

UC Irvine

UC Irvine Previously Published Works

Title

Computer applications in cultural anthropology

Permalink

<https://escholarship.org/uc/item/88d8g416>

Journal

Language Resources and Evaluation, 5(1)

ISSN

1574-020X

Author

Burton, Michael L

Publication Date

1970-09-01

DOI

10.1007/bf02404256

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Computer Applications in Cultural Anthropology

Author(s): Michael L. Burton

Source: *Computers and the Humanities*, Vol. 5, No. 1, Annual Survey of Recent Developments (Sep., 1970), pp. 37-45

Published by: Springer

Stable URL: <http://www.jstor.org/stable/30204004>

Accessed: 15-02-2017 22:18 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Computers and the Humanities*

Computer Applications in Cultural Anthropology

Michael L. Burton

Cultural anthropology falls into an ambiguous position among academic disciplines in that it is considered by some practitioners to be a science, by others to be of the humanities, and by still others to be both. This ambiguity is understandable, since much of cultural anthropology consists of the scientific study of human symbolic systems. It includes folklore, primitive religion, values, semantics, and cognitive aspects of social organization, all of which could be considered to be "humanistic" topics. In studying them according to the scientific method, the cultural anthropologist constructs models of empirical phenomena, and creates rules of correspondence between entities in the models and observable phenomena in the world. This requires collection of data according to standardized procedures, so that any investigator can, by similar procedures, replicate the findings of another with the same kind of data. Data-collection, then, is more than simple observation; it is observation with the goal of arranging observed facts into a "system." At the highest level of scientific advance the system is a complex mathematical model which requires numerical inputs. Cultural anthropology is in the early stages of scientific evolution, and is just beginning to grapple with the problems of quantification. It is very much concerned with the development of ways to measure phenomena which have previously seemed too soft and elusive to express quantitatively.

Since the publication of *The Uses of Computers in Anthropology* (Hymes, 1965), which was based on a symposium sponsored by the Wenner-Gren Foundation in 1962, computer applications in anthropology have expanded greatly, both in scope and in frequency. Most of the material which appeared in that volume is now obsolete, both in its reference to kinds of computer hardware and in the mathematical models to which computers were applied. In 1962 only a very small number of anthropologists had experimented with computers. Now, computer usage is so far beyond the experimental stage that applications by anthropologists occur in the preliminary stages of data analysis, and are never mentioned in print. Also, it is no longer possible for a single person to be conversant with all of the techniques which may find computer applications in anthropology. For these reasons, this article is slanted towards a few areas, which happen to be those that I know best—the areas of data analysis and scaling.

The computer can be used as a processor of numbers or as a processor of lists of symbols. Data analysis and scaling fall within the

Michael L. Burton is an assistant professor of Anthropology at the University of California, Irvine.

former category. I wish to emphasize here that some of the most interesting developments in computer applications to anthropology have been in the latter category, which I have omitted. Within this area I would include uses of the computer to simulate social structure (Gilbert and Hammel, 1966; Randolph and Coult, 1968), for the analysis of text (Colby, 1966), and for the manipulation of genealogical systems, including computerized componential analysis (Coult and Randolph, 1965; Kreps, 1964; Kronenfeld, 1967). In these applications, the primary variables are linguistic or cultural symbols which the computer counts or manipulates. Although these may require statistical computations which use numbers, they also presume the existence of some important nonnumerical inputs.

Classification of Data Matrices

In all cases, data analysis and scaling begin with a matrix of numbers. I have chosen to proceed by classifying the computer applications according to the kind of entities represented by the rows and columns of the matrix. These entities I call actors, objects, and variables. Actors include individuals, households, families, lineages, and other culturally relevant categories of social grouping. Objects include artifacts, concepts, parcels of property, beliefs, plants, and other culturally defined partitionings of the universe. Variables are scientific concepts which are imposed upon phenomena by the anthropologist in accordance with some model of human behavior. According to the logic of combinations, six kinds of matrix are possible:

	ROWS	COLUMNS
1.	variables	variables
2.	objects	objects
3.	actors	actors
4.	actors	variables
5.	objects	variables
6.	actors	objects

The first of these does not occur with raw data, for it is derived from the data through computation. An example is the correlation matrix. The other five represent interactions of data which occur in the preliminary stages of analysis.

