LA-UR- 97-210
CONF-970493--1

VISUALIZING DATA FOR ENVIRONMENTAL ANALYSIS

Jan Benson
Facility for Information Management, Analysis, and Display
Los Alamos National Laboratory
PO Box 1663 MS D452
Los Alamos, NM 87545

ABSTRACT

The Environmental Restoration Project at Los Alamos National Laboratory (LANL) has over 11,000 sampling locations in a 44 square mile area. The sample analyses contain raw analytical chemistry values for over 2300 analytes and compounds used to define and remediate contaminated areas at LANL. The data consist of 2.5 million records in an Oracle database. Maps are often used to visualize the data. Problems arise when a client specifies a particular kind of map without fully understanding the limitations of the data or the map. The ability of maps to convey information is dependent on many factors, though all maps are data dependent. The quantity, spatial distribution, and numerical range of the data can limit use with certain kinds of maps.

To address these issues and educate our clients, several types of statistical maps (e.g., choropleth, isarithm, and graduated symbol such as bubble and spike) used for environmental analysis were chosen to show the advantages, disadvantages, and data limitations of each. By examining both the complexity of the analytical data and the limitations of the map type, it is possible to consider how reality has been transformed through the map, and if that transformation accurately conveys the information present.

INTRODUCTION

Cartographers and GIS professionals are constantly striving to create the optimal map for the task at hand. However, interactions and work with colleagues from diverse disciplines make the task at hand both challenging and ambiguous. Maps are designed to answer questions. Where is the phenomenon of interest? How much of the phenomenon is there? What are the limits of the phenomenon? Is there regularity in the distribution of the phenomenon?

How these questions are answered are reflected in map design and visualization. In the pas, most maps were created as a presentation device to show an abstract view of some portion of the world (MacEachren, 1994). In 1990 David DiBiase suggested a graphic model of map and graphic use in scientific research. Visual thinking and visual communication are two ends of a continuum along which lead to differences in design and symbolization. Visual thinking seeks to prompt insight, reveal patterns in data, and highlight anomalies through exploration and confirmation. Through synthesis and

presentation visual communication seeks to make a point by communicating what is known. This model lends itself well to an orderly process of research and decision-making. Yet, what happens when diverse professional researchers and decision-makers with different view points and agendas interact on these processes? Exploration, confirmation, and synthesis happen within the same map. Maps designed to make a point are perceived as political and suspect. Instead, the data must state the issue. Common ground is reached by presenting unaltered data as clearly as possible. This process is a Pandora's box as few people have formal training in reading and understanding maps. This paper seeks to address some of the conflicting problems in map design and visualization of environmental data.

DATA

Historically, traditional cartographers were given data and requested to make a map. With the onset of GIS technology, cartographers started building and maintaining data in digital format. Graphic data, digital descriptions of map features including coordinates and their relating symbols, and non-graphic data, attributes or characteristic qualities and relationships of geographic locations, were stored separately due to their different characteristics, maintenance, and use in different software systems (Antenucci, et al, 1991). In some instances, non-graphic data evolved into large relational database systems sometimes warehoused for communal sharing.

The cartographer is still requested to show the attribute (non-graphic) data on the map (graphic data). However, now the cartographer must understand the data structure of the attribute database as well as the graphic database, and how to relate the two together. How attribute data is defined is critical to how it can be used in visualization. The data model is an abstract representation of the key features of the phenomenon. The abstraction process begins with identifying the spatial dimension and the level of measurement (MacEachren, 1994).

Spatial dimension can be categorized into positional (order 0), linear (order 1), area (order 2), volumetric (order 3), and space-time (order 4). Map symbols can be classed as point (0 dimension), line (one dimension), area (two dimension), and volume (three dimension). However, map scale and intended map use precludes matching symbol dimensions with phenomenon dimensions as seen in plotting a city as a point on a map of the world, but as an area on a state map. Because traditional maps are two dimensional, only points, lines, and areas are used to symbolize data. These form the basis of map symbols.

