Chapter 6 TN 32: PRACTICAL APPLICATIONS OF CLUSTER ANALYSIS[1]
By Jay Beaman and Sandy Lindsay

Purpose

The purpose of these notes is to present to planners and decision-makers some practical implications of results that can be obtained using cluster analysis. It also provides insight into why factor analysis should not be used to determine "activity packages" for a heterogeneous population.

Introduction

This paper concentrates on presenting different ways cluster analysis results may be used by planners and decision-makers. A paper on the *methodology* of cluster analysis by Romsa et al. has already been received by Parks Canada and some results of Romsa's methodological investigation have been published. Specifically, it is in the *Journal of Leisure Research* (Romsa, 1973) that Romsa and his colleagues at the University of Windsor have reported on the use of cluster analysis for deriving activity packages from the CORDS National Survey Information. Given Romsa's methodological effort and given the Burdge and Field (1972) and Tatham and Darnoff (1971) general discussions, a very practical discussion is now required concerning the ways in which Romsa's results, or the results that could be arrived at by others, can be used by planners and decision makers.

Since there is a difference between the kind of cluster analysis carried out by Romsa and his colleagues, and the kind of work that has some times been called cluster analysis by other researchers, the reader should look carefully at the example presented in the next section in order to understand the kind of cluster analysis referred to. Briefly, the cluster analysis approach discussed here *does not look at intercorrelations between activities,* but rather clustering refers to a relation defined by individuals participating in similar collections of activities: clustering is of individuals on the basis of activities in which they participate, rather than on any intercorrelation between activities over some collection of people (see Figure 1).

Concern is not with people grouping *themselves* in the sociological sense of forming social groups in which there is interaction; at least this is not necessarily the case. Here, grouping refers to people being in collectivities; people are indicated to be members of particular collectivities defined by participation in a certain set of activities. The set of activities that is related to a collectivity is said to *be* or *define* the activity package for the activity. In the following discussion the existence of collectivities defined on the basis of particular people participating in a particular collection of activities defined by their "activity packages" is accepted.

The Use of cluster Analysis with Park Users Survey Information to Obtain Results That May Be Of Use to Planners

One of the problems that arises in trying to understand the results prepared from park user surveys data on activities participated in is (1) the results presented rarely reflect the total loading of facilities and (2) the results, e.g. tables prepared from raw data, do not give a clear picture of the multi-faceted participation characteristics of the users of facilities. Assume here that the user information is collected as part of an entrance or exit survey. In reality one of the major dimensions of analysis concern may be the mix of activities in which people participate. Information on the collection of activities in which an individual participates and on the characteristics of participants simply loses the dimension of activity mix when presented on an activity by activity basis. Even cross-tabulations of activity against activity only give a very limited insight into individuals' activity mixes.

Participation analysis that goes beyond a simple analysis of facility loading requires a technique that reflects the interrelationships between the activities in which individuals participate. The use of factor analysis to examine intercorrelations defined by people's participation in a number of activities

violates the basic assumption of defining "activity packages" (see the Appendix). Rather, one must adopt a true cluster analysis technique that will break the user population into relatively homogeneous groups on the basis of those activities in which individuals participate.

Figure I: HYPOTHETICAL CLUSTER ANALYSIS RESULTS FOR A PARK USER ENTRANCE OR EXIT SURVEY ACTIVITY INFORMATION

Cluster #	Activity / Person #	INCOMPATIBLE 1, 2, . . ., N	N+1, . . ., M	HIGHLY COMPATIBLE M+1, . . ., Z-1, Z
1	(10% of People)*	▨	**	
2	(20% of People)	▨	▨	
3	(20% of People)		▨	
4	(20% of People)	▨		▨
5	(20% of People)		▨	▨
6	(10% of People)			▨

* Person #'s are not given , but one could consider that say persons 1, 8, 18, etc. are in cluster 1.
** Single line indicates the primary activities of importance in defining the cluster indicated.

For some people the preceding discussion may be rather abstract and even appear obtuse. So, assume that three types of activities are provided in a park: (1) activities which are not compatible with the theme or general purpose of the park; (2) other activities that are not offensive to the purpose of the park but not exactly in line with its theme; and (3) activities that are consistent with the purpose of the park. If a cluster analysis of park user survey data on the basis of the activities in which visitors participate were carried out for a given park , then one could see how people 's participation is divided among the sets of activities just noted.