Actors Measured on Variables

Variables may be measured on nominal, ordinal, or interval scales. The nominal scale is a simple categorization, the ordinal is a rank ordering, and the interval corresponds to the real numbers, with or without a zero point. All three kinds of data commonly occur with census or survey work. Thus, individuals may be categorized as to kind of residence, rank ordered as to

prestige, and measured (in money units) as to wealth. It is also conceivable that the interval scale would be the result of some more complicated procedure, such as a psychological test.

In cross-cultural studies, cultures are the actors. They are usually coded on nominal scales such as "presence or absence of slavery," or "patrilocal or matrilocal residence." These may be ordered nominal scales such as "degree of social stratification: low, middle, high."

The procedures used for analysis of this kind of data are fairly well standardized, at least from the user's point of view. The methods for computation of variance analysis or contingency tables may change from time to time, but the basic operations and concepts remain constant. Usually, it is possible to refer to a large program package such as the UCLA Bio-Medical series, or *Data-Text* for all of one's needs. The format for punching cards is also standardized, as are some of the rules for coding the data for card punching. Thus, the area of analysis of actors with respect to variables is the easiest for the new user of computers. Consequently, most computer applications in anthropology fall within this category.

The most common use of nominal scale data is the contingency table analysis. Although a number of other ingenious tests are possible with nominal data (Siegel, 1958) the contingency table remains the workhorse of this type of data. It is also the only test that is commonly found in the large program packages.

Similarly, the most common ways to analyze ordinal- and interval-scale data are rank order correlations and the Pearson's correlation coefficient. This is not a prescription of what people should do with their data, but a description of what seems to be commonly done. Here again, the availability of correlation programs and the ease of their use may be having their effects. For example, one of the more interesting questions to ask of interval data is whether different groups of people have important differences on the variables. This is a problem of analysis of variance. Although computer programs for variance analysis are also readily available, they require more statistical sophistication from the user. The difference is enough to exclude many anthropologists, who may have learned statistics in graduate school, but who seem to have difficulty in perceiving how it is relevant to their own data.

A problem with computer programs that generate contingency tables or correlation tables is that they allow people to perform too much computation. A small amount of data can easily lead to hundreds of pages of contingency tables, or to thousands of correlation coefficients. With contingency tables it is possible to run every triplet of variables, using the third variable as a control. Such computations are very cheap, and there is a tendency to run every variable against every other variable lest some important relationship escape notice. This procedure is of dubious scientific validity, since it does not really test a theory, and therefore cannot contribute to the evolution of scientific theory. The more practical

consequence is that there is often too much output to read; since most of the tables go unread, it is as if they had never been computed.

Given this tendency to run all possible combinations of contingency tables, Textor's *Cross-Cultural Summary* (1967) is a great timesaver. Textor computed all two by two contingency tables for more than 500 variables on the 400 cultures of the *World Ethnographic Sample*. His computer program produces well labeled output, which includes all tables that are statistically significant. The researcher who wants to test some relationship from this sample can simply look it up in the *Cross-Cultural Summary*. If it isn't there, it isn't significant. This book suffers from one important deficiency: it does not include any computations of three variables at a time. These are important in order to test whether an observed relationship is genuine, or whether there is some third factor which explains it. It would have been particularly useful, and a good control for diffusion, to have tested some of the most significant relationships within each of the major culture areas.

With cross-cultural data, the computer has also been used for factor analysis (Sawyer and Levine, 1966; Driver and Schuessler, 1967; Gouldner and Peterson, 1962). The two recent studies are based on the same corpus of data, Murdock's 1959 version of the *World Ethnographic Sample*. It contains thirty cultural characteristics such as "social stratification," "agriculture," "exogamy," and "differentiation of cousins and siblings." Sawyer and Levine recoded the variables so as to produce ordered category scales, each with three categories (low, medium, high). The differences between the Sawyer and Levine study and the Driver and Schuessler study are in the next step. Each used a different measure of association as input to the factor analysis. Of nine factors, the three most important are "presence of agriculture," "presence of animal husbandry," and "patrilineality."

One problem with both factor analytical studies is that they use less sophisticated scaling methods than were available at the time the research was done. Some theories of cultural evolution predict a nonlinear development of some traits. Driver and Schuessler mention this possibility (1967:351):

The positive correlation of bilateral descent with hunters and collectors and modern Western society makes the point that a variable common in an early stage may be largely rejected in an intermediate stage, only to reappear in a later stage of social evolution.

Another example is independence training (a socialization variable) as contrasted to obedience training. Independence training occurs with hunters and gatherers and with industrial society, but not in agricultural societies (J. Whiting, personal communication). Ordinary factor analysis cannot account for such nonlinearities, but there are nonlinear factor analytical models which would have been appropriate to this kind of data.

