Collecting Data in Two-Dimensional Euclidean Metric Space

Linton C. Freeman
Institute of Mathematical Behavioral Sciences,
University of California, Irvine

Both cognitive anthropologists and social network analysts typically collect data by asking subjects to perform some more or less standardized task. Subjects are sometimes asked simply for their free recall of names of objects from some domain like fruits, musical instruments, or personal friends (Rosch et al. 1976; Romney et al. 1979; Freeman et al. 1981; Gatewood 1983, 1984; Freeman et al. 1987; Freeman and Romney 1987).

Sometimes subjects are presented with a list and asked to choose the objects from the list that fall into the domain under investigation (Freeman and Freeman 1980; Krackhardt 1984; Fagan 1992). Or they are presented with the names of objects—each on a separate card—and asked to sort the cards according to some criterion (Miller 1969; Boster 1986, 1987; Freeman et al. 1988). Or they may be required to perform a triads task (Romney and D'Andrade 1964; Burton 1972; Weller and Romney 1988; Romney et al. 1993), or to rank them (Miller 1974; Killworth and Bernard 1976; Bernard and Killworth 1979), or perhaps to assign each of them a numerical position on a scale (Andrews and Withey 1974; Thompson 1975; Bernard and Killworth 1979).

Typically, data that have been collected by any of these means are used to construct some measure of the perceived distance (or closeness) between pairs of objects. When n objects are studied, the result is a square, $n - n$, object-by-object matrix in which each cell contains an index of the distance between the object designated by the row and that designated by the column.

Matrices of this sort are usually fairly complex. Often it is difficult or impossible to grasp their overall patterning (Friendly 1979). But if their structure is simplified—if the objects and their arrangement are projected onto an ordinary two- or three-dimensional Euclidean space—we can actually see their patterning. The aim, of course, is to retain as much information about the original distances between objects as possible while we are transforming them into this simpler form.

The use of various standard methods to perform such simplification is common. These methods include MDS, principal components analysis, and correspondence analysis (Weller and Romney 1990). They differ in detail, but they all transform distance data into Euclidean form.

Using any of these transformations rests on the faith that (1) the subjects' cognitive organization of the domain in question actually is both Euclidean and of small dimension, or (2) if it is not, the small-dimensional Euclidean representation imposes no serious distortion on that cognitive organization (whatever form it really takes).[1] But, if either of these beliefs is true, we could streamline the process that leads us to collect our data in one form, scale them, and then transform them to an entirely different form. Perhaps we could begin by collecting our data directly in Euclidean form.

At least for the two-dimensional case, collecting data in Euclidean form turns out to be both simple and straightforward. Following a suggestion by David Morgan,[2] I programmed a micro computer to collect Euclidean data directly from subjects. The program is called MAP.[3] It performs three tasks: (1) it instructs the subject, (2) it presents stimuli, and (3) it records detailed data about subjects' responses.

The program requires three files: (1) a labeling file to guarantee that each subject has a unique identification, (2) a file containing instructions for subjects and (3) a file containing the stimulus list. Each time the program is run, it updates the labeling file and it produces a uniquely labeled data file for a subject.

There is, of course, no restriction on the domains from which we might draw objects for study by this method. It could be used to see how subjects arrange objects of any sort, and it would be natural to use it to study subjects' perceptions of proximities or similarities when the objects were human beings. Thus, it could be used to study perceived group structure or the perceived structure of roles, traits, or social positions. Suppose, for example, we want to collect data about subjects' images of a domain made up of the names of 21 animals (Henley 1969). We would write a file containing a number (say 1), another containing instructions, and a third containing the list of animal names. Then the program would be run and a subject asked to sit down at the computer and to strike the key marked <Enter>. The computer might display an instruction screen like that shown in Figure 1.

Then, when the subject strikes a key, a work area is framed on the left of the screen and the stimulus list is randomly ordered and presented on the right. The cursor is displayed prominently in the center of the work area, and reminders of the commands are listed at the bottom. A working screen is shown in Figure 2.

At this point, only the four arrow keys, the G and D keys and the <ESC> key are enabled. Pushing any other key produces a BEEP sound and no other action. Data recording begins when the subject moves the cursor so that it covers the letter associated with an animal's name and hits the G key. At that point, the label for the animal dims and a copy of that animal's letter becomes locked to the cursor. That letter is moved with each use of an arrow key. It will follow the cursor movement until it is dropped by pressing the D key.

As the task progresses, data are recorded in a file. Records include both (1) the name of each object picked up along with the time at which the pick up occurred and (2) the location at which that object was put down as well as the time of the put down.

The most important feature of the MAP method is that it requires no data transformation. Each subject produces a set of distances between each pair of objects. And these distances are explicitly both two dimensional and Euclidean. Moreover, this new method provides an additional benefit by recording more and more detailed data than other procedures. It provides a record of the order in which objects are chosen by the subject, it records information on the number of times each object is relocated, and it retains precise information on the subject's timing in picking each new object up as well as in placing that object down in its chosen location.