Figure I presents hypothetical results of a cluster analysis of summer activities in a given park, suggesting that there are six clusters and thus six types of park visitors. The figure shows that 10 per cent of the people visiting the park are in Cluster 1, activities incompatible with the park's theme. The figure also shows that 20 per cent of visitors show an interest both in incompatible activities and in activities that do not tie strongly to the theme of the park. The third group indicated is not involved, to any large extent, either in incompatible or in highly compatible activities. The fourth cluster of people noted are those that present an extremely problematic pattern of clustering by participating in both incompatible and highly compatible activities. The planner might wish to serve these visitors by moving incompatible activities out of the park (to be supplied by the private sector) while leaving the highly compatible activities in the park.

Finally, the last two clusters indicated in the figure involve respectively 20 and 10 per cent of the park visitors. These are the clusters made up of people whose participation focuses on *highly compatible activities.* However, cluster 5 involves a high level of participation both in activities that are not highly related to the park theme (but not incompatible), and in incompatible activities.

One should note that while Figure I may clarify how the results of a cluster analysis of park visitor survey data might be of use to a planner, the figure oversimplifies the kind of clustering pattern that might be expected in a cluster analysis. This oversimplified clustering pattern results in an "obvious" interpretation of the results, which would rarely occur. Certain sets of activities containing some highly compatible activities and some activities of the other two types may turnout to define

activity packages for people. The activity package just alluded to may have few activities in common with another activity package that also involves activities that run from incompatible to highly compatible. The existence of two clusters that both involve incompatible and highly compatible activities, but have no overlap in activities, may reflect an orientation of some people toward physically demanding activities and others toward passive activities when *both incompatible activities* and *compatible activities* are *considered.*

In terms of the example just cited, one can see that a dimension of clustering not suggested by the figure may be relevant for consideration in planning. However, one may wish to ignore the active-passive split and, for planning analysis, combine clusters to match the pattern shown in Figure I by combining cluster s that are identified by a computer analysis of park user survey data.

Cluster Analysis of Household Survey Information and Planners: A Perspective

The planning considerations to which the development of activity packages from data collected from people at their homes is claimed to be important, are (a) the equity factor; (b) planning for specific are as or population; (c) future research; and (d) predictions of behaviour. This data can relate to participation in activities at a particular park or to participation in all recreation activities. This paper is working in the latter context-data collected on participation patterns at any park, private facility or at home.

The reader is asked to refer to Figure II, which is very similar to Figure I, except that the focus of attention is not on activities that take place in a given park but rather, on a set of activities which for some reason was chosen to be considered in a National or Regional Survey. Thus, in contrast to the way activities are divided in Figure I, the list of activities dealt with may reasonably be broken into active or passive; into social, cultural or athletic; or into what ever classification is appropriate for the policy related research and /or planning of the Administration Unit involved in a given analysis.

Figure II: HYPOTHETICAL CLUSTER ANALYSIS RESULTS FOR AN ANALYSIS OF NATIONAL SURVEY DATA

Cluster #	Person #	ATHLETIC		CULTURAL		SOCIAL	
		Active	Passive	Active	Passive	Active	Passive
		...N	N+1...M	M+1...P	P+1...R	R+1...	...Z
1	(7.5%)*	▓		▨	▨	▓	▓ ***
2	(7.5%)	▓	▨ **			▓	▓
3	(15% of People)		▨	▨	▨		
4	(20% of People)			▨	▨	▨	
5	(20% of People)		▨	▨			▨
6	(15% of People)			▨	▨		▨
7	(15% of People)		▨	▨		▨	▨

* Person #'s are not given, but one could consider that, say, persons 1, 8, 18, etc., are in cluster I.
** Single line and /or crosshatched indicates the primary activities of importance in defining the cluster indicated.
*** In the text reference is made to tent camping, trailer camping, fishing, hunting and driving for pleasure. The dark gray areas indicate participation in these activities. Tent camping, fishing and hunting are assumed to be ATHLETIC-ACTIVE, so to distinguish participation in these activities the dark gray areas appears under ATHLETIC-ACTIVE for cluster s I and 2.

