Visualizing Metadata: Design Principles for Thematic Maps

Joshua Comenetz
Assistant Professor
Department of Geography
University of Florida
P.O. Box 117315
Gainesville, FL 32611
comenetz@geog.ufl.edu

This paper argues that the best way to transmit metadata to thematic map users is through cartography rather than text notes or digital means. Indicators such as reliability diagrams, common on maps derived from air photos or satellite images, are rarely included on thematic maps based on the census or other socioeconomic data sources. These data not only suffer from an array of quality problems but also are widely distributed among the general public in cartographic format. Metadata diagrams for thematic maps based on human variables therefore must be clear and concise, so as to be comprehensible by the non-specialist. Principles of good metadata diagram creation are proposed, with attention to the balance between clarity and space constraints.

Key words: data quality, reliability diagram, thematic map, visualization

METADATA DISPLAY

Metadata should be included on thematic maps for the same reasons they appear on maps of topography or geology: to help users evaluate data quality, ascertain whether maps are appropriate for their intended use, and compare them with other maps. Among the most common ways to provide metadata are through text notes or metadata files, but when a map is based on data that vary spatially in quality or source, cartography is an efficient means of transmitting them. Previous research and published maps indicate that cartographers have essentially four options for the graphic display of metadata (Beard *et al.* 1994; MacEachren 1992; McGranaghan 1993; van der Wel *et al.* 1994). They may be displayed by means of:

(1) A visualization technique on the main map (metadata overlaying main map)
(2) A visual variable on the main map (enhancing existing map layer)
(3) Animation or other computer-based technique
(4) A separate map, often an inset map

Option 1 is frequently employed to illustrate basic data limitations. It has received relatively little attention from researchers, probably due to the limitations imposed by multivariate cartography. If display of metadata requires use of an additional visualization technique, then it will confuse those users who are unused to maps composed of multiple layers. The best-known example of this kind of metadata indicator is the use of the color purple by the US Geological Survey to indicate revisions on the 1:24,000 series of topographic maps. This is effective because that color is employed solely for updates and communicates only a single dichotomous item of metadata (whether or not an area has been updated). The same principle underlies the use of blank space to inform users that data were unavailable or fell below a minimum threshold for a given area, as in Brewer and Suchan's (2001) maps of racial and ethnic group distribution.

Pickle *et al*. (1996) employed a separate cartographic technique to map data availability. In a series of choropleth maps of disease-specific death rates by health service areas, they indicated "sparse data" by means of double hatching. This consists of "parallel white and black hatch lines [which] allow visibility of the hatching over light and dark colors." The choropleth colors are sufficiently distinct and the hatched lines narrow enough that one can easily perceive both layers (data and metadata). Numerous methods were considered during production of this atlas before selection of the hatching technique, including gray scales, textures, dots, and point symbols such as asterisks (MacEachren and Brewer, 1995). Hatching is effective as an indicator of the existence of a quality problem (*e.g.* sparse data), but would be less practical for more complex problems, as the use of multiple widths or colors of hatching would confuse the main map.

Previous research has focused on option 2. Maps constructed in this way do not need to have multiple layers, though they will be multivariate. Instead, existing symbols or techniques are enhanced by the use of an additional visual variable. As shown by examples in Edwards and Nelson (2001), metadata added through symbol enhancement have the virtue of not adding additional objects to the map. For example, if the size of graduated circles are used to represent data values, one may add metadata by coloring the circles. The problem with this method is that it may be impossible to add metadata to all parts of the map. It is difficult for users to determine the color of very small graduated circles and therefore metadata will only be transmissible for those regions whose circles are above a size threshold. A similar problem hinders the use of the third dimension to indicate data quality. Because taller prisms may obscure shorter ones, "several views may be required to gain a complete picture of the data" (Beard and Mackaness 1993). This is practical online but not on paper where space is at a premium.

MacEachren (1992) proposes communicating the uncertainty level of data through direct manipulation of map objects. If data values are uncertain or the limits of regions undefined, this lack of certainty can be communicated by making map objects and symbols "fuzzy" or adding "fog" to the map, *i.e.* eliminating or reducing the sharpness of edges and instead using color gradations to indicate regions, boundaries, or attribute values in dispute. One limitation of this method is that it is itself uncertain: users have no means for quantifying the degree of uncertainty displayed (a difficulty previously encountered in studies of continuously-variable gray shading for choropleth maps). Nor is it always possible for cartographers to quantify a single variable called "uncertainty". Use of different or multiple cartographic techniques for display of fuzzy data (Plewe 1997) may mitigate this problem, but, on a more basic level, people without cartographic training (most people) may interpret foggy images as publication errors rather than intentional features of a map.