Thus, as compared with standard techniques of data collection, the MAP procedure produces a more direct measurement of Euclidean distance and at the same time yields much richer and more detailed data about the subjects' responses. Rather than imposing a Euclidean space on data that are possibly non-Euclidean, MAP automatically yields a true Euclidean metric. At the same time, it generates enough detailed data to permit us to examine many of the micro aspects of the process in which objects are located in Euclidean space.

Notes

[1] This assumption is open to question. Indeed, to the degree that Eleanor Rosch (1978) and others are correct in characterizing cognitive structures as hierarchically arranged sets of categories, their structure approaches that of a mathematical form called ultrametric. But ultrametric structures—even simple ones—cannot usually be embedded in Euclidean spaces of small dimension.

[2] Personal communication, 1985.

[3] The MAP program was written in Q-Basic and compiled to run under DOS. On request, I will send a copy of both the source deck and the executable code. Write me at the School of Social Sciences, University of California, Irvine, CA 92717. Phone: 714-856-6698. Best of all, Internet: LIN@ARIS.SS.UCI.EDU.

Figure 1: An Instruction Screen.

```
This experiment is concerned with how people think about
animals.  You will be presented with a screen. The names of
several animals will be listed on the right.  Each is
identified by a letter, like "F".  On the left, is a blank
screen. You use the arrow keys to move the cursor around.
First move to an animal's letter, and hit G to grab it.  Then
move it to some spot on the screen and hit D to drop it.
Locate all the animals anywhere you want on the screen--the
rule is simple.  If a bunch of animals seem to be similar to
each other--if they resemble one another in an important way--
put them close together on the screen.  Animals who seem to
have less in common should be located farther apart on the
the screen.  Once you drop an animal's letter, you can always
go back and grab it again using G.  Move it to where you want
then ise D to drop it.  When you are satisfied that all the
animals are arranged in a pattern that reflects the patterning
of similarity, just hit the key marked "ESC" to finish.



Hit any key to continue...
```

Figure 2: The Working Screen.

```
                                   A BEAVER
                                   B HORSE
                                   C TIGER
                                   D ANTELOPE
                                   E SHEEP
                                   F CHIMPANZEE
                                   G CAT
                                   H CAMEL
                                   I GIRAFFE
                                   J CHIPMUNK
                ■                  K COW
                                   L GOAT
                                   M DEER
                                   N ELEPHANT
                                   O RAT
                                   P GORILLA
                                   Q LION
                                   R DOG
                                   S MONKEY
                                   T RABBIT
                                   U ZEBRA

MOVE = ↑ ↓ → ←    GRAB = G    DROP = D    FINISHED = <ESC>
```

References

Andrews, F. M. and S. B. Withey 1974. Developing Measures of Perceived Life Quality: Results from Several Surveys. *Social Indicators Research* 1:1-26.

Bernard, H. R. and P. D. Killworth 1979. Informant Accuracy in Social Network Data II. *Human Communication Research* 4:3-18.

Boster, J. 1986. Can Individuals Recapitulate the Evolutionary Development of Color Lexicons? *Ethnology* 25:61-74.

Boster, J. 1987. Agreement Between Biological Classification Systems Is Not Dependent on Cultural Transmission. *American Anthropologist* 88:429-436.

Burton, M. L. 1972. Semantic Dimensions of Occupational Names. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Shepard, R. N., A. K. Romney and S. Nerlove, eds., pp. 55-77. New York: Seminar Press.

Fagan, K. R. 1992. Influences of Memory and Cognition on Estimates of Acquaintanceship Volume. PhD dissertation, University of California, Irvine.

Freeman, H. E., A. K. Romney, J. Ferrera-Pinto, R. E. Klein, and Tom Smith 1981. Guatamalan and U.S. Concepts of Success and Failure. *Human Organization* 40:140-145.

Freeman, L. C. and S. C. Freeman 1980. A Semivisible College: Structural Effects on a Social Networks Group. In M. M. Henderson and M. J. MacNaughton, eds., *Electronic Communication: Technology and Impacts*, pp. 77-85. Boulder, CO: Westview Press.

Freeman, L. C., S. C. Freeman, and A. G. Michaelson 1988. On Human Social Intelligence. *Journal of Social and Biological Structures* 11:415-425.

Freeman, L. C. and A. K. Romney 1987. Words, Deeds and Social Structure: A Preliminary Study of the Reliability of Informants. *Human Organization* 46:330-334.

Freeman, L. C., A. K. Romney, and S. C. Freeman 1987. Cognitive Structure and Informant Accuracy. *American Anthropologist* 89:311-325.

Friendly, M. 1979. Methods for Finding Graphic Representations of Associative Memory Structures. In C. R. Puff, ed., *Memory Organization and Structure*, pp. 85-129. New York: Academic Press.

Gatewood, J. B. 1983. Loose Talk: Linguistic Competence and Recognition Ability. *American Anthropologist* 85:378-387.