Tables of Interaction between Actors

Here both row and column entries represent actors. When the actors are individuals, the entries in the table could measure the amount of a particular kind of interaction occurring between any two, for example, the amount of money that the row person would contribute to the column person's bride payments (an asymmetrical matrix), or the number of times the two persons were observed to interact during some time span (a symmetrical matrix). Another kind of asymmetrical matrix would be one which contained observations on how often the row person performed a particular kind of behavior (hit, insulted, dominated, was aggressive, gave help, asked for help) to the column person. This sort of data occurs with Whiting's studies of child behavior.

With this kind of data, one possibility is to search for natural clusterings of individuals. The interaction matrix could be considered to be a matrix of similarities between individuals. It would then be proper input to a cluster analysis procedure such as Johnson's hierarchical clustering program (Johnson, 1967). I know of no cases where this has actually been tried. A second possibility would be to test hypotheses about the sociological determinants of the interactions. For example, one might predict that older children would aggress on younger children more frequently than the reverse, or that young girls would express dependence more often than young boys. John Whiting is currently engaged in a computer analysis of these kinds of questions, using what he calls a "target analysis" procedure.

When the actors are social groups, the most important contribution has been Romney's model of endogamy (Romney, 1970). Here the actors are communities or social classes, or other endogamous groups, and the entries in the cells of the table are the numbers of marriages between the two communities. The entry in row i , column j represents the number of times that a man from community i married a woman from community j . Often the various communities have different populations. When the populations of two communities differ greatly, the effect of innate preferences for endogamy or exogamy becomes confounded with the effect of community sizes. If one community is large and the second is small, there may be many marriages between the two relative to the number of marriages within the second community because the smaller community is forced by its small population to look for mates from the larger community. Romney uses an iterative algorithm, which is programmed for the computer, and which compensates for the effects of differences in community size. The iteration program starts with the observed populations and works toward a table for which all communities have the same population. The entries to this iterated matrix then provide information as to the actual preferences for endogamy.

Actors in Relationship to Cultural Objects

With this type of data, the cell for actor i and object j provides

information about the actor's attitude towards the object, or about whether he owns the object, or practices it (where it is a religious observance), or believes it (where it is a belief). A common analysis with this sort of scheme is the Guttman scale. Although computer programs are readily available, I know of no published accounts of such an application. Two examples which did not use a computer are Kay (1964) and Goodenough (1965). Kay studied material culture in Tahiti and found that it could be scaled by the Guttman scale. Goodenough has an analysis of sexual behavior permitted with different categories of kin in Truk, which demonstrates that the sexual prohibitions can be ordered along a Guttman scale. Both studies demonstrate that if more people would work with it, the method could be of great importance to anthropology.

Cultural Objects Measured on Variables

As with the first case, the variables may be of nominal, ordinal, or interval degree of measurement. The measurement may be made either directly, as when an archeologist measures the important physical dimensions of an artifact, through some task that is given to informants, as in the semantic differential, for which each concept (cultural object) is rated by each person on a number of rating scales (Osgood, et al., 1957). The data for each object are its average positions on each of the rating scales. In all cases, cultural objects are measured on variables as a means to an end, which is the study of the structure of interrelationships of the objects. In archeology, this structural study is called seriation: the arrangement of the artifacts in a plausible temporal sequence. In cognitive anthropology, it produces a model of the cognitive structure of a semantic system, or of a system of beliefs. Typically, as input to a computer program, the object-variable matrix produces measures of association among the objects. One such measure is the correlation coefficient. There exist a wide variety of other measures, which do not make the restrictive assumptions of correlational analysis, and which are often more appropriate to anthropological data. The resulting tables of measures of association among objects are instances of the last kind of data matrix, for which both rows and columns are cultural objects.

Matrices of Similarity or Difference between Objects

Many techniques have been developed during the past few years for obtaining data on the relative similarities of cultural objects. Much of this development has been a response to a breakthrough in multidimensional scaling (Shepard, 1962, 1966; Kruskal, 1964), the invention of nonmetric multidimensional scaling, in which a data matrix contains measures of similarity or dissimilarity among objects. It assumes that the data contains an accurate rank ordering of the similarities or dissimilarities. From such a rank ordering it produces a spatial representation of the objects in one, two, three, or more dimensions. This is usually done in Euclidean space,

although certain non-Euclidean representations are also possible. It is then possible to examine the spatial representation to test hypotheses about the structure of the objects. For example, binary features, as in phonemics or semantics, would be represented by a dimension of the space for which the objects were bimodally distributed. A hierarchy (taxonomy, key, or tree) would be represented by a group of discrete clusterings of objects. Comparative scales such as prestige, power, evaluation, or purity would appear as an alignment of objects along one of the axes of the space, after suitable rotations had been performed.