> *". . . on a more basic level, people without cartographic training (most people) may interpret foggy images as publication errors rather than intentional features of a map."*

Another potential problem for maps intended for wide dissemination is that some users may be confused by the addition of metadata to a map. According to one study, the level of confusion is highest with non-graphic metadata indicators (i.e., notes in text format) (Edwards and Nelson 2001). However, this difficulty has not been fully quantified by studies of cartographic perception, in part because study subjects are often students in geography classes—a group that by definition has a higher level of interest in and knowledge of maps than the general public.

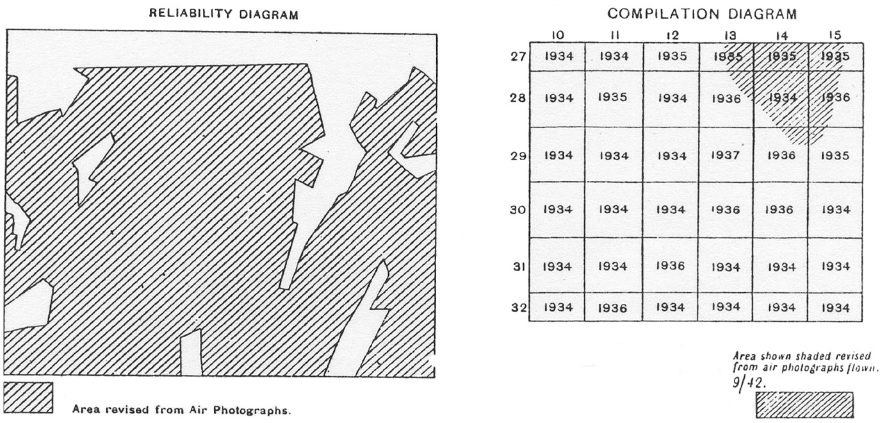
As cartography has become a digital endeavor, new methods (Option 3) for display of metadata have been developed. Van der Wel *et al.* (1994) review an array of techniques involving time, sound, and computer graphic effects. Perhaps the most intriguing animation technique is described by Fisher (1993). In his "error animation" technique, pixels flicker between possible alternate attribute values with each value being present with a relative frequency equivalent to its probability of actually occurring at any given point, thus avoiding the use of mixed or gradational symbols. In a similar vein, Evans (1997) produced a map in which a reliability layer flickered over a land use layer. Veregin (1993) and Fisher (1994) employed sound to transmit quality information, using a variety of audible signals to indicate the existence of different quality problems and the level of reliability of the source data. Though these methods allow for simultaneous communication of data and metadata, the majority of thematic cartography continues to be directed toward paper publication. Even on the internet, most maps are essentially digital equivalents of paper maps[2]. Thus if one is to transmit metadata to all users in an easily-accessible format, it is still necessary to think in terms of static (whether paper or digital) rather than dynamic display of information.

"Even on the internet, most maps are essentially digital equivalents of paper maps. Thus if one is to transmit metadata to all users in an easily-accessible format, it is still necessary to think in terms of static (whether paper or digital) rather than dynamic display of information."

Certain types of inset maps (Option 4) are widespread. These are often called reliability diagrams because they are intended to help users assess aspects of data quality. While they have been in use for more than 50 years, in the vast majority of cases they have been applied only to maps based on airphotos or satellite images. Reliability diagrams commonly indicate the scale, resolution, or date of acquisition of source images or maps, or the revision date (Figure 1). Yet as Wright (1942) noted, the "basic reliability" of government topographic maps is hardly in doubt, but one cannot "take on faith the reliability of the average . . . statistical map." Why, then, are metadata indicators so common on the former and absent from the latter?

"A literature review revealed only one example of a demographic map with reliability diagrams."

A literature review revealed only one example of a demographic map with reliability diagrams. Porter (1956) conducted a census of Liberia, using airphotos to count housing units. After estimating the number of people per housing unit for different regions of the country, he was able to obtain an estimate of the total population of Liberia. He created a detailed population map, on which he included two inset maps (Figure 2). The first rates the accuracy of the housing unit counts while the second evaluates the probable error in the estimates of people per housing unit. The first is



	10	11	12	13	14	15
27	1934	1934	1935	1935	1935	1935
28	1934	1935	1934	1936	1934	1936
29	1934	1934	1934	1937	1936	1935
30	1934	1934	1934	1936	1936	1934
31	1934	1934	1936	1934	1934	1934
32	1934	1936	1934	1934	1934	1934

Figure 1. Sample reliability diagrams from topographic maps. Diagram on left from Ordnance Survey (1943) indicates revised areas with hatching. Diagram on right from Ordnance Survey (1944) also provides dates of compilation.