Continued on page 12

Continued from page 11

Gatewood, J. B. 1984. Familiarity, Vocabulary Size, and Recognition Ability in Four Semantic Domains. *American Ethnologist* 11:507-527.

Henley, N. M. 1969. A Psychological Study of the Semantics of Animal Terms. *Verbal Learning and Verbal Behavior* 8:176-184.

Killworth, P. D. and H. R. Bernard 1976. Informant Accuracy in Social Network Data. *Human Organization* 35:269-286.

Krackhardt, D. 1984. A Social Network Analysis of the Effects of Employee Turnover: A Longitudinal Field Study. PhD dissertation, University of California, Irvine.

Miller, G. A. 1969. A Psychological Method to Investigate Verbal Concepts. *Journal of Mathematical Psychology* 6:169-191.

Miller, M. L. 1974. A Comparison of Judged-Similarity, Trait Inference and Trait Rating Tasks with Regard to the Multidimensional Structure of Personality Traits. PhD dissertation, University of California, Irvine.

Romney, A. K., D. D. Brewer, and W. H. Batchelder 1993. Predicting Clustering from Semantic Structure. *Psychological Science* 4:28-34.

Romney, A. K. and R. G. D'Andrade 1964. Cognitive Aspects of English Kin Terms. *American Anthropologist* 66:146-170.

Romney, A. K., T. Smith, H. E. Freeman, J. Kagan, and R. E. Klein 1979. Concepts of Success and Failure. *Social Science Research* 8:302-326.

Rosch, E. 1978. Principles of Categorization. In E. Rosch and B. B. LLoyd, eds., *Cognition and Categorization*, pp. 28-48. Hillsdale, NJ: Earlbaum.

Rosch, E., C. Simpson, and R. S. Miller 1976. Structural Bases of Typicality Effects. *Journal of Experimental Psychology: Human Perception and Performance* 2:491-502.

Thompson, R. W. 1975. Gratification Patterns in Buganda: An Explanation of Intracultural Diversity. *American Ethnologist* 2:193-206.

Weller, S. C. and A. K. Romney 1988. *Systematic Data Collection*. Newbury Park, CA: Sage.

Weller, S. C. and A. K. Romney 1990. *Metric Scaling: Correspondence Analysis*. Newbury Park, CA: Sage.

NEW PRODUCTS FOR TEACHING

✔ For teaching statistics, consider a package called *Social Statistics Using MicroCase* published by MicroCase Corporation. The package includes a text and a lab workbook (by William Fox), a statistical analysis program, and a diskette containing two data sets (one with data on 81 variables about the 50 states in the U.S. and the other with data on 103 variables from the U.S. General Social Survey of 1990). The price is $47.50, including shipping. Contact MicroCase Corporation, 1301 120th Ave. NE, Bellvue, WA 98005. Tel: 800-682-7367. Fax: 206-635-0953.

✔ For teaching intro to cultural anthropology or for a course on cross-cultural research, take a look at another MicroCase product called *Exploring Preindustrial Societies*. This is an instructional data base with data on 94 variables for the 186 societies in the Standard Cross-Cultural Sample, plus data on 37 variables for 124 societies in North America. This package costs $95.00. The data can be displayed with MicroCase's *ShowCase* presentational software. The presentation software contains a statistical analysis program, so you can ask questions about the data as you lecture and the answers are projected onto the display screen. The *ShowCase* software also costs $95. During 1994, readers of CAM are offered both the presentation software and the package on exploring preindustrial societies for $95. Contact MicroCase Corporation, 1301 120th Ave. NE, Bellevue, WA 98005. Tel: 800-682-7367. Fax: 206-635-0953.

✔ For building your CAI (computer assisted instruction) materials, consider *PC-CAI* from TexaSoft. This easy-to-use program lets you create your own interactive tutorials and tests. You use a simple editor to build your own CAI programs, and the package comes with sample programs so you can see what the finished products should look like. You can use sound effects and graphics to make tutorials interesting. The Plus version of the program has a module that lets you build a data base of multiple-choice questions that you can pull into interactive tests. Then, students can sit at the computer and take tests at their own pace. The program keeps track of students' scores. TexasSoft publishes a newsletter with helpful hints for users of this program. Cost is $69 plus shipping for the basic version, and $99 plus shipping for the Plus version. Contact TexaSoft at P.O. Box 1169, Cedar Hill, TX 75104. Tel: 214-291-2115. Fax: 214-291-3400.

PUBLICATIONS

The Center for Electronic Texts in the Humanities (CETH) was established a couple of years ago at Rutgers and Princeton Universities. The Center produces a newsletter. Contact CETH@ZODIAC.RUTGERS.EDU on the Internet.

The Anthropology and Sociology Section (ANSS) of the Association of College and Research Libraries (ACRL) publishes a newletter for its members. For information on joining ANSS, contact DIXON@HARVARDA on Bitnet.

The *Qualitative Research Methods Newsletter* is published by the Department of Health Services Studies, Tata Institute of Social Sciences. Write to P.O. Box 8313, Deonar, Bombay, India.