Geography 125
W. R. Tobler

INFORMATION THEORY AND CARTOGRAPHY

Consider the representation of phenomena which can take on several states, in the sense of cybernetics. A convenient convention is to write the number of states in the so-called binary number system, which uses only two symbols, zero and one. If there are two possible states these can uniquely be identified by the tags 0,1. If there are four states these can be labelled with the four symbols 00, 01, 10, 11. Eight states can be distinguished if one uses the names or labels

000, 001, 010, 011, 100, 101, 110, 111.

Each state, of eight, is thusly identified by the two symbols, arranged in three positions. It can then be said that three binary digits are required to specify the eight states. The two words binary digits, are usually contracted to the one term, bits, so that a scheme of eight states is specified by three bits, one of sixteen states by four bits, and so on. In general the number of bits is equal to the exponent of two required to yield the necessary number of states, e.g.,

8 states $= 2^3 \rightarrow$ 3 bits

16 states $= 2^4 \rightarrow$ 4 bits

32 states $= 2^5 \rightarrow$ 5 bits.

The bit is a measure of the amount (not importance) of information contained in the states, in the sense that it tells one the number of symbols required to specify the state when the binary coding scheme is used. This is especially appropriate if all of the states actually occur with equal probability.

One cartographic application is as follows: Suppose some census data are given in the form of a variable which can take on several states. An example would be a one decimal digit for each census tract. There are ten decimal digits so that the number of bits would be

$$10 = 2^{bits}, \text{ or taking logarithms, } \ln_2 10 = \ln_2 2^{bits} = bits \ln_2 2 = bits$$

(recalling that $\ln_2 2 = 1$). From a table one finds that $\ln_2 10 = 3.3219$. Thus the number of bits of information in a single decimal digit is 3.3219. A classification into ten land use types (a nominal variable) would yield the same number (3.3219) of bits. A two digit decimal number contains $\ln_2 100 = 6.64$ bits, and so on.

Now suppose that the number of possible states pertaining to the census data is 128, a convenient 7 bits. Suppose further that the cartographer produces a choropleth map from those data using four grey levels (a convenient two bits). One interpretation is that five bits (7-2=5) of information have been lost, for a compression ratio of 7/2 = 3.5. The calculation here uses $2^2 = 4$; 4 grey levels yields a two bit map, and $2^5 = 32$ levels are contained in five bits, and 128/4 = 32.

If there are N census tracts, and the values for each tract are independent of each other, then the measure of information for the entire data set is the number of tracts times the information for each tract. If each tract can take on any one of 128 values the information contained in the data is 7N bits; that of the map is 2N bits. Another way of looking at this is to say that if there are N tracts, each taking on one of 128 values then there are 128^N possible data sets. The number of possible maps, using four grey levels, is only 4^N. The number of possible maps is thus much smaller than the number of possible data sets. This means that some (32N) data sets could not be distinguished by viewing the choropleth maps.

If, in addition to reducing the number of states (e.g., grey levels), one also combines census tracts to form N^* "regions", then the information content is reduced again; from 7N bits to 2N bits to $2N^*$ bits.

The assumptions that (a) all states are equiprobable, and (b) that neighboring census tract states are independent of each other, are usually not true for geographical data. A consequence of (a) is that it is theoretically possible to construct more efficient codes; i.e., select a special symbol for frequently occurring states. This is discussed in advanced works on information theory. The consequence of (b) is that one must consider the conditional probabilities of the states, somewhat like the fact that the letters "th" occur in combination in the English language more frequently than one would expect from the separate occurances of these letters. Again, advanced information theory covers these cases, albeit somewhat poorly.

The geographical reason for (a) seems to be that geographical data need support: Valleys (and low elevations generally) must occur more frequently than mountain tops because the latter must rest on the former. Much geographical data has this hierarchical structure. Similarly (b) occurs because geological materials cannot exceed their angle of repose (try to make a "cliff" in a pile of dry sand). Similar autocorrelation principles seem to hold for the arrangements of people, income, land use, etc., although these have not been studied carefully. Of course this is what allows one to make contour maps from scattered observations, predicting unknown values from nearby known values.

It is known that the channel capacity (bits per second) of humans is not terribly large (estimates vary but it seems to be circa 3 bits). One interpretation in cartography has been that the number of grey levels on a choropleth map should be kept small. The foregoing equation ($7N \rightarrow 2N^*$) however, shows that a reduction in the number of bits can also be achieved by aggregating tracts. This is equivalent to reducing the spatial resolution, perhaps by differing amounts in different parts of the data. (Average spatial resolution can be defined as the square root of the total map area divided by the number of observations; e.g. for the contiguous USA with data by state, the average resolution is 408 km and one cannot expect to see features smaller than 816 km in size). Some have suggested that people are able to reduce the bits/second by reducing resolution when examining pictures. By squinting at a picture you can reduce the detail; this form of spatial filtering sometimes helps make the picture more understandable. Others have suggested that only the difference in grey levels from a local neighborhood are used by people in studying pictures. Since the number of difference-states would be expected to be less than the total possible states (by (b) above) fewer bits would be needed. Probably people are more sophisticated than this. A black and white television set may have 5 bits (32 greys) or more but I don't look at these; rather I break the scene into "background", "face", "tree" and similar high level concepts, possibly not at a rate exceeding 3 bits/second. It can be expected that information theory will be useful in clarifying this and other aspects of cartography in the future. The attempts to date have been very crude (and often wrong).