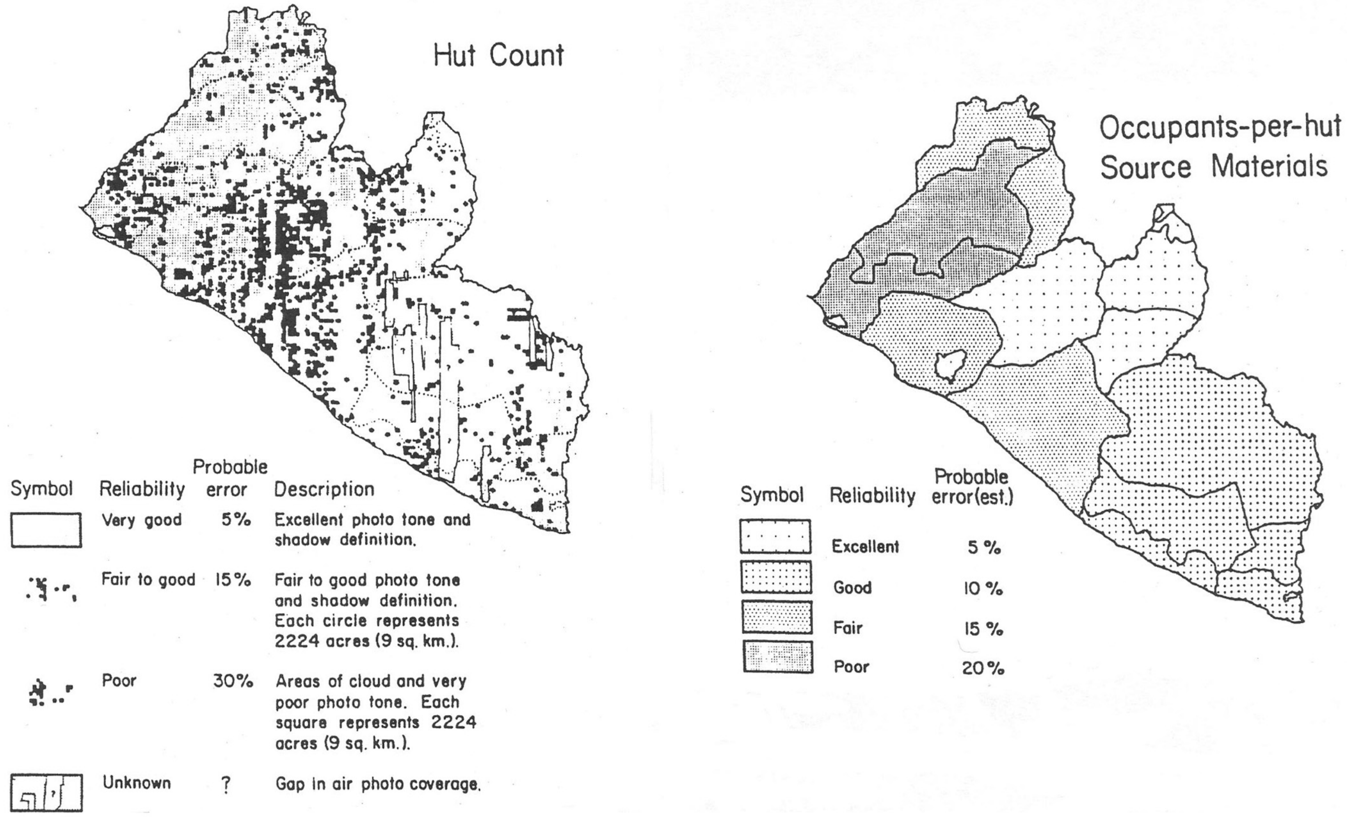


Figure 2. Reliability diagrams from map of the population of Liberia (Porter 1956). Diagram on left assesses quality of source data (airphotos). Diagram on right evaluates quality of demographic information (persons per housing unit).

particularly useful to map users because it indicates the quality of the airphotos–based on clarity and the level of cloud cover–used as data sources and shows areas where no airphotos were available.

The advantage of inset maps is that they are separate from the main map layer, without consuming precious space[1]. Within space constraints, inset maps can be made as complex as needed without risk of visual conflict with the primary map layer. At the same time, as McGranaghan (1993) observes, separation from the main map also introduces the primary deficiency of inset maps or reliability diagrams: they must be "mentally overlaid" on the main map to be properly interpreted. When the main map is a topographic or geological map and the reliability diagram shows information about airphoto or survey coverage, this will be difficult because coverage zones of data sources do not correspond to mapped areal units. The edges of a survey or photo do not follow contour lines or lithologic contacts, making it difficult to match regions and boundaries depicted in the reliability diagram with those shown on the main map. Thus if one is to argue in favor of inset maps in preference to the other options, it must be with the proviso that insets are to be designed so as to ease mental overlay. This can be accomplished with socioeconomic data because data and metadata are reported for the same areal units, or for units at different levels in a standard hierarchy. For example, if the main map shows data for counties, metadata may be available for counties or states. If one includes state boundaries on the main map, users will have no trouble comparing state-level metadata with a main map based on counties.

"The advantage of inset maps is that they are separate from the main map layer, without consuming precious space. Within space constraints, inset maps can be made as complex as needed without risk of visual conflict with the primary map layer."