Equity

Cluster analysis has value in considering equity because by breaking up the population according to the activity packages in which they participate; one can see the equity with which activities are provided to various collectivities of people. It is not suggested that all the activities listed in a survey should be provided in such a way that each collectivity associated with an activity package receives the

same dollar input from a given agency's Parks and Recreation Expenditures (receiving the same dollar input is not even necessarily possible); there is no suggestion that equity is reflected if planning involves allocating money to activities so that some kind of balance is kept in the money benefits that people with various activity packages "receive". However, if one is placing money into a number of activities, then one can look at activity packages for the collectivities into which a population is divided and see in which packages the activities just noted occur and, in which packages they do not: one can see *what* proportion of the people are being served by a given set of activities and in some sense *how*.

If the activities noted in the last paragraph occur in only two activity packages involving 15% of the population (see Figure II, Clusters 1 and 2 marked by cross hatched) *and these are the only activities into which government puts money,* then there is obviously some inequity. But if one government agency puts money into supporting tent camping, trailer camping, fishing, hunting and driving for pleasure while there are other activities into which other government agencies put money or if, for example, provincial responsibility is related to athletic activities and the federal agency is examining the activities into which they are putting money, then inequities that may be reflected by supplying facilities for activities in which only 15 per cent of the population participate may well be irrelevant to the national level planner responsible for planning provision of specific facilities.

The point is that, by defining groups of people according to the activities in which they participate, one gets a clear and multi-dimensional idea of who is receiving what support when money is put into several activities or into a large number of activities. Thus, when cluster analysis results that define activity packages are available, one avoids the error of thinking that because 15 per cent of the people in a population participate in one activity (e.g. an activity common to both Cluster s 1 and 2) and 22.5 per cent participate in another (an activity common to both Clusters 2 and 3), then monetary input in the two activities services 37.5 (22.5 per cent plus 15 per cent) per cent of the population. (Actually, in the example cited, 30 per cent, that is 7.5 per cent plus 7.5 per cent plus 15 per cent, of the population is served.) The author maintains that this kind of fallacy permeates the thinking of many officials who believe that planning for activities should be done on the basis of the total man-days of participation in each activity or *on the basis of the proportion of the population participating in each of a number of activities.* But there is a need to distinguish between planning based on a fallacious use of information and allocation of resources where equity considerations may be relevant from a geographical distribution of expenditure but not from some other perspective.

Given the points made above, the author claims that when cluster analysis results are available for a given city, it *can be* abundantly clear if the cluster analysis results indicate that the recreation plan for the city endorses or results in a physically active minority *of* the people receiving drastically disproportionate amounts of recreation dollars, in comparison to those people who may have activity packages which contain relatively few, say, passive activities. Romsa 1973 presents activity packages for Quebec that show the possibility that most recreation dollars go to a few people.

Planning Implications

Planning implications that can be derived from cluster analysis that do not relate to equity considerations (a policy consideration) may be understood easily, given the preceding discussion. If a population can be broken into the kind of collectivities just described (collectivities based on activity packages), then the population has been broken into natural units. This is because the various activities in an activity package may inter-relate. There may be a trade-off between activities (see Hendee and Burdge, 1974; Romsa, 1973; Bishop and Witt, 1972;Burton, 1971; Gillespie, 1973 as well as articles cited in these sources) or some other relationship (Beaman and Leicester, 1970 discuss the use of complementarity). An understanding of such relationships between activities, as well as knowledge of the sizes of those collectivities that have an activity package making them a market for a facility package, allow the planner to recognize facility packages that are wanted by various segments of the population. Also there is the possibility that the planner can recognize the kind of facility or activity

"deficiencies" that may reasonably be left when one activity is allowed to substitute for another (Burton, 1971).

With the activity packages for the various "groups" in the population in mind, the planner can begin to see why large segments of a population will not participate in certain facilities if a certain set of activities is provided. Given such information, planners can inform politicians of the "real" significance of a situation. When a political ruling (policy) on what will or will not be provided is to be examined, the planner who has cluster analysis results relevant to be given policy, can give the politician a clearer idea of *what* he is really asking about and *why,* with regard to a given policy, than is usually possible with out cluster analysis results. The important point is that the planner who has cluster analysis results can inform the politician and/or manager of the consequences of given actions in a much more intelligent way than can be done when the only figures that can be provided are figures that show that some people do not participate in one activity or another. Cluster analysis thus allows the planner to better understand the structuring of behavior. By allowing a better understanding of the structuring of behavior, cluster analysis makes possible a realistic and comprehensive discussion of planning options. Thus politicians and /or managers/decision makers can make better decisions regarding facility provision than are usually made today (at least the potential is there).

