Introductory Comments on Information Theory and Cartography

Waldo Tobler
Professor Emeritus
Geography Department
University of California
Santa Barbara, CA 93106-4060
tobler@geog.ucsb.edu

"The context of this work was the measurement of channel capacity required to send a message, but it provides a counting mechanism for any type of information, including symbols on a map."

"An example would be a one decimal digit for each census tract."

Maps contain information. A method by which the amount of information contained on a map can be measured is given by Claude Shannon's theory of information (1948, 1949). The context of this work was the measurement of channel capacity required to send a message, but it provides a counting mechanism for any type of information, including symbols on a map. More recent works include Cover and Joy (1991), and Crowley and Mitchell (1994). It is necessary, however, to recognize that this important theory does not imply anything concerning the veracity or value of the information. It just measures how large a quantity of information is contained in the message or on the map. To explain this, consider the representation of phenomena that 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 then these can be uniquely identified by the tags 0,1. If there are four states (perhaps called A, B, C, D) these can be labeled with the four symbols 00, 01, 10, 11. Eight states (e.g., A, B, C, D, E, F, G, H) can be distinguished if one uses the eight names or labels 000, 001, 010, 011, 100, 101, 110, 111. Each state, of the eight, is thus 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, ***bi**nary digi**ts**, 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. Many desk top computers use a system of bytes, made up of eight bits to yield 256 possible symbols. The newer computers use 16 or 32 bits and can represent larger numbers of symbols, for example for international languages.

The number of bits is equal to the exponent of two required to yield the necessary number of states or symbols, e.g.,

2 states = 2^1 -> 1 bit
4 states = 2^2 -> 2 bits
8 states = 2^3 -> 3 bits
16 states = 2^4 -> 4 bits
32 states = 2^5 -> 5 bits,
etc.

The bit then 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 that 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 as a variable that 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 $10 = 2^{bits}$, or taking the base two logarithms of both sides of this equation, gives $\log_2 10 = \log_2 2^{bits}$. This converts to the correct number of bits, recalling that $\log_2 2 = 1$. From a base two logarithmic table one finds that $\log_2 10 = 3.32$. 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 (3.32) of bits. A two-digit decimal number contains more bits (6.64 or twice as many, as one would expect), and so on. The number 3.32, although it represents the information contained in a single one digit number, does not seem very informative. One would normally round this up to the next higher whole number, i.e., to four bits.

The meaning of bits can be explained somewhat further by using a cartographic example. For this example suppose that the number of possible states pertaining to some integer census data is 128, a convenient seven bits. This means that there are 128 classes into which an observation may fall. Suppose further that the cartographer produces a choropleth map from these data using only four grey levels, a convenient two bits. One interpretation is that five bits, (namely 7 bits - 2 bits = 5 bits) of information have been lost, for a compression ratio of 7/2=3.5. The calculation here uses $2^2 = 4$ (four grey levels yields a two bit map), and five bits ($2^5 = 32$) have been lost. The 128 states have been reduced to only four grey levels, and the amount of information loss is by 32 times.

"The meaning of bits can be explained somewhat further by using a cartographic example."

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

If, besides reducing the number of states (e.g., to four grey levels), one also combines census tracts to form M "regions", with M less than N, then the information content is reduced even further, from 7N bits to 2N bits to 2M 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 or code for frequently occurring states (Hamming, 1986). This is discussed in advanced works on information theory, e.g., Roman (1992). Morse code provides a simple textual example for the English language. The more frequently occurring letters use the simpler codes. In the choropleth example the cartographer would likely use unequal class intervals to more closely approximate the data distribution.

"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."

The consequence of (b, above) 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 occurrences of these letters. Again advanced information theory covers these cases, albeit less adequately with respect to two dimensional phenomena; but see Hammer (1995).

The geographical reason for (a) is 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 geographic arrangements of people, income, land use, etc., although these have not, until recently, been studied as carefully. Of course this is what allows one to make contour maps from scattered observations, predicting unknown values from nearby known values. Extending the bit count to continuous images, such as elevations on a topographic map, presents additional challenges.

More advanced information theoretic studies introduce the concept of entropy, which allows for expectations or probabilities of occurrence. Entropy then measures the amount of new information, beyond what was previously known. A paper by Pipkin (1975) explores this in the context of cartography, as does a discussion paper by Sheppard (1975) in the more general context of geography (also see Thomas, 1981).

"Extending the bit count to continuous images, such as elevations on a topographic map, presents additional challenges."

It is known that the channel capacity (bits per second) of humans is not terribly large (estimates vary but it seems to be circa three bits). One interpretation in cartography has been that the number of grey levels on a choropleth map should be kept small. The foregoing equation (e.g., 7N -> 2M) shows, however, that a reduction in the number of bits can also be achieved by aggregating spatial units, or tracts. This is equivalent to reducing the spatial resolution, perhaps by differing amounts in different parts of the area. Average map 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 circa 400 km and one cannot expect, from the sampling theorem (Jaehne, 1991, p. 45-52), to see features smaller than 800 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 (or map) one 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 different 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; compare Marr (1982). A black and white television set may have five bits (32 greys) or more, but one does not look at these. Instead I break the scene into "background," "face," "tree," and similar high level concepts, possibly not at a rate exceeding three bits per 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 emphasized the cognitive aspects of maps instead of measuring the amount of information contained therein (MacEachren, 1995, provides a good summary). An exception is in the related field of picture processing, where the bulk of the work has concentrated on image compression for electronic transmission.

"It can be expected that information theory will be useful in clarifying this and other aspects of cartography in the future."

As an exercise for students, let them calculate how many distinct images could result from a ten cm by ten cm format aerial photograph, with a crude resolution of only ten lines per millimeter (for a total of 10,000 pixels), and with 256 grey levels (8 bits) possible for each pixel. This could turn into a computer exercise, perhaps using a real digitized aerial photograph, reducing the resolution by combining pixels, or by collapsing the grey levels, and recomputing the information amount. Eventually the scene will no longer be recognizable.

REFERENCES

Cover, T. and Joy, A., 1991, *Elements of Information theory*, Wiley, New York.

Crowley, D. and Mitchell, D., 1994, *Communication Theory Today*, Stanford U. Press, Stanford.

Hammer, E., 1995, *Logic and Visual Information*, CSLI Publications, Stanford.

Hamming, R. 1986, *Coding and Information Theory*, 2nd ed., Prentice Hall, Englewood Cliffs.

Jaehne, B., 1991, *Digital Image Processing*, Springer, Berlin, 383 pp.

MacEachren, A., 1995, *How Maps Work*, Guilford, New York.

Marr, D., 1982, *Vision*, Freeman, San Francisco.

Pipkin, J.C., 1975, "The Map as an Information Channel: Ignorance Before and After Looking at a Choropleth Map," *The Canadian Cartographer*, 12,1:80-82.

Roman, S., 1992, *Coding and Information Theory*, Springer, New York.

Shannon, C.E., 1948, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27:379-423,623-656.

Shannon, C.E., and W. Weaver, 1949, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Sheppard, E., 1975, *Entropy in Geography*, Discussion paper # 18, Department of Geography, Toronto.

Thomas, R., 1981, *Information Statistics in Geography*, CATMOG # 31, Norwich.