

TECHNIQUE FOR DATA MAPS
Griffith Feeney
July 21, 1992

(1) We begin with (i) a map showing the area of a country or a region within a country divided into numerous areal units and (ii) a list of numerical values corresponding to these units. Our objective is to produce a choropleth map (Robinson et. al. 1984:343-366; see also the discussions in Cleveland 1985: 282-289 and Monmonier 1991: 40-42) showing the spatial variation in the given values. This means deciding on several "cut points" that divide the given numerical values into several groups and then coloring or shading in the same way all areal units with values lying in a given group.

(2) Areal units will usually be identified by names, both on the map and on the data list. Index numbers for areal units are easier to work with than names, both because names may be unfamiliar, and because index numbers provide an order that can help us find units on the map. Thus we prefer a map that shows index numbers, with a list showing the correspondence of index numbers to names to one side, and a data list that shows the same index numbers as well as names.

(3) Index numbers on the map proper should be bold enough to be easily read through tracing paper and should be positioned off center. Maps on standard size paper (8 1/2 by 11 inches) are most convenient, but the areal units must be large enough to work with. We need to be able to write a two digit number inside each area without undue crowding.

(4) The first step is to make a stem-and-leaf plot (Mosteller and Tukey 1977; Tukey 1977) of the data to be plotted to get a sense of the distribution of the values. This means rounding to two significant figures, and this will usually be sufficient precision for mapping. Certainly we will not use more than three significant figures.

(5) Next, secure a piece of tracing paper over the map with paper clips and write down the two (or three) digit value for each unit on the tracing paper over that unit. Record the values first, before making any other marks, excepting an indication of what data is being plotted, on the tracing paper. Boundaries, legend, and other items should be drawn in later.

(6) We should not forget to label the tracing paper to indicate what data is plotted, preferably with a detailed indication of source, including table and page numbers. An obvious point, but one easily overlooked in the heat of execution. There is something to be said for writing title and source notes on Post-It pads, rather than on the tracing paper directly. In the case of

changes, then, we may simply pull off one note and replace it with another, saving the time and messiness of erasing.

(7) It is useful to be able to see both the value, on the tracing paper, and the index number of the unit, on the map beneath. This is why we want index numbers on the map off center.

We use tracing paper superimposed over the map so as to be able to work with index numbers rather than names, while at the same time being able to record values directly on the map. If we work on map copies directly, index numbers will get hopelessly confused with the values plotted, whence we must work with names.

(8) Recording values directly on the map is important for several reasons. When we are constructing the map to begin with, it facilitates the identification of spatial patterns. When the map is completed, it allows us to see what the effect of changing the dividing lines between categories would be. Most generally, recording values increases the information content of the display without impairing its visual effectiveness.

(9) Recording values is of course practical only if we are dealing with relatively small numbers of reasonably large units, tens or hundreds. Census grid square maps (Tufte 1990:40-41), which consist of many thousands of small areas, are useful representations, but obviously not ones that can be supplemented by adding numerical values.

(10) Now identify an area, i.e., a collection of areal units, in which to look for a pattern. We may look at all units, but we may also look at an area less than the whole. In Vietnam, for example, we may want to look at the north and the south separately.

(11) We are looking for a spatial pattern of some sort, almost always involving contiguity, i.e., we expect see, when we are done, a number of groups of contiguous units colored similarly. Other patterns are of course possible, e.g., checkerboard, but are unlikely to be encountered in practice.

(12) Beyond contiguity, we are likely to be looking for a gradient in a particular direction, a core-periphery pattern corresponding either to a central place and its outlying areas or to a Skinnerian region, or some other preconceived pattern.

(13) Find the areal unit with the lowest value and shade in this unit with (say) a pink highlighter. Now look for the unit with the next higher value and consider what will happen if we shade this second point the same way. The second unit will either (i) adjoin the first, giving two continuous units, or (ii) not adjoin the first, giving us a "jump" to a non-contiguous unit.

(14) Now look for the unit with the next higher value and

consider what will happen if we shade this unit with the same color. Proceeding in this way, we build up a group of units with values at the lower end of the distribution. We may of course begin with the highest value and proceed down, rather than at the lowest value and proceeding up, according to context.

(15) What we want to decide is how far to go, how many units to put in the first category. There are several criteria. First, we want an informative visual representation of the areal data, and this argues against putting too many units together. A map that groups all provinces but one, for example, will be minimally informative. If we are to have four groups, we are likely to want at least some of them to have roughly one fourth of the total number of areal units.