Data are merely recorded facts or observations. Measurement attempts to structure observations about reality (Dent, 1996). A measurement system can be one of four levels: nominal, ordinal, interval, and ratio. The measurement system of the data is important because it defines what mathematical operations can occur. Likewise, these mathematical operations define what statistical methods and tests are valid as well as how the data can be visualized on a map.

MAPS

Thematic maps may be divided into qualitative or quantitative. Qualitative maps show location characteristics of phenomenon where, quantitative maps show location and quantity of a phenomenon. Map symbolization should be matched with spatial dimension, or geographical phenomenon. Likewise, the map symbol should reflect its distribution form: discrete, sequential, or continuous. Yet, how data is collected and measured may force deviations from traditionally accepted cartographic practices. At LANL much of the environmental restoration data is in the form of occurrences at a point. In an attempt to visualize point data, three different quantitative map types are examined: choropleth, isarithm, and graduated symbol (e.g., bubble and spike).

The International Cartographic Association defines the choropleth technique as "A method of cartographic representation which employs distinctive color or shading applied to areas other than those bounded by isolines. These are usually statistical or administrative areas."* Choropleth maps can be thought as a stepped statistical surface or three dimensional histogram.

Choropleth maps are used to determine an actual value associated with a geographic area, to examine the overall geographic pattern of the mapped variable, and to compare one choropleth map pattern to another (Dent, 1996). Data must be collected from defined enumeration units or administrative political subdivision. Use of continuous data is inappropriate since the geographic phenomenon is not controlled by the unit. Data derived from enumeration units is of two kinds: totals and derived units. It is not generally accepted to map total values when using the choropleth map because the enumeration units can be of unequal area. Figure 1 shows a map deviating from the traditional choropleth map. How data are collected affects how it can be visualized. Another researcher working on a similar project chose to count shrapnel using a similar grid, but instead assigned the values to the lower left corner of each grid cell.

Isarithmic maps represent a planimetric graphic of a three dimensional volume. Isarithmic map constraints include data that are continuous, that data be in the form of a geographical volume, or assumed to be voluminous, and that the cartographer fully understands the distribution being mapped. Figure 2 is an example of such a map. Isarithmic maps show both the absolute magnitude of the scale and information on the surface gradient. All surface heights are related to a single, fixed datum from which all phenomenon points are measured. Isarithmic maps show the geographic phenomenon orthogonal, or as viewed

---

* E. Meynen, ed., *Multilingual Dictionary of Technical Terms in Cartography* (International Cartographic Association, Commission II - Wiesbaden, West Germany: Franz Steiner Verlag, 1973), p. 123.

from above. Trends and distributions can be seen and read if values are color coded. The most important factors in final appearance are the number of isolines and the interval used (Unwin, 1990). Too many lines clutter the map making it difficult to read and obscures underlying features; too few lines provide a poor relief picture. The contour interval chosen depends on scale as well as map use and the nature of the phenomenon. The interval can act as a filter on the phenomenon by masking out values just under the interval value. Lastly, the algorithm used to produce the map may be simple or quite complicated to use. Isarithmic maps work well for the trained eye.

Graduated symbol maps allow the size of the chosen symbol to vary systematically thereby showing relative magnitudes of phenomenon at specific geographic locations. Ordinal, interval, and ratio scaled data are suitable for this map type. Data must occur at a point, or be aggregated to a point. Circles are a popular symbol due to their compact geometric form, visual stability, and ease of scaling. Figure 3 shows a graduated circle (bubble) map. Proportional symbols may represent a range of values, or be proportional to the data such that each value is represented by a symbol. Graduated circle symbol maps are popular with statisticians when the circle area is proportional to the data value. In this form they are perceived as an honest map. Graduated circle symbol maps make a statement and help define trends and patterns in the data.