Futures Research

The preceding section could have started with the statement that: to say *people* behave in such a way as to define activity packages is almost a tautology. The import of the assertion that people 's participation in activities defines activity packages really becomes clear when one recognizes that *for purposes of futures research, activities do not have a future.* Rather, activity packages have a future. It is the people living today and having similar patterns of behaviour who, as they move into the future, may be expected to modify their behaviour so that activity packages shift as new activities shift into activity packages, or some activities shift out of some or *all* activity packages.

The basic fallacy in much current recreation/leisure oriented futures research is that it approaches making projections into the future *by projecting the future of activities, activity-by-activity:* researchers approach the future of an activity by asking what is going to happen to this activity and, in doing so, fail to recognize that it is not the activity that goes into the future and behaves in some way. It is people who, by living from day to day and year to year, go into the future and either change or maintain their behaviour and attitudes! Thus, cluster analysis of participation data results in appropriate units, the activity package and the collectivities associated with activity packages, with which to confront Delphi panels when they are asked to make predictions of the future.

In other words, it is claimed that the task of a Delphi panel should be to make projections for activity packages (1) in terms of the number of people in the population that will be in given collectivities associated with given activity packages and (2) make projections that indicate what activities may drop from a package, and what new activities may be expected in the package. they must also suggest totally new kinds of activity packages that may arise. Then the panel must put the results of projections for all activity packages (new or existing) together to get an activity by activity perspective for the future, if for some reason there is concern with having an activity-by-activity perspective.

The simple activity-by-activity approach to projection confuses the issue of what is really happening over time by removing the insight that may be gained in recognizing a structuring of participation in activities in our society. For this reason, in order to make reasonable projections, one needs to recognize the fact that certain collectivities can be defined and used in making futuristic projections.

Predicting Behaviour

Finally, a consideration related to the points made above is that cluster analysis offers insight into that is important in predicting behaviour. The predictions that one needs to make in planning or

policy evaluation may not be predictions of the future. Examination of the characteristics of a population in an are a where survey results are not available may indicate that people are such that certain activity packages may be expected to exist if certain facilities are created (i.e. if certain policies are followed!). Specifically, until we understand the behaviour of people in terms of activity packages, there will be a tendency to treat the prediction of behaviour on the basis of individual activities regardless of whether a large array of alternative activities is supplied.

At present it does not seem clear that there is a methodology by which cluster analysis results can be used to make predictions of what will happen in terms of people 's participation in certain activities in a given are a under various policy options. Never the less, cluster analysis is the only methodology that defines the kind of activity packages that are here argued to be of importance. Thus, the results of cluster analysis offer information that is of value in making policy related predictions and therefore it is presently the only methodological tool known to the authors that has any promise of improving our productive ability by increasing our understanding of activity substitutability and complementarity.

Conclusion

The discussion in the preceding sections has presented considerations related to a number of practical applications of cluster analysis. Though it would be desirable if all of these are as of application could be explained in terms of practical examples of what *has actually been done,* the practical application of cluster analysis to define activity packages to be used in the ways described has simply not taken place. In fact, it is the purpose of this paper to prompt practical planning work using cluster analysis results by setting down the suggestions and guidelines noted here.

Appendix

To understand why factor analysis should not be used to derive "activity clusters" one need only understand a few basic considerations. Factor analysis should be used only when the data for analysis are considered to have underlying dimensions common to all people (recall the early studies on single or multiple dimensions of intelligence) or when any sub-group of the population selected for analysis has the same dimensions as the population as a whole. Data should not be considered as having an internal structure such as the one suggested by the clustering illustrated in Figure I. Figure I suggests that the people in a community or nation (as suggested by Romsa's research on Canada Quebec results reported in the Journal of Leisure Research, 1973) may be broken up into collectivities of people on the basis of the activities in which individuals participate.

Each one of the collectivities derived by cluster analysis is characterized by the intercorrelations among the activities that define the "cluster of people" that have the *activity package* (defined by the activities in which people in a given collectivity participate). Given that a number of collectivities encompass all the activity packages of a community, it may be asked what the intercorrelations of people's participation intercorrelations tell one about participation in activities. If correlations for the population are processed in factor analysis values depend on the relative sizes of various collectivities in a population. Therefore, they are not yielding information about the individual collectivities per se. If certain subgroups of the population of a city are selected for analysis, one does not get the same factor structure for the population as a whole unless dealing with a very particular kind of sample designed to be representative of the city as a whole: when a population is subdivided on the basis of age, sex, education, income and other variables, the relative balance between collectivities changes in various sub-groups. This fact is confirmed by cluster profiles derived by Romsa (1973) and results presented in Currie (1973).

