On Household Groups and Housing Occupancy

by

John R. Miron

Research Paper No. 107

The Centre for Urban and Community Studies

University of Toronto

May, 1979
Reprinted June 1980

ISSN 0316-0068
ISBN 0-7727-1266-2

ABSTRACT

A major problem in forecasting future housing demand concerns the way in which such forecasts are based on present patterns. In a typical exercise, households are divided into groups, the present occupancy pattern of each household group assessed, and forecasts made on the basis of those contemporary occupancy patterns applied to forecasts of the number of future households in each group. While numerous studies have been made of the housing occupancy patterns (e.g., type of dwelling, tenure) of particular household groups, there has been little research concerning the designation of these household groups. Presumably, a household group should include households with similar patterns of housing occupancy, should have an occupancy pattern which is relatively stable over time, and should be distinct from other household groups.

In this paper, the problem of choosing appropriate household types is considered. A new clustering technique is introduced: the SR method. This method is based on an iterative binary splitting and recombination of clusters of households. Two data sets are used; the 1971 Census of Canada Public Use Sample and the 1974 CMHC Survey of Housing Units. A two-way classification, mode of tenure by type of dwelling occupied, forms the basis for the identification of household groups. It is concluded from these two empirical studies that (i) a very parsimonious set of variables may be used to define household groups which are substantially different from one another, (ii) these groupings are quite similar for both the 1971 and 1974 data sets, (iii) at the same time, there remains a substantial amount of intra-group variation in occupancy choices, and (iv) the SR method of household group identification appears to offer substantial promise.

On Household Groups and Housing Occupancy*

In recent years, there has been a small explosion of empirical research on occupancy patterns. Using large microdata samples, researchers have looked both at the tenure and at the type-of-dwelling patterns of household. These studies typically relate occupancy patterns to such characteristics of the household as the age, sex, and marital status of the head, number of persons in the household, and family or non-family status of the household.

Although similar in this respect, these studies vary widely in their specification of household groups. Consider the age of head characteristic as one example. Maisel (1966) uses two age classes; those headed by a person under 65 and those headed by one 65 or older. Shafer (1978) uses three classes; those headed by a person under 30, by a person 30-59, and by someone over 59. Struyk and Marshall (1975) use four classes; the under 30's, the 30-44's, the 45-65's, and the over - 65's. Similar kinds of differences can also be found when looking at other household characteristics.

That different specifications are used by different researchers raises the question of how much the researcher's choice of household groups conditions the results obtained. Presumably, the objective behind grouping is to identify clusters of households having similar occupancy patterns. An inappropriate grouping can possibly mask an important source of heterogeneity. However, in these studies, one typically has no way of assessing the homogeneity of the groupings used.

It was in regard to this concern that the research described in this paper has its origins. The original question posed was "what groupings of

---

*The financial support of the Centre for Urban and Community Studies and the outstanding research assistance of M.H. Matusoff are gratefully acknowledged.

households can be identified which have similar tenure and type of dwelling patterns?". As with most research, upon investigation this original question became intertwined with the following set of more specific questions:

Q1: In what sense is it meaningful to think of homogeneous household groups?

Q2: What alternative methods are available for the identification of homogeneous household groups?

Q3: What kinds of homogeneous groups can be identified and on what basis can their existence be rationalized?

Q4: How stable are these groupings over time?

In this paper, each of the above questions and the issues raised by them will be discussed in turn in the following four sections. The first two questions will be considered at a conceptual level. Several subsidiary issues are raised by them and these will be assessed in Sections 1 and 2.

In Section 3, a new clustering technique is introduced; one which is referred to as the Split-Recombination (SR) Method. The SR Method uses a sequence of binary splits and recombinations in deriving housing occupancy groups. From a 1974 Survey of Housing Units (SHU) sample, 14 groups of households are identified with this method. These groups partition the set of sampled households, are all statistically distinct, and are such that no additional distinct groups can be formed by a binary splitting of an existing group. In this sense, one can say that a relatively parsimonious set of distinct household groups has been identified.

In Section 4, the SR Method is applied to a 1971 Public Use Sample to derive 19 distinct household groups. These groupings are broadly similar to the SHU groupings. At the same time, there are differences between the two sets. In fact these differences can be rationalized in terms of differences between the samples. However, the question of temporal stability of the groupings is not resolved.