Graduated circle symbol maps have disadvantages. First, untrained map users may read the size of the circle as the areal extent of the phenomenon. Second, the data range may be too great to show both small and large values on a single map. If the same phenomenon is shown across multiple maps with the same proportional scale, finding small circles may be impossible. Third, readers may also underestimate the size of large circles (Flannery, 1956). Circles often lie on top of one another. Groop and Cole, 1978, found readers make errors of perceptual judgment with cut-out symbols in proportion to the obscured amount of the circle. Transparent circles are perceived as well as non-overlapping circles. Yet, cut-out circles add dimension to the map where transparent circles look flat and uninteresting. Lastly, lay readers may compare the radii instead of the areas of the circle thereby missing the crucial information of the map (Tyner, 1992). The uncertainty of circle estimation has lead to the graduated square symbol because proportional area is more accurately judged than circle areas (Heino, 1995). However, the rectangularity can interfere with map design.

In Figure 4 a line symbol at a point is graduated to show the proportional measurement of a point phenomenon. This is rarely seen in traditional cartography books. Ratio data are scaled to show dispersion, pattern, and relative concentrations above background. The line is a bi-value symbol showing those concentrations above background and up to screening action level (SAL-black), and those concentrations above SAL (gray). Patterns in the data and the underlying base map are easy to see. The design of the map clearly points out the discontinuity of the data. Yet, dense distribution of the data may preclude the use of this map form. Spikes (i.e., lines) may superimpose from overlapping map points and do exhibit foreshortening. Large discrepancies among data points require scaling. If a logarithmic scale is used, it may be difficult for the user to interpret. The user should be sophisticated and know the data well.

ANALYSIS

The maps seen in figures 1-4 were all designed to help researchers better understand their data. The actual visualization of data includes many factors. In Figure 1 point data are seen as an area. The discontinuous nature of the data is lost in the map design. Yet, patterns and trends are easily observable. Though the data is gridded, only parts of the grid are sampled. So, could the data be visualized differently? Isarithms are inappropriate due to the unevenness of the sampled grid. Lines would stop at the grid data boundaries and appear discontinuous on the map. The discontinuous nature of shrapnel distribution also makes the isarithmic map a poor choice. Graduated symbol maps would plot on top of each other and portray an unnatural distribution. Overall, the map conveys the data, and the personal preference of the map user was for a density map.

The isarithm map in Figure 2 appears very cluttered. The density of the contour interval and the base map information adds to this clutter. The original map is color range-graded into ten classes ranging from -2305 gamma (blue) to 4099 gamma (red). The black and white rendition of this map is poor due to the number of ranges and its smaller size. The data consists of 4380 points in an area 3235 feet long and 560 feet wide, which worked well for creating a three-dimensional surface and interpolating isolines. The data could have been put into grid cells, but individual data values would be concealed. Areas with high data values are easily seen as well as any distribution pattern in the data. Less detail would make the map easier to read, however, the end user wanted it all.

The graduated symbol map is one of the more popular forms of analysis map (Figure 3). Even with all the problems inherent in a graduated circle symbol map, they consistently appear in reports and official documents. For the trained user, a glance can say it all. In most cases the researcher is looking for elevated values and patterns in the data. Though the elevation contour lines clutter the map, the distribution and range of the data is clear.

Figure 4 which uses a line as a graduated symbol for sample values provides the user with good base data and a good feel for the distribution of the data. The lines, or spikes, are excellent at depicting the discontinuous nature of Pu-239. The bi-value nature of the symbol allows the viewer to see the extent of a value above background, and to see how far the value is beyond a screening action level. On larger plots the sample id is plotted next to the point symbol allowing researchers to look up the elevated values in the table. One obvious problem with the map is the scale of the symbol. The legend needs to display a range of data values with the associated scaled line symbol to support visualization.

CONCLUSION

There are many other ways to visualize point data. What has been discussed are some of the more common map forms used in the Environmental Restoration project at LANL. Their effectiveness is as dependent on the map user as on the cartographic design. The demands of map users sometimes collide with cartographic practices and design. Yet, this visualization of the data promotes sound management decisions and material for defending decisions and practices.