The point being stressed here is that factor analysis is the most appropriate tool to use in looking for structure in data only when (see Horst, 1965 or Harmon, 1959) the invariance of factor structure for subgroups in the population condition holds. Specifically, if the assumption of invariance of correlations based on people 's participation in activities is to hold, the population that is being subjected to a factor

analysis must not be structured as Romsa's cluster analysis shows the population of Canada to be structured with respect to its participation in outdoor recreation activities. When such a structuring exists, factor analysis is not an appropriate technique for learning something about the structuring of participants in the population of concern.

As well as the fairly general theoretical considerations just noted, a more detailed critique of the factor analysis "clustering technique" could be based on reference to a number of points. In particular, the instability (loose definition) of structures defined by varimax rotations or other factor rotation algorithms is one point that should be considered. Along this line, it can be noted that algorithms have been developed to relate factor structures derived on one set of data with factor structures derived on another set of data. However, it has also been shown that because of the nature of the transformations involved, factor structures from two sets of data can often be related even if there is little relationship.

Notes

1. Burdge and Field (1972) have cited the need for a cluster analysis approach to the processing of participation data. Tatham and Darnoff's work on cluster analysis (1971) deals with some of the issues presented here. Their work is of particular relevance for *planning implications.* Notions on *equity,* however, are not developed in quite the same way as in the present paper. Their discussion does not deal with *prediction of behaviour and futures research* in a way that is similar to the perspective endorsed here.

2. The Technical Note will be available in the fall of 1974. However, a "Preliminary Report on CORD Technical Note No. 10 a Search for Structure in the patterns of Participation of Canadians in Outdoor Recreation Using Cluster Analysis Methods: The Windsor Work is now available.

3. Van Doren and Heit (1973) in their review of the first three years of the JLR note that few papers have been published in the major content area of " planning applications of research and design". Hopefully the content of this article is" planning application" rather than methodology, even though there is a major implicit concern with the use of a methodology "cluster analysis".

4. There is abundant literature in sociology that has to do with socialization, role learning and how these relate to social organization, and social persistence (as opposed to social change). Literature in psychology on learning, related to how "needs and desires develop" suggests that projections should be based on what people *will do* rather than what will "happen" to an activity.

References

Beaman, J. and J. Leicester, 1970. "Supply and Use of Facilities-Theory and Modeling Considerations Applied to Recreation Facilities". Prepared for Research Division of parks Canada, Ottawa.

Bishop, D. W. and P. Witt, 1972. "Report to Department of Lands and Forests Re TORPS (discussion of substitutability concepts as they relate to the TORPS Model)."Toronto: Tourism and Outdoor Recreation planning Committee.

Burdge, R. J. and D. R. Field, 1972. "Methodological Perspectives for the Study of Outdoor Recreation". Journal of Leisure Research, 4 (winter): 63-72.

Burton, T. L. 1971. *Experiments in Recreation Research.* London: George Alien & Urwin Ltd.

Gillespie, R. 1973. "A Definition of Canadian Outdoor Recreation Types Through the Application of Factor Analysis to Participation data: An experiment in Recreation Research". University of Waterloo: Thesis submitted in partial fulfillment of the requirements of the degree of Bachelor of Arts.

Hendee, J. C. and R. J. Burdge. 1974. "The Substitutability Concept; Implications for Recreation and Management". Forthcoming in the Journal of Leisure Research.

Romsa, G. 1973. "Deriving Outdoor Recreation activity Packages for Canada". Published in the Journal of Leisure Research, fall. 1973, Vol. 5 (based on a larger report of the same title that was prepared while under contract to Parks Canada, Ottawa).

Tatham, R. L. and Dornoff, R. J., 1971.*Market Segmentation for Outdoor Recreation,* Journal of Leisure Research, 1971.

Van Doren, C. S. and M. J. Heit, 1973."Where It's At: A Content Analysis and Appraisal of the

Journal of Leisure Research". Journal of Leisure Research, 5 (winter): 6773.