METADATA AND DATA QUALITY

The term “reliability diagram” implies the transmission of information about data quality. Inset maps can, however, display geographic variation in any data that describe the source data, whether or not related to quality. Therefore the term “metadata diagram” is more appropriate, because metadata include all information about data (or data about data) that users need to fully understand the data, e.g. sources and processing methods. Inset maps will be most appropriate whenever metadata should be displayed spatially but need not appear at the same scale as the main map, that is, in situations where the metadata are neither very basic nor very complex. In the former case, a text note or visualization technique (Option 1 above) is sufficient, while the latter case calls for an independent, full-size metadata map (see Note 1).

“Inset maps will be most appropriate whenever metadata should be displayed spatially but need not appear at the same scale as the main map, that is, in situations where the metadata are neither very basic nor very complex.”

In principle, there is no reason why metadata diagrams should be limited to source and revision information on image-based maps. Many types of thematic map are based on multiple data sources, or on data of different vintages. One cannot, for example, create an international demographic map without recourse to multiple data sources. Even the uniform data publications of international agencies depend on reports from national statistical agencies. These in turn depend on census and survey results obtained at different points in time in each country, by different methods. Date of acquisition and revision information should be shown on demographic maps where possible to allow users to assess reliability or fitness for use. But these are not the only items of metadata that may be of interest to map users. Definitions of mapped variables (*e.g.* ethnic groups or the poverty level) may differ among sources, errors may be introduced through the cartographic method, and accuracy can vary widely even within a single data set. While a satellite sensor will have a fixed pixel size, resolution, and bandwidth, the same is not true for the human equivalent, the census form. The level of undercount in the census is not uniform across the US, despite the use of a uniform survey instrument. If users can be informed of geographic variation in these metadata by cartographic methods that do not detract from the main map, they will surely be able to make better informed choices about reliability and fitness for use.

“If users can be informed of geographic variation in these metadata by cartographic methods that do not detract from the main map, they will surely be able to make better informed choices about reliability and fitness for use.”

One reason for the scarcity of metadata indicators on demographic maps is the absence of published frameworks of data quality attuned to the needs of socioeconomic data producers and users. Though data quality is discussed in Census Bureau reports and demographic reference publications (*e.g.* Shryock and Siegel 1971), the demography literature does not include anything comparable to the discussions of the elements of data quality found in Buttenfield and Beard (1994), Godwin (1997), and Guptill and Morrison (1995). These frameworks are perhaps too comprehensive for socioeconomic data. Many of their elements, in particular those directed towards positional accuracy and resolution, are of lesser interest for thematic mapping of demographic variables because they were clearly derived with topographic or image mapping in mind. Human data suffer from an array of problems related to semantics and human error that do not arise in endeavors of physical science. For example, while physical geographers and geologists may argue over the specification of soil types, it is at least possible to arrive at common definitions after discussion. In contrast, it is impossible to create a rigorous definition of a religious or ethnic group that will be accepted by all, because there is no scientific basis for such delineation[3].

METADATA DIAGRAM DESIGN

To achieve its purpose, a metadata diagram must communicate information about data (1) rapidly and (2) without confusion to (3) all users, regardless of background in cartography. Cartographers do not have the luxury of designing graphics solely for their colleagues, as is the case in some fields: a good map can be used and enjoyed by experts and the general public alike. A metadata diagram must therefore be (1) graphically simple, (2) easy to understand, and (3) without jargon. Because a diagram will always be smaller than the map it enhances, simplification is essential. This means eliminating every unnecessary detail or graphic element: as Tufte (1983, p. 51) suggests, "Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space." All maps must be understood if they are to communicate information, but clarity is of even greater importance when designing metadata diagrams. Rapid comprehension implies the use of simple techniques, colors, and patterns; concise text; and omission of all extraneous detail. Successful communication with all users necessitates use of a simple vocabulary, omitting technical terms relating to data quality elements.

"Cartographers do not have the luxury of designing graphics solely for their colleagues, as is the case in some fields: a good map can be used and enjoyed by experts and the general public alike."

Though these may sound like good guidelines for all mapmaking, the fact is that many demographic maps have a high level of geographic detail and/or are graphically complex. While good design can compensate for the complexity of source data, this option is not available for metadata diagrams because of the time constraint of the map reader's attention span. One hopes that the user will, on viewing a complex but interesting map, devote extra time to understanding it. On the other hand, the user is likely to ignore a complex reliability diagram because it is not the center of attention: it is part of the "ground" rather than the "figure". This is all the more likely with thematic maps created for distribution to the general public, as opposed to topographic or geological maps whose audience is largely composed of people with a higher level of cartographic literacy. As with source notes, subtitles, and scale bars, a metadata diagram should not detract from the main message of a map but rather provide information to those who seek it. If readers skip over wordy titles or metadata notes, they will certainly pass over a dense graphic whose purpose is not clear at a glance. Thus the first priority in the design of metadata diagrams must be simplicity, followed by care in designing graphics and text that will be readily understood by a broadest audience.