REFERENCES

Antenucci, J, K. Brown, P. Croswell, M. Kevany, & H. Archer (1991). *Geographic Information Systems A Guide to the Technology*, Chapman & Hall, N.Y., NY

Dent, B. (1996). *Cartography Thematic Map Design.* Wm. Brown, Dubuque, IA.

DiBiase, D. (1990) Visualization in the earth sciences. *Earth and Mineral Sciences*, Bulletin of the College of Earth and Mineral Sciences, PSU 59(2):13-18.

Dunn, R. (1988). Framed rectangle charts or statistical maps with shading: An experiment in graphical perceptions. *American Statistician,* 42(2):123-129.

Flannery, J. (1956) *The Graduated Circle: A Description, Analysis, and Evaluation of a Quantitative Map Symbol*, unpublished Ph.D. dissertation, Department of Geography, University of Wisconsin.

Groop, R. & D. Cole (1978). Overlapping graduated circles: Magnitude estimation and method of portrayal. *Canadian Cartographer,* 15:114-122.

Heino, A. (1995). The presentation of data with graduated symbols. *Cartographica*, 32(1):43-50.

Jenks, G.F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7:186-190.

MacEachren, A. (1994). *Some Truths With Maps*, Association of American Geographers, Washington, D.C.

MacEachren, A. & D. Taylor. (1996). *Visualization in Modern Cartography*, Pergamon/Elsevier, Tarrtown, NY.

Morrison, J. (1971). Method produced errors in isarithmic mapping. Technical Monograph No. CA-5, American Congress on Surveying and Mapping, Washington, D.C.

Tomlin, C. (1990). *Geographic Information Systems and Cartographic Modeling*, Prentice Hall, Englewood Cliffs, NJ.

Tyner, J. (1992). *Introduction to Thematic Cartography*, Prentice Hall, Englewood Cliffs, NJ.

Unwin, D. (1981). *Introductory spatial Analysis*, Methuen, N.Y.,NY.