1. Some Conceptual Issues

Let us turn our attention to the first question raised above. "In what sense is it meaningful to think of homogeneous household groups?" There are at least three aspects to this question. One aspect has to do with the meaning and measurement of occupancy patterns. Another has to do with the attributes of the household used in identifying groups. The third aspect has to do with relationships among the number of attribute classes to be considered, the size of the sample, and the number of household groups to be identified. These are now considered in turn.

1.1. The Meaning and Measurement of Housing Choice

When we say that the households in a group have similar occupancy patterns, what is meant? There are at least three elements in such a statement; a basis for defining the group, a notion of causality, and finally a formulation of the occupancy pattern being considered.

Let us turn to the first of these. Suppose that a group of 100 households is somehow defined. Suppose further that we are interested in the tenure pattern of these households and find that 85 households are owner-occupiers while the other 15 are renters. On the basis of this it might be concluded that the households have "similar" tenure patterns with a high propensity to be owner-occupiers. A critic might look at this and say 'But wouldn't the group be even more homogeneous and have "more similar" tenure patterns if somehow the 15 renter households could be allocated to some other group?'. Yes, you might say, but not if in so doing we end up creating by tautology only two groups; one of renters and one of owner-occupants.

By such an argument, we have come to recognize that a group must be predefined (and on some basis other than occupancy pattern) before one can speak meaningfully of groups of households who have similar occupancy patterns

This problem will be readdressed in the following section. However, it is noteworthy here that groups are commonly defined in terms of family life cycle attributes such as age, sex, and marital status of head, size of household, and ages of children present. In other words, groups are defined on the basis of those household attributes which are usually seen to create a demand for a particular kind of housing.

This introduces the second element; the issue of causality. In a demand-based approach to household occupancy patterns the household is generally taken as given. It is commonly assumed in these studies that the supply of available housing does not affect household formation. The effect of the pricing and availability of housing on a household's choice of housing has been studied but the effect on household formation itself has been largely ignored.[1] While this supply effect may be negligible for a household consisting of a family of parents and young children, it can be very important in the living arrangement choices of (i) families not presently maintaining their own households, (ii) non-family individuals, and (iii) maturing children at home-leaving ages.[2] Thus, the standard (demand-based) approach is to see two distinct processes, household formation followed by housing choice, with the supply of housing affecting only the latter. A more comprehensive approach would see household formation and occupancy patterns as at least partly interdependent with both affected by supply considerations. In the empirical portion of this paper, the charade of the demand-based approach is maintained though it rests somewhat uncomfortably. An example of this issue will be raised in Section 3.

---

[1] The only major exception is Beresford and Rivlin (1966)

[2] We adopt here the Census definition of a family which is (i) a married couple with or without never-married children still living at home or (ii) a parent with one or more never-married children still living at home.

The final element has to do with the formulation of the occupancy pattern being considered. When we say that households in a group have similar occupancy patterns, what do we mean? Does one mean, for instance, that each household resides in an owner-occupied bungalow of 150-175 m^2 constructed between 1950 and 1960 with a two-car attached garage, a lot area of 400-500 m^2, an in-ground swimming pool, central air conditioning, three bedrooms, two bathrooms, a finished basement, natural gas heating, and a paved driveway? Typically, a researcher in this area would say that this is not what is meant by a "similar" occupancy pattern. Maisel (1966), Struyk and Marshall (1974, 1975), Kain and Quigley (1975), Doling (1976), and Li (1977) have looked only at tenure patterns and at that for only two classes (owner-occupiers, renters). Shafer (1978) has looked at the split between single dwelling buildings and multiple dwelling buildings. David (1962) alone among these authors has investigated more than one aspect of housing occupancy: he examined the tenure, size of dwelling (number of rooms), and dwelling value per room choices. Thus, for the most part, researchers have looked at just one aspect of occupancy.

It is also clear that household groupings that are relatively homogeneous (or appropriate) for one type of occupancy pattern may be heterogenous (or inappropriate) when looking at another type. For example, the groupings used by Shafer to look at the type of dwelling occupied may rightly be quite different from the groups Li would use in looking at tenure. At the same time however, many of these occupancy patterns have a high covariance. For example, owner occupancy is quite often associated with single dwelling buildings and larger size of dwelling. Thus, although Shafer and Li for example would purport to look at different kinds of patterns, they may in fact be studying essentially the same phenomenon in which case the same household groupings are appropriate.

In the empirical work in this paper, an attempt is made to identify households with similar types of dwelling and tenure patterns. Two tenure statuses (owner-occupier and renter) are considered. Dwelling units are described as either low-density (detached, semi-detached, or row dwellings) or high density (duplex, flats, or apartments). Thus, there are four different occupancy categories

(1) Owner-occupied, low density

(2) Owner-occupied, high density

(3) Rented, low density

(4) Rented, high density

The most frequented categories are (1) and (4) although for some specific household types there are relatively high incidences of (2) and (3). By considering both tenure and type of dwelling, it is hoped to bridge some of the earlier empirical work.