It is also possible to do cluster analysis from matrices of similarity or dissimilarity. One of the best procedures for cluster analysis, a relatively new one called hierarchical clustering (Johnson, 1967), produces a complete binary tree from a matrix of dissimilarities. Often it is helpful to do both the cluster analysis and the multidimensional scaling, in order to test whether the cluster model or the spatial model gives a better correspondence to the data.

Much of the pioneering work in the uses of multidimensional and clustering models for cultural data has been done by Volney Steffle (see Steffle, in press). Particularly interested in item-by-use data, he uses procedures for generating all of the kinds of statements that can be made about a cultural domain. He then has informants fill out a matrix for which rows are statements ("X is good for you," "X is something you do in the morning," "X is a bright, happy food") and columns are things (diseases, foods, products, role terms, colors, etc.). Subjects check all the statements that can be made about each thing. By correlating the things across statements, or by using some other measure of similarity, a matrix of similarities can be obtained. This then becomes input to the multidimensional scaling program. Steffle has been working with this kind of data for several years, and it has proven to be very successful for making predictions in applications to marketing research and to political polls.

A variant on item-by-use data occurs in a study of disease concepts in Mexican-Spanish and American English (D'Andrade, et al., in press). Here the original data matrix has disease terms for the columns and beliefs about diseases for the rows ("X is a children's disease," "X should be under a doctor's care," "X is a fatal disease," "Fat people are more prone to X," etc.). The entry in the matrix for disease *i* and belief *j* tells how many subjects thought belief *j* was true for disease *i*. D'Andrade has a factor analysis of diseases and beliefs, hierarchical clustering (using a method slightly different from Johnson's), and multidimensional scaling for which both beliefs and disease appear in the same structure. This paper is one of the best examples of all of the different things the computer can do for anthropological research. The researchers even used a computer editing program to make revisions and to produce multiple copies for distribution to colleagues.

Two other types of data for scaling and clustering are triads data

(Romney and D'Andrade, 1964; Romney and Wexler, in press), and sorting data (Burton, 1968; Burton, in press). In the triads test, subjects are presented with three things at a time and asked to judge which is most different. This tells which two of the three are most similar. The average across all triads for all subjects of number of times two concepts are most similar is the entry in the table of similarities. The primary example of the triads test in anthropology is Romney and D'Andrade's study of English kinship. In the sorting test, names of things are written on small cards. Subjects are then requested to sort the cards so that similar things go in the same group. (Technically, they are asked to partition the set of things.) The major example of this kind of data in anthropology is the multidimensional scalings of role terms and occupation terms by Burton, using data collected by himself and by John Brim.

Other applications of multidimensional scaling in anthropology include a study of Ixil role terms (Harding, 1969) and several applications to archeology by George Cowgill (Cowgill, 1967). The rapid proliferation of multidimensional scaling studies in a period of about two years will soon make this technique one of the most important reasons for anthropologists to use computers.

Summary

This paper covers important developments in the use of computers for quantitative research in cultural anthropology, particularly in areas which (unlike statistics) are uniquely anthropological. These fall into statistical topics and topics in scaling and measurement. By far the largest single usage of computers by cultural anthropologists is for statistical summaries of field data and for simple statistical tests such as the *chi*-squared for the analysis of field data or for cross-cultural studies. As the discipline develops this situation will remain the same. In fact, the proportion of people who use the computer primarily for contingency tables, frequency counts, and correlation analysis may very well increase, since there are many potential users who would fall in this category and only a few potential users who would perform other operations such as multi-dimensional scaling or simulation. The few other computer techniques that would be relevant to anthropology, and for which the technology already exists, include linear regression, as practiced by economists, and linear programming (also practiced by economists), both of which could be extremely useful in the study of peasant economy. Careful research with such models could dispel some of the controversy which has been hindering the development of economic anthropology for the last fifteen years. The training of anthropologists who can understand the relevance of such models to their work may be far in the future, since the majority of them are still skeptical of most formal methods and of the computers which make them work.

