example, population density). Data that are discrete and abruptly changing should be displayed with graduated symbols (for example, TRI point source emissions).

## Areas of Possible Continued Research

As a simple continuation of these AMLs, there are many possibilities of combinations of data types (particularly for area data) and visual variables that are not presented here. This expanded set of AMLs, could be developed on an as needed basis.

A potentially valuable addition to the ARC/INFO repertoire is the development of animation. Time can be used to show temporal changes in data, for touring a data set, or for display of some other variable. These abilities can be used to both impress (as in visual communication) and highlight otherwise obscure relations (as in visual thinking).

## Acknowledgments

I would like to thank my colleagues at the National GIS Program: Tommy Dewald, Jeff Booth, Dave Wolf, and Ed Partington for seeing the need for an ARC/INFO cartographic guideline and reviewing the text. Dave Rejeski originally foresaw the need for a guideline, from which this has developed.

At the Pennsylvania State University, Deryck Holdsworth has provided valuable guidance, for the version of this guideline that is my master's thesis. Anthony Williams also provided helpful comments.

# References

Arnheim, Rudolf. 1969. *Visual Thinking*. Berkeley, CA: University of California Press.

Aspaas, Hellen R. and Stephen J. Lavin. 1989. "Legend Designs for Unclassed, Bivariate, Choropleth Maps." *The American Cartographer*. 16(4). 257-268.

Bertin, Jacques. 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. William J. Berg, trans. Madison, WI: University of Wisconsin Press.

Brewer, Cynthia A. 1989. "The Development of Process-Printed Munsell Charts for Selecting Map Colors." *The American Cartographer*. 16(4). 269-278.

Carstensen, Laurence W., Jr. 1982. "A Continuous Shading Scheme for Two-variable mapping." *Cartographica*. 19(3). 53-70.

-------1986. "Hypothesis Testing Using Univariate and Bivariate Choropleth Maps." *The American Cartographer*. 13(3). 231-251.

Cort, J. Wilmont; Clinton M. Rowe and William D Philpot. 1985. "Small-Scale Climate Maps: A Sensitivity Analysis of Some Common Assumptions Associated with Grid Point Interpolation and Contouring." *The American Cartographer*. 12(1). 5-16.

Coulson, Michael R. C. 1987. "In the Matter of Class Intervals for Choropleth Maps: with Particular Reference to the Work of George F. Jenks." in Patricia Gilmartin (ed.) Studies in Cartography: A Festschrift in Honor of George F. Jenks. *Cartographica*. 24(2): Monograph 37. 16-39.

DeLucia, Alan A. and Donald W. Hiller. 1982. "Natural Legend Design for Thematic Maps." *The Cartographic Journal*. 19(1). 46-52.

Dent, Borden. 1985. *Principles of Thematic Map Design*. Reading, MA: Addison-Wesley.

DiBiase, David. 1990. "Scientific Visualization in the Earth Sciences." *Bulletin of the College of Earth and Mineral Sciences*. 59(2). 13-18. University Park, PA: The Pennsylvania State University.

DiBiase, David; John Krygier; Catherine Reeves; Alan M. MacEachren and Alan Brenner. 1991. *Elementary Approaches to Cartographic Animation*. University Park, PA: video publication by the Deasy GeoGraphics Laboratory, Department of Geography, The Pennsylvania State University.

Dobson, Michael W. 1983. "Visual Information Processing and Cartographic Communication: The Utility of Redundant Stimulus Dimensions." in D. R. F. Taylor (ed.) *Graphic Communication and Design in Contemporary Cartography*. New York: John Wiley and Sons.

Environmental Protection Agency. 1991. *Locational Data Policy Implementation Guide*. Washington, D. C.

-------1992. *Spatial Data Management Plan*. Washington, D. C.

Environmental Systems Research Institute. 1992. *Understanding GIS: The ARC/INFO Method*. Redlands, CA.

-------1992. *ARC Command References*. 2 vols.

-------1992. *ARCPLOT Command References*. 2 vols.

-------1992. *AML User's Guide*.

-------1992. *GRID Command References*. 2 vols.

-------1992. *GRID Class Manual*.

-------1992. *Map Display and Query*.

Edwards, Betty. 1989. *Drawing on the Right Side of the Brain*. Los Angeles: Jeremy P. Tarcher, Inc.

Eyton, J. Ronald. 1984. "Complementary-Color, Two Variable Maps." *The Annals of the Association of American Geographers*. 74(3) 477-490.

Federal Geographic Data Committee. 1992. *Manual of Federal Geographic Data Products*. Washington, D. C.

Gershmehl, Philip J. 1990. "Choosing Tools: Nine Metaphors of Four-Dimensional Cartography." *Cartographic Perspectives*. Number 5, Spring 1990. 3-17.