1.2. Relevant Household Attributes

In this section, the problem of choosing attributes of households which can then be used to define groups is discussed. Most the literature in this area tends to focus on the one-family household (i.e., a household consisting of one census-defined family without any additional persons present). These currently form about 2/3 of all households in urban areas and thus are quite important. However, the remaining 1/3 of households are (i) one-family households with additional non-family (though often related) persons, (ii) multiple family households with or without additional non-family persons, (iii) one-person households, and (iv) multiple nonfamily person households. Of these, the most common currently tends to be the one-person household which forms from about 1/10 to 3/10 of all households. There is very little research literature though on the formation and housing occupancy patterns of either

the one person household or any of these other types.

1.2.1. The One Family Household

Let us therefore turn our attention first to the housing occupancy patterns of one-family households. A commonly-held view is that when a family moves it makes a choice that is based principally on a tradeoff between the "space" it would like to occupy in its new housing and what it can afford. The space a family might like to occupy is often seen to be related to the family's size, characteristics, and wealth.

In terms of household size, the presence (or anticipation) of children commonly brings about a change in the preferences of the family unit. Usually, the arrival of children creates a need for additional bedrooms, both indoor and outdoor recreation areas, and access to schools and other public facilities. Some researchers such as Shafer (1978) and David (1962) emphasize that these needs change with the ages of the children and therefore divide one-family households into those with any young children and those with older children only. Other researchers assume that the presence and ages of children can be proxied by the age of the head of household.

Another characteristic of the family unit which affects household size and housing choice is the presence or absence of a spouse. This of course can occur through widowhood, divorce, or separation. It is common in the case of divorce or separation for the wife to retain custody of children and for the husband to then become a nonfamily individual. Widowhood of course can result in either spouse being left with children although there is a low incidence of widowhood for men who are still young enough to be a head of family. In addition, it is noted that the practice of designating the husband as "head" whenever he is present had continued up until the 1976 Census of Canada. Therefore, when examining census data prior to 1976, a

good indication of the presence of only one spouse is the headship of a family unit by a woman.

The final set of attributes usually considered relate to income and wealth. There are several problems in trying to relate the income or wealth attributes of a family either to housing choice when the family moves or to the current occupancy pattern of a group of families. Among these are the following:

(1) Because of the general absence of data on such measures as a family's Net Worth, one can at best develop only rough estimates of household wealth.

(2) With the rapid escalation of housing prices in recent years, real estate equity has become a large part of a typical family's wealth. It is however no longer clear whether the household chooses owner-occupancy (for example) because of its wealth or if the household is well-off in fact because it happended to choose owner occupancy when housing prices were increasing.

(3) Others argue that wealth is not important in the tenure decision in any case. Researchers such as Smith (1971) argue that the size of downpayment and the monthly carrying costs are the major concern of a purchaser of housing. These may be tied more closely to the income (or recent history of income) of the family than to its wealth.

(4) There has been a considerable discussion over the formulation of an income measure. Particularly for young and elderly households, current annual income may be an inadequate guide to the household's ability over the subsequent years to pay for a particular housing choice. Instead, a 'permanent' income measure which although unobservable is intended to approximate the household's ongoing ability to purchase.

(5) The effects of income and wealth measures will depend on the housing choice being analyzed. Consider the tenure choice as an example. Given that there is a wide range of price possibilities for both purchases and rentals, the decision to own or to rent may be quite insensitive to income above some threshold level. On the other hand, if one was examining the dollar value of the housing purchase, one might expect this to be quite sensitive to the income or wealth of the family.

Because of these problems, different researchers have resorted to different approaches in measuring the impact of income and wealth on housing choice and occupancy patterns. In addition to the approaches mentioned above, there has been a tendency to let age of the head be a proxy for income and wealth as well and this approach is adopted for the empirical work in this paper.

As indicated earlier, little work has been done on the occupancy patterns of other kinds of households. In addition, the kinds of other households considered have varied widely. Struyk and Marshall (1974, 1975) distinguish just three types of households; those headed by primary (i.e., nonfamily) individuals. The latter two were subdivided by race (black, white) but not by any other characteristics. Maisel used a similar approach based on the following characteristics of the head; married with spouse and any number of other persons present, head with no spouse but at least