Diagram Components

"In the interest of conciseness, a metadata diagram should have a minimum of separate graphic elements."

In the interest of conciseness, a metadata diagram should have a minimum of separate graphic elements. As the main map layout will already include a title, the author's name, a north arrow, etc., these can be omitted from the diagram. A scalebar is also unnecessary because it should be obvious that the diagram covers the same area as the main map, and scale is not relevant to the message of the diagram. There are at most two essential elements: the diagram itself and a legend describing the symbols, colors, or patterns used in the diagram. The title of the legend can substitute for a larger map title (as in Figures 3 and 4b). Symbology should be kept as simple as possible; choropleth shading or line patterns (as in Figures 1, 3 and 4a) will work best in many cases. Point symbols (graduated symbols, shapes) may also be effective but only in cases where the areal units in the metadata diagram are relatively few. Otherwise there is the risk of overlap, illegibility due to reduced size, or confusion from the clustering

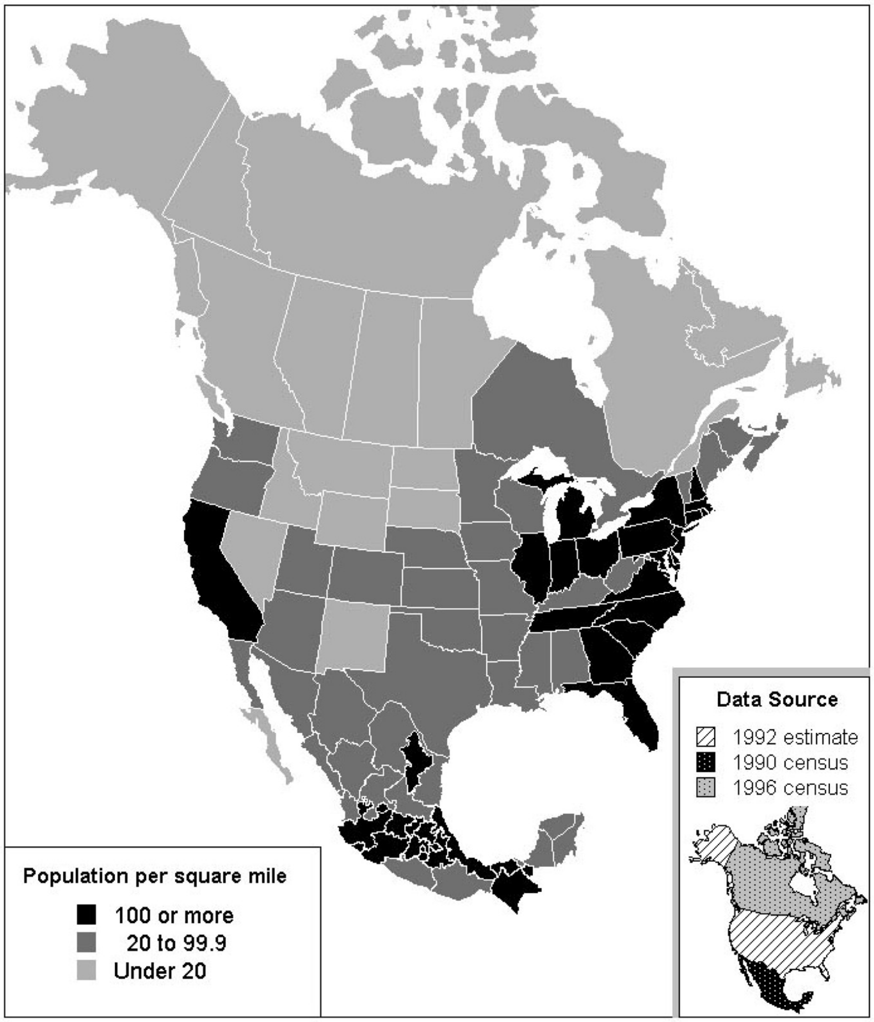


Figure 3. Population density by state or province in the USA, Canada, and Mexico. Metadata diagram showing data source and date.

of many different objects in a small space. Where the diagram need only transmit a handful of items of information, such as a few dates or quality ratings, one may even omit the legend completely and print the information within the map units (Figure 5)–hardly a recommended practice in standard cartography, but useful when space is at a premium and one is not striving to communicate detailed patterns.

"Where the diagram need only transmit a handful of items of information, such as a few dates or quality ratings, one may even omit the legend completely and print the information within the map units–hardly a recommended practice in standard cartography, but useful when space is at a premium . . ."