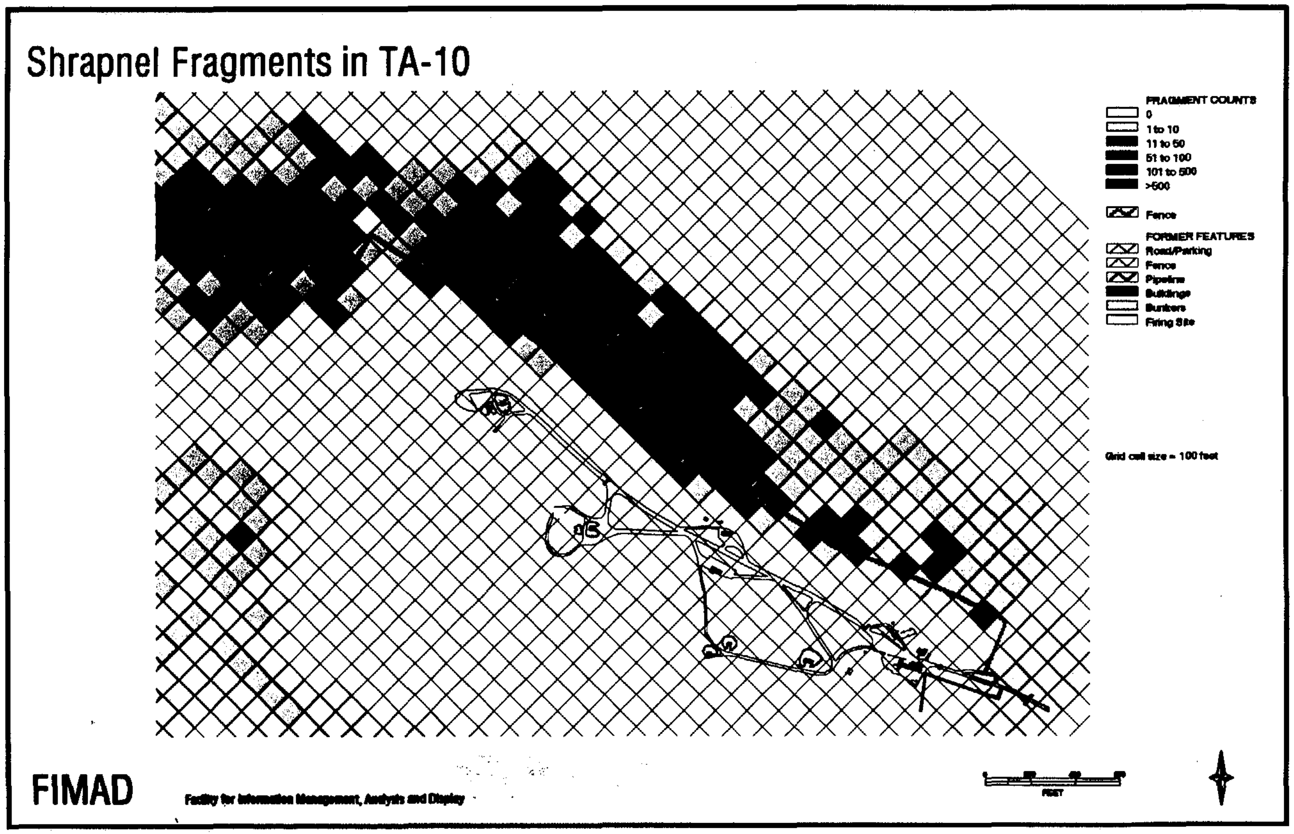

Figure 1. Choropleth (Density) Map

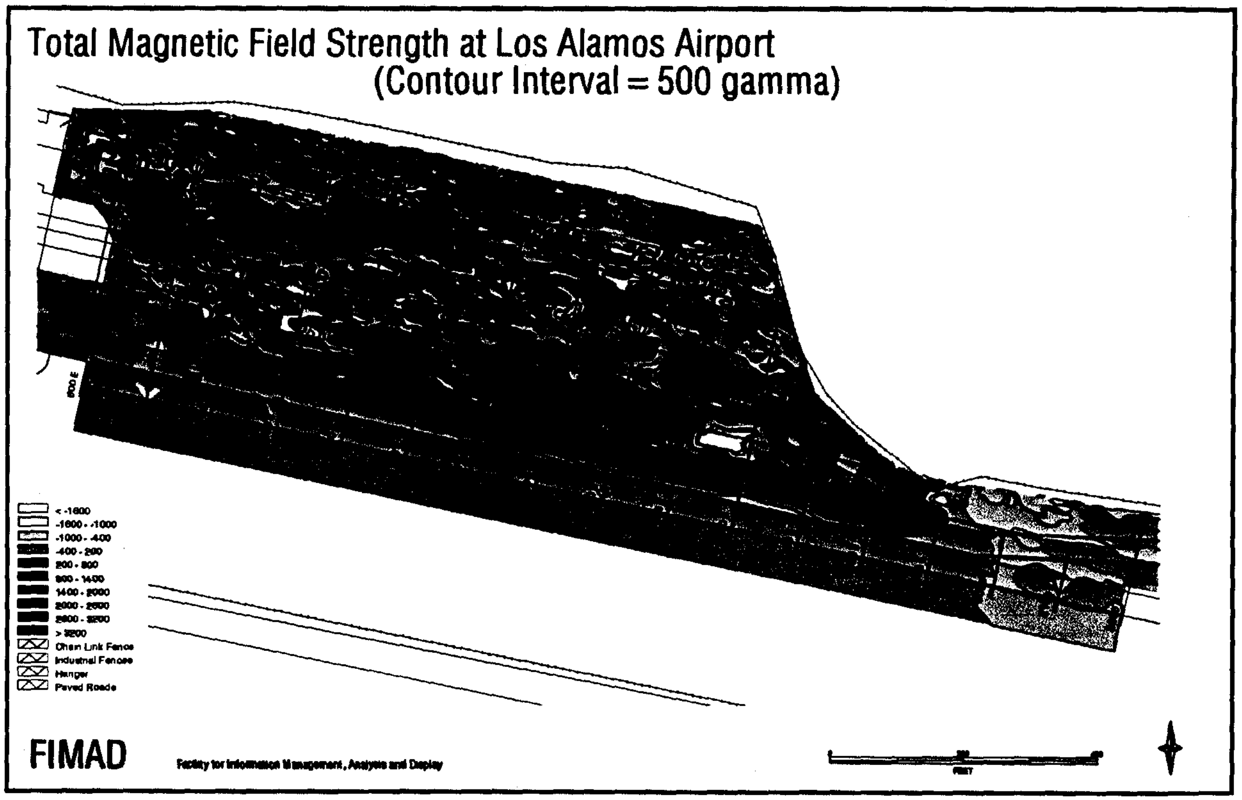