(16) A second criterion is contiguity, whence we are more likely to stop adding units when the next candidate for inclusion is a jump. If at any stage we do make a jump, we will aim to continue adding units to regain contiguity, insofar as possible.

(17) A third criterion is to have units with similar values grouped together and units with dissimilar values in different groups. Thus we are more likely to stop adding units when the difference in value between the last and the potential next unit is relatively large. Thinking of the data as ordered along a scale (assuming ordering is appropriate), we want to divide the data between points that are far apart, and not, preferably, between points that are close together. In particular, we don't want to divide the data in such a way that a slight change in the division line makes a big change in the pattern shown on the map.

(18) Highlighters have two advantages over colored pencils or monochrome patterns (horizontal, vertical, and slanting lines, crosshatch, and so on). First, areas on the map may be colored much more rapidly and easily with highlighters. Second, pink and yellow highlighters conveniently give four colors, pink, yellow, gold (generated by overlaying yellow on pink), and white (no shading).

(19) Having decided on a first group of spatial units and corresponding interval of values, we proceed on to higher values and form a second group and interval, and then to a third, and perhaps a fourth.

(20) An additional advantage of using highlighters in this connection is that we may continue on shading pink (say) through two potential groups of values, and then backtrack, overlaying yellow on the units comprising the first group. The end result is that the first group is colored gold, the second group pink. This allows us to see the emerging pattern for the first group through a series of potential cut points before committing finally to one.

(21) For working as opposed to presentation maps, the advantages of highlighters far outweigh the disadvantages, which are, first, that mistakes can't be erased, and second, that color copies can be made only at great expense (20 or more times the cost of monochrome copies).

(22) If the ranges of values grouped together are clearly shown in a legend, monochrome copies may be made and the appropriate groupings of areal units deduced. Note however that having values recorded on the map is essential for this reconstruction.

(23) Particular units may fail to fall into a general pattern formed by other units. We may think of such units as "spatial outliers" and use a separate medium for marking them, a diagonal line pattern, for example. In this way we use the map to call attention both to the general pattern and to one or a small number of exceptional cases.

(24) The number of categories into which the data is classified depends partly on initial disposition and partly on the pattern displayed by the data as it is mapped. There are certainly situations in which we should use two groups only, as (implicitly) in Skinner's indication of high population density areas within China's macroregions (Skinner 1977: 214-215). Using more than four groups makes it difficult to obtain a reasonable correlation between the values of the data mapped and the intensity of the colors used to represent the categories into which they are grouped. Three or four categories will probably be the best choice in most situations.

(25) This process of data mapping should be conceived of as an intelligent and judgmental search for patterns. We aim to make maps that show the most clearly defined pattern possible, recognizing and accepting that this is a matter of judgement and that the same data may reasonably be mapped in more than one way.

(26) While the movement from the data to the data map is not a purely objective and determinate transformation, neither is it purely subjective and therefore arbitrary. If the data do indeed show a spatial pattern, there will be a strictly limited number of ways to represent it. It is useful in this respect to have several people map the same set of data independently, or to do it oneself in several times aiming for maximally different but all reasonable results, and then to assess and compare results.

(27) If the data are not spatially patterned, it will not be possible to create the appearance of a pattern no matter how hard we try. It would be instructive, however, to assign random numbers to areal units and see how far it is possible to go in the

direction of creating patterns where there are none.

REFERENCES

- Cleveland, William S. 1985. The Elements of Graphing Data. Monterey, California: Wadsworth Advanced Books and Software.
- Monmonier, Mark. 1991. How to Lie with Maps. Chicago: The University of Chicago Press.
- Mosteller, Frederick, and John W. Tukey. 1977. Data Analysis and Regression: A Second Course in Statistics. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Robinson, Arthur H., Randall D. Sale, Joel L. Morrison, and Phillip C. Muehrcke. 1984. Elements of Cartography. Fifth edition. New York: John Wiley & Sons.
- Skinner, G. William. "Regional Urbanization in Nineteenth-Century China." Pages 211-249 in G. William Skinner, ed., The City in Late Imperial China. Stanford, California: Stanford University Press.
- Tufte, Edward R. 1990. Envisioning Information. Cheshire, Connecticut: Graphics Press.
- Tukey, John W. 1977. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Publishing Company.