EXAMPLES

Figure 3 is a map of population density by state or province in the United States, Canada, and Mexico. Data quality is an issue because the data were gathered at different dates, and the population figures for Canada and Mexico are from censuses while the US data are estimates. The metadata diagram can, however, be greatly simplified and reduced in size relative to the main map because all data within each country were gathered at the same time and none of the three countries is very small relative to the others. Thus it is only necessary to show four areal units in the inset (continental US, Alaska, Canada, Mexico). The resulting diagram indicates both the date and method of data acquisition while occupying only a small fraction of the map area.

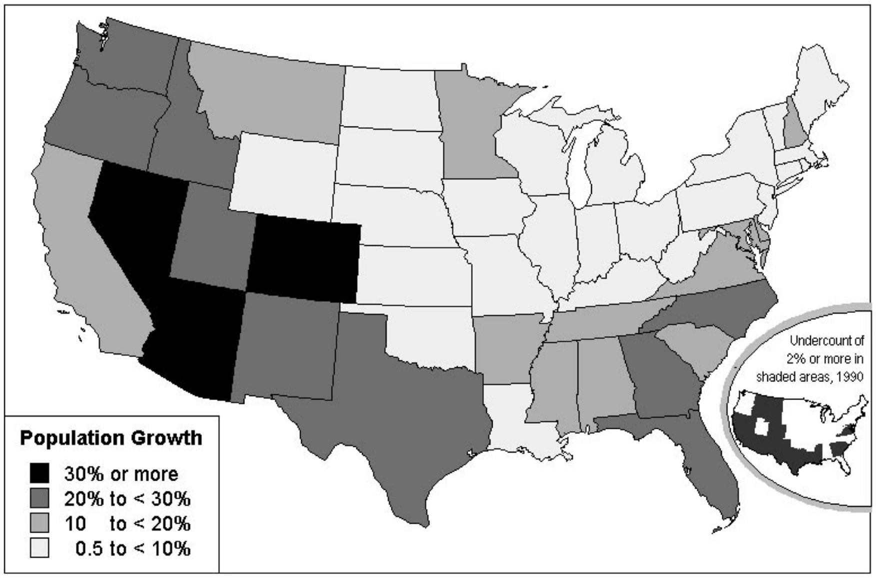


Figure 4a. Population growth from 1990 to 2000. Metadata diagram indicating undercount estimates for 1990.

Census undercount information can easily be added to a population map with a metadata diagram. Figure 4a shows the percent change in the US population from 1990 to 2000. The estimated net undercount in 1990 is shown in the inset map. In this case, the geographical units (states) are the same, but the reliability diagram uses a reduced number of numerical classes to ensure that it is still readable with a size reduction. State boundary lines are also eliminated to create a dasymetric image. Figure 4b is a less-simplified alternative version of the same diagram. Note that there is a correlation between states with higher undercount rates and states with higher population growth rates. This may in part reflect the reduction in the net undercount rate between 1990 and 2000. High growth rates from 1990 to 2000 are due not only to actual growth during the decade but also a fuller counting of people already present in 1990. Here, the metadata diagram does more than serve as an indicator of data reliability: it also enriches the message of the main map.

A metadata diagram can also provide information that helps the user interpret the main map, as opposed to assessing its accuracy. This stretches the definition of metadata, as it does not involve information "about" the data on the main map but rather adds to the user's understanding of the spatial pattern revealed by mapping the data. However, the provision of this additional information has the same goal as the provision of orthodox metadata: to further inform the user about the data. In Figure 5, the main map shows the starting salary for teachers by township in a hypothetical region, while the metadata diagram shows the cost of living index by county (the smallest unit for which cost of living data are generally available). Users can readily observe where salaries have not kept pace with cost of living. County boundaries are indicated by heavy lines in the main map to reduce or elimi-

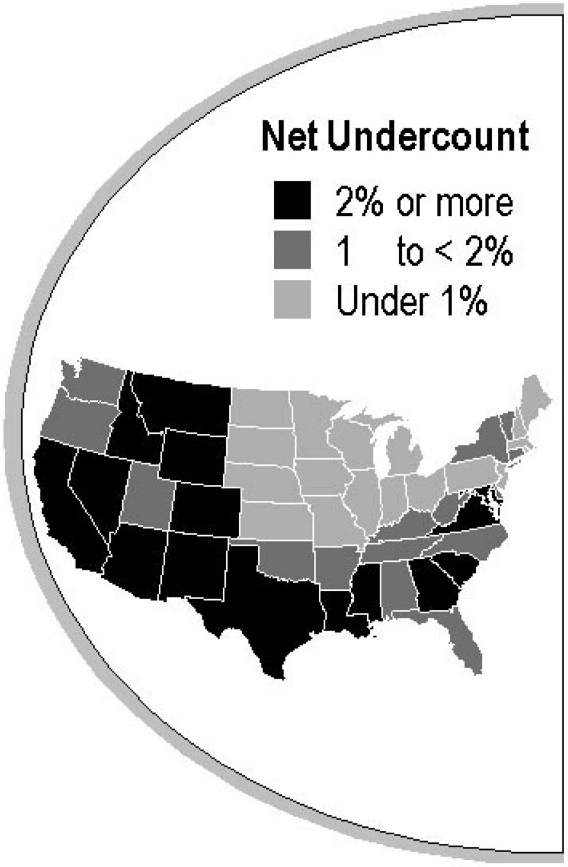


Figure 4b. Alternate metadata format for 1990 undercount data.