Figure 2. Isarithmic Map

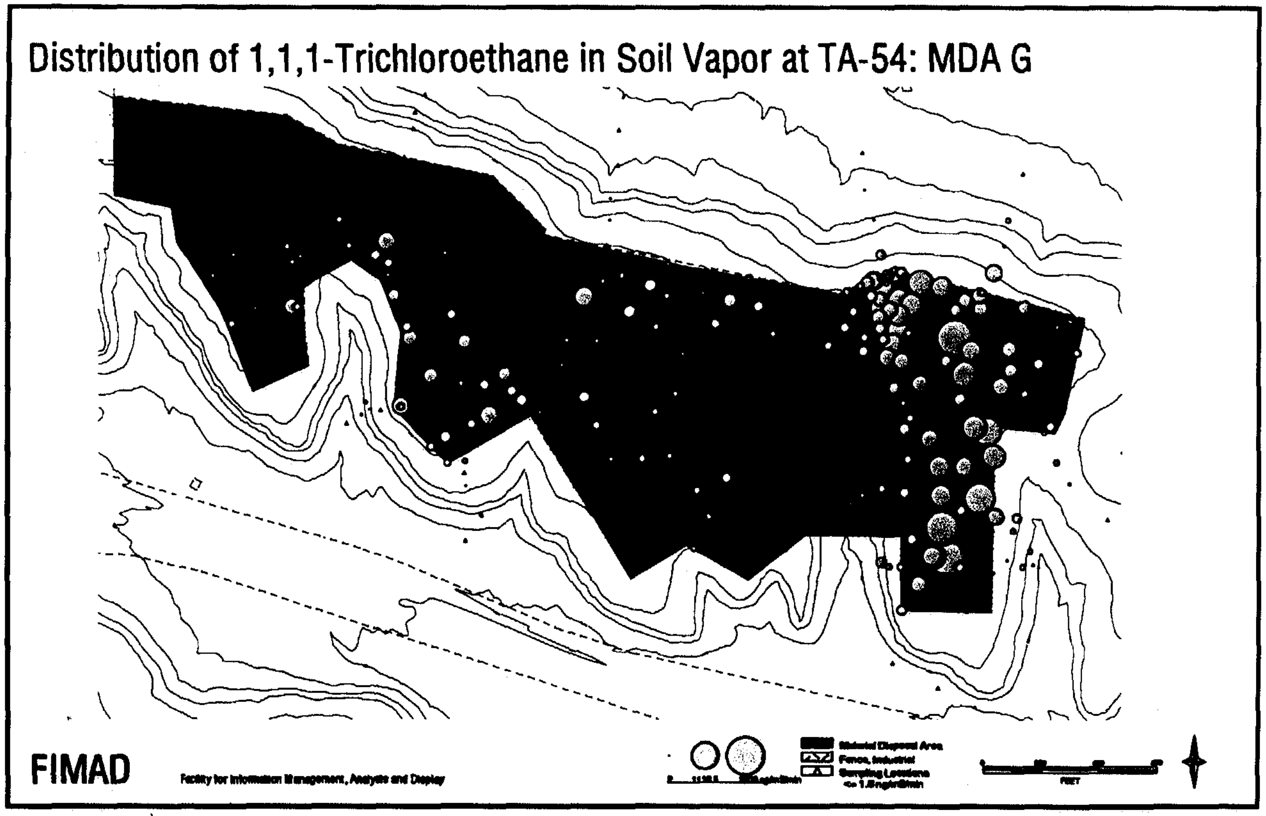

Figure 3. Graduated