References

- Burton, Michael L. *Multidimensional Scaling of Role Terms*. Stanford University Ph.D. dissertation, 1968. Ann Arbor, University Microfilms No. 69-8160.
- . "Semantic Dimensions of Occupation Terms." In A. K. Romney, Roger Shepard, and S. B. Nerlove, eds. *Multidimensional Scaling in the Behavioral Sciences*. In press.
- Colby, Benjamin N. "The Analysis of Culture Content and the Patterning of Narrative Concern in Texts." *American Anthropologist* 68, i(1966): 473-88.
- Coult, Allan D. and Richard R. Randolph. "Computer Methods for Analyzing Genealogical Space." *American Anthropologist* 67(1965):21-30.
- Cowgill, George. "Computer Applications in Archeology." *AFIPS Conference Proceedings*, vol. 31. Washington, D.C.: Thompson Book Co., 1967.
- D'Andrade, Roy G., N. R. Quinn, S. B. Nerlove, and A. K. Romney. "Categories of Disease in American-English and Mexican-Spanish." In A. K. Romney, Roger Shepard, and S. B. Nerlove, eds. *Multidimensional Scaling in the Behavioral Sciences*. In press.
- Driver, Harold and Karl F. Schuessler "Correlational Analyses of Murdock's 1957 Ethnographic Sample. *American Anthropologist* 69(1967):332-52.
- Gilbert, John P. and E. A. Hammel "Computer Simulation and Analysis of Problems in Kinship and Social Structure." *American Anthropologist* 68:71-94.
- Goodenough, Ward H. "Rethinking 'Status' and 'Role': Toward a General Model of the Cultural Organization of Social Relationships." In M. Banton, ed. *The Relevance of Models in Social Anthropology*. London: Tavistock Publications, 1965.
- Gouldner, Alan W. and Richard A. Peterson. *Notes on Technology and the Moral Order*. Indianapolis: Bobbs-Merrill, 1962.
- Harding, Joe R. "Multidimensional Scaling of Ixil Role Terms." Paper presented at meetings of the American Anthropological Association, New Orleans, 1969.
- Hymes, Dell. *The Uses of Computers in Anthropology*. The Hague: Mouton and Co., 1965.
- Johnson, S. C. "Hierarchical Clustering Schemes." *Psychometrika* 32(1967):241-54.
- Kay, Paul. "A Guttman Scale Model of Tahitian Consumer Behavior." *Southwestern Journal of Anthropology* 20,ii(1964): 160-67.
- Kreps, Theodora. "Computer Analysis of Uto-Aztecan Kinship Systems." Stanford University Ph.D. dissertation, 1964.
- Kronenfeld, David. "Computer Analysis of Lineal Kinship Systems." M.S., Stanford University, Department of Anthropology, 1967.
- Kruskal, J. B. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nometric Hypothesis." *Psychometrika* 29(1964):1-27.
- Osgood, C. E., G. J. Suci, and P. H. Tannenbaum. *The Measurement of Meaning*. Urbana: University of Illinois Press, 1957.
- Randolph, Richard R. and Allan D. Coult. "A Computer Analysis of Bedouin Marriage." *Southwestern Journal of Anthropology* 24,i(1968):83-99.
- Romney, A. K. and Roy G. D'Andrade. "Cognitive Aspects of English Kin Terms." *American Anthropologist* 66,iii,part 2(1964):146-70.
- . "A Model for the Analysis of Endogamy." In Paul Kay, ed. *Explorations in Mathematical Anthropology*. Cambridge: Massachusetts Institute of Technology Press. 1970.
- , and Ken Wexler. "Individual Variations in Cognitive Structures." In A. K. Romney, Roger Shepard, and S. B. Nerlove, eds. *Multidimensional Scaling in the Behavioral Sciences*. In press.
- Sawyer, Jack and Robert E. Levine "Cultural Dimensions: A Factor Analysis of the World Ethnographic Sample." *American Anthropologist* 68(1966): 708-31.
- Shepard, Roger N. "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function." *Psychometrika* 27(1962):125-39.
- . "Metric Structures in Ordinal Data." *Journal of Mathematical Psychology* 3(1966):287-315.
- Siegel, Sydney. *Non-Parametric Statistics*. New York: McGraw-Hill, 1958.
- Stefflre, Volney. "Some Applications of Multidimensional Scaling to Social Science Problems." In A. K. Romney, Roger Shepard, and S. B. Nerlove, eds. *Multidimensional Scaling in the Behavioral Sciences*. In press.
- Textor, R. B. *A Cross-Cultural Summary*. New Haven: Human Relations Area Files Press, 1967.