Johnson, Branden B. and Vincent T. Covello (eds.). 1987. *The Social and Cultural Construction of Risk: Essays on Risk Selection and Perception*. Boston: D. Reidel Pub. Co.

Jones, Kelvyn. 1991. "What's Hiding Behind Statistical Maps." *Society of University Cartographers Bulletin*. 24(1). 23-30.

Keats, J. S. 1982. *Understanding Maps*. New York: Halsted Press.

Knoble, Herman D. 1990. *Reality of REAL Arithmetic's Really Wrong Results*. Transcript Share 75, FORTRAN Project, Session A722. University Park, PA: Pennsylvania State University.

Leonard, Jeff J. and Barbara P. Buttenfield. 1989. "An Equal-Value Gray Scale for Laser Printer Mapping." *The American Cartographer*. 16(2). 97-107.

MacEachren, Alan M. 1992. *Primer on Cartographic Symbolization and Design for GIS*. unpublished U. S. Environmental Protection Agency document.

MacEachren, Alan M. and John V. Davidson. 1987. "Continuous Geographic Surfaces: Sample Resolution, Intermediate Value Estimation Accuracy, and Isometric Mapping." *The American Cartographer*. 14(4). 299-320.

MacEachren, Alan M. and David DiBiase. 1991. "Animated Maps of Aggregated Data: Conceptual and Practical Problems." *Cartography and Geographic Information Systems*. 18(4). 221-229.

MacEachren, Alan M. and John H. Ganter. 1990. "A Pattern Identification Approach to Cartographic Visualization." *Cartographica*. 27(2). 64-81.

McLeary, George F. 1981. *How to Design an Effective Graphics Presentation*. Harvard Library of Computer Graphics, 1981 Mapping Collections. 17. 15-64.

-------1983. "An Effective Graphic 'Vocabulary'." *IEEE CG&A*. March 46-53.

Monmonier, Mark. 1975. "Class Intervals to Enhance the Visual Correlation of Choropleth Maps." *The Canadian Geographer*. 12.

-------1976. "Modification of a Choropleth Technique to Communicate Information." *International Yearbook of Cartography*. 160-164.

-------1990. "Strategies for the Visualization of Geographic Time-Series Data." *Cartographica*. 27(1). 30-45.

-------1991. *How to Lie With Maps*. Chicago: University of Chicago Press.

Monmonier, Mark and Branden B. Johnson. 1990. *Design Guide for Environmental Maps*. Trenton: New Jersey Department of Environmental Protection, Division of Science and Research.

Morgan, M. Granger and Max Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge: Cambridge University Press.

Morison, Joel L. 1974. "A Theoretical Framework for Cartographic Generalization with the Emphasis on the Process of Symbolization." *International Yearbook of Cartography*. 14. 115-127.

Muller, Jean-Claude. 1979. “Perception of Continuously Shaded Maps.” *The Annals of the Association of American Geographers*. 69(2). 240-249.

Olson, Judy M. 1981. “Spectrally Encoded Two-Variable Maps.” *The Annals of the Association of American Geographers*. 71(2). 259-276.

Patton, Jeffery. C. and Paul. V. Crawford. 1978. “The Perception of Hypsometric Colours.” *The Cartographic Journal*. 15. 115-127.

Rejeski, David. 1991. *Confessions of an Eco-risk Manager: Thought Piece for the Workshop on Ecological Risk*. U.S. Environmental Protection Agency.

Rejeski, David and Jacques Kapuscinski. 1990. *Risk Modeling with Geographic Information Systems; Approaches and Issues*. U.S. Environmental Protection Agency.

Simpson, J. A. and E. S. C. Weiner, eds. 1989. *The Oxford English Dictionary*, Volume XVIII. Oxford: Oxford University Press.

Star, Susan L. 1985. “Scientific Work and Uncertainty.” *Social Studies of Science*. 15. 391-427.

Toffler, Alvin. 1990. *PowerShift*. New York: Bantam Books.

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

-------1990. *Envisioning Information*. Cheshire, CT: Graphics Press.

Upton, Barbara L. and John Upton. 1989. *Photography: Adapted from the Life Library of Photography*. Glenview, IL: Scott, Foresman and Company.

Weible. Robert and Barbara P. Buttenfield. 1988. "Map Design for Geographic Information Systems." *GIS/LIS '88 Proceedings: Assessing the World*. Volume 1. 350-9.

Woodward, David. 1991. Personnal Communication to David DiBiase. Referenced in MacEachren, 1992.

# Subject Index

## V



## W

# ARC/INFO Commands and AML Programs

## T



## U

# Author Index