Symbol (Bubble) Map

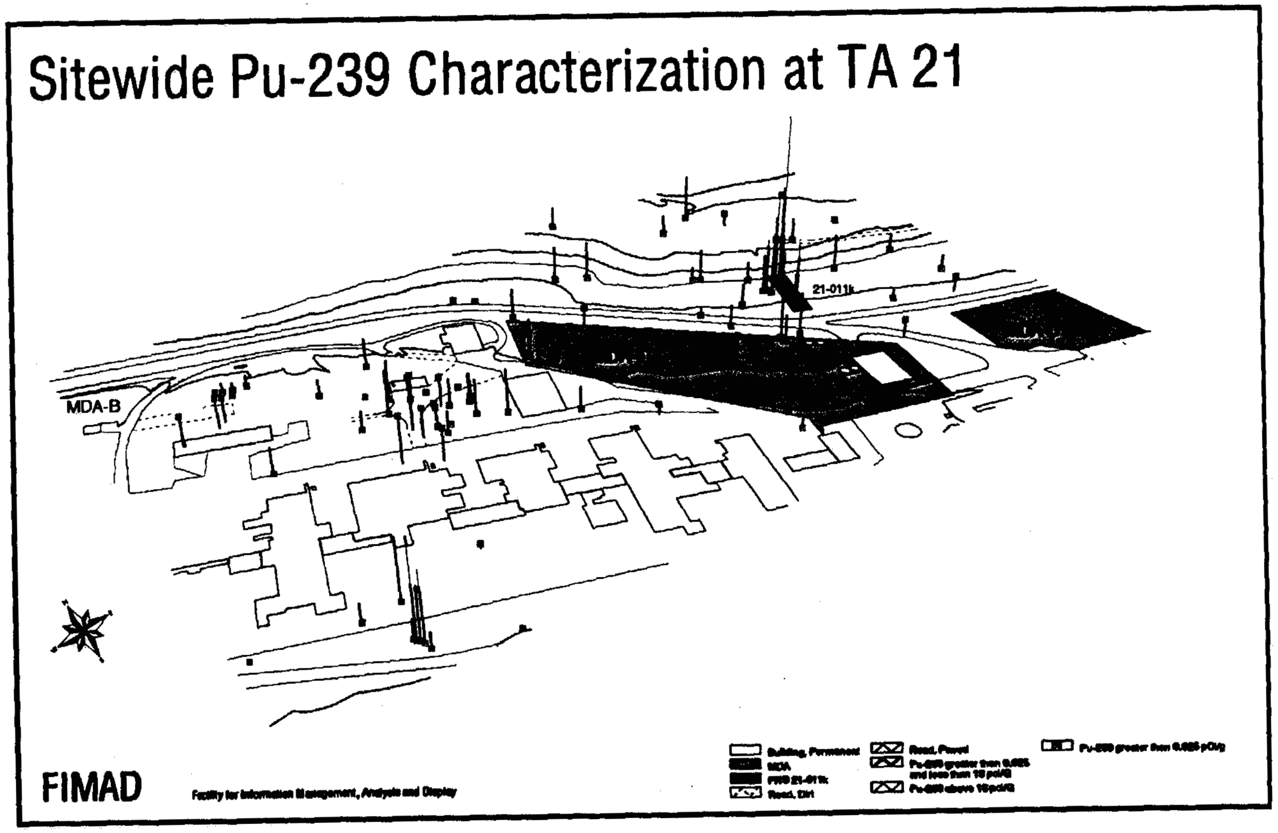

Figure 4. Graduated Symbol (Spike) Map