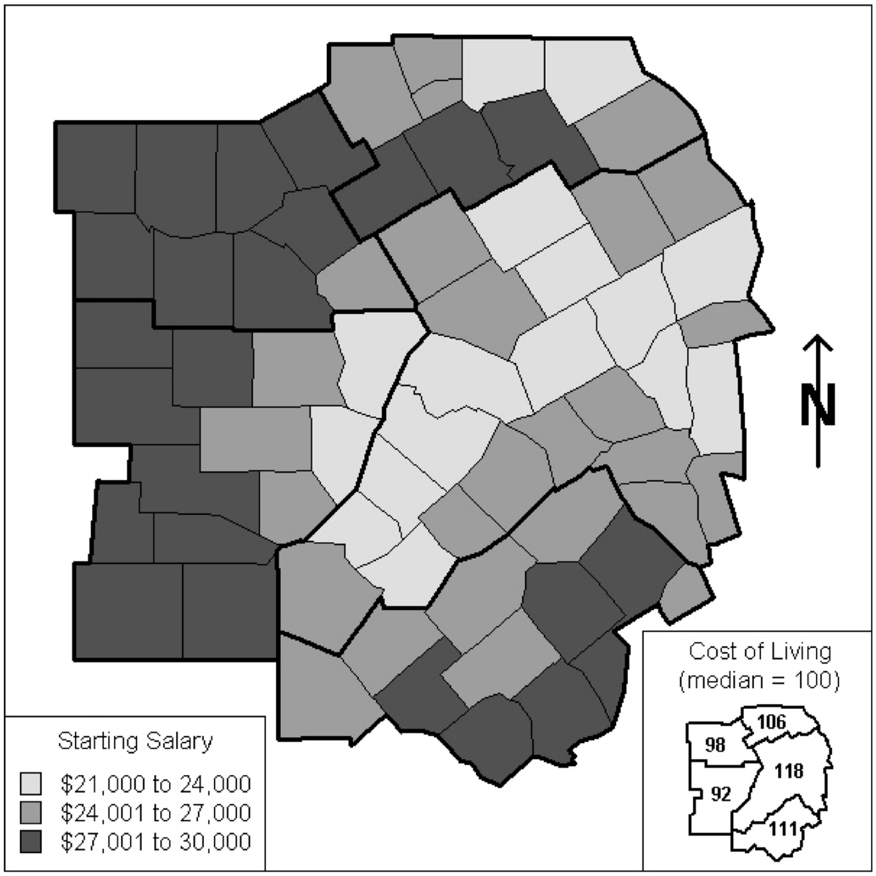


Figure 5. Starting teacher salary by township. Text-label metadata diagram showing cost of living.

nate the mental overlay problem noted by McGranaghan (1993). As this map has very little extra space, the metadata diagram has been skeletonized to county boundaries, a title, and printed text. Though this would be inadvisable for the main map, it is effective in the metadata diagram because of the small number of areal units.

DISCUSSION AND RECOMMENDATIONS

> *"A metadata diagram can be added to most thematic maps with a minimum of extra effort on the part of the cartographer."*

A metadata diagram can be added to most thematic maps with a minimum of extra effort on the part of the cartographer. Sometimes, as in the undercount and population density examples above, the information needed to create the reliability diagram was readily available from the same source as the data for the main map. For example, undercount information is available on the Census Bureau's website (www.census.gov). In all of the above examples, simplicity and clarity have been emphasized as the primary principles of design. To be effective, a metadata diagram must communicate quality or background information efficiently and without ambiguity. Simplicity is equally important to the cartographer facing a production deadline. If metadata diagrams can be created quickly, they have the potential to become more common and to enhance the cartographic understanding of more users.

More complex quality problems than those presented here do exist, but the selected examples illustrate several of the most common problems. In this situation–when the benefits to the map user (better communication of information, improved understanding of data limitations) clearly outweigh the cost to the cartographer–metadata diagrams should always be included. In situations where it is more difficult to obtain geographically-referenced quality information, cartographers should consider whether the quality problem is serious enough to merit the extra investigation or data processing. Where appropriate, these diagrams have the potential to improve cartographic communication by enhancing the comprehension of the map-viewing public.

NOTES

1. Ideally the metadata map would be parallel to and of the same size as the data layer, as illustrated in Howard and MacEachren (1996). This works online but not on paper, where publication costs dictate that space must be conserved. The importance of "conservation of space" as a motivating factor in cartographic design has yet to be fully investigated. Additional online space has no marginal cost. Thus cartographic design for the Internet need not fit into the confines of a rectangular sheet of paper. Most maps posted online today are replicas of paper originals, but entirely new principles of cartographic design may be expected to arise as online mapping becomes more important.

2. Most maps currently available on the internet are non-animated images that may be readily printed on paper. Examples include online road maps and maps of census data. Printing an online map is, in fact, a good test of whether it is a static or dynamic display of information: static maps look the same in either format, something clearly untrue about maps with dynamic elements.

3. Further uncertainty is introduced by the ability of humans to respond falsely to surveys. Measurements with mechanical instruments can be redone, but one cannot "repeat the experiment" to check the results of a human study because people can choose to respond differently each time.

REFERENCES

Beard, M.K., B.P. Buttenfield, and W.A. Mackaness. 1994. *Research Initiative 7: Visualization of the Quality of Spatial Information Closing Report.* Santa Barbara, CA: National Center for Geographic Information and Analysis.

Beard, K., and W. Mackaness. 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartographica* 30(2/3):37-45.

Brewer, C.A., and T.A. Suchan. 2001. *Mapping Census 2000: The Geography of U.S. Diversity.* Washington, DC: US Census Bureau.

Buttenfield, B., and M.K. Beard. 1994. Graphical and Geographical Components of Data Quality. In *Visualization in Geographical Information Systems*, vol. 1, ed. H.M. Hearnshaw and D.J. Unwin, pp. 150-157. New York: Wiley.

Edwards, L.D., and E.S. Nelson. 2001. Visualizing Data Uncertainty: A Case Study Using Graduated Circle Maps. *Cartographic Perspectives* 38:19-36.

Evans, B.J. 1997. Dynamic Display of Spatial Data-Reliability: Does It Benefit the Map User? *Computers and Geosciences* 23:409-422.

Fisher, P.F. 1993. Visualizing Uncertainty in Soil Maps by Animation. *Cartographica* 30(2/3):20-27.

Fisher, P.F. 1994. Hearing the Reliability in Classified Remotely Sensed Images. *Cartography and Geographic Information Systems* 21:31-36.

Godwin, L. 1997. Coordinating Standards on Data Quality: An Important Ingredient for Cartographers. *Proceedings of the 18th International Cartographic Conference:* 533-540. Stockholm.

Guptill, S.C., and J.L. Morrison. 1995. *Elements of Spatial Data Quality*. Kidlington, England: Pergamon.

Howard, D., and A.M. MacEachren. 1996. Interface Design for Geographic Visualization: Tools for Representing Reliability. *Cartography and Geographic Information Systems* 23:59-77.

MacEachren, A.M. 1992. Visualizing Uncertain Information. *Cartographic Perspectives* 13:10-19.

MacEachren, A.M., and C.A. Brewer. 1995. Mapping Health Statistics: Representing Data Reliability. *Proceedings of the 17th International Cartographic Conference*: 311-319. Barcelona.

McGranaghan, M. 1993. A Cartographic View of Spatial Data Quality. *Cartographica* 30(2/3):8-19.

Ordnance Survey. 1943. Maeseyck. GSGS 2364 series. Britain: Ordnance Survey. Reprinted by Army Map Service, Washington, 1944, as Belgium and NE France 1:100,000 series, map sheet 4.

Ordnance Survey. 1944. Oldenburg. GSGS 4416 series. Britain: Ordnance Survey. Reprinted by Army Map Service, Washington, 1948, as series M641 (Germany 1:100,000), map sheet M2.

Pickle, L.W., M. Mungiole, G.K. Jones, and A.A. White. 1996. *Atlas of United States Mortality*. Hyattsville, MD: US Department of Health and Human Services.

Plewe, B. 1997. The Cartographic Representation of Gradation: Fuzzy Maps from Fuzzy Data. *ACSM/ASPRS Annual Convention 1997*: 83-92. Seattle.

Porter, P.W. 1956. *Population Distribution and Land Use in Liberia*. Doctoral dissertation (Geography), University of London.

Shryock, H.S., and J.S. Siegel. 1971. *The Methods and Materials of Demography*. Washington, DC: Bureau of the Census.

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

Van der Wel, F.J.M., R.M. Hootsmans, and F. Ormeling. 1994. Visualization of Data Quality. In *Visualization in Modern Cartography*, ed. A.M. MacEachren and D.R.F. Taylor, pp 313-331. Kidlington, England: Pergamon.

Veregin, H., P. Krause, R. Pandya, and R. Roethlisberger. 1993. Design and Development of an Interactive 'Geiger Counter' for Exploratory Analysis of Spatial Data Quality. *GIS/LIS Proceedings*: 701-710. Minneapolis.

Wright, J.K. 1942. Map Makers are Human: Comments on the Subjective in Maps. *The Geographical Review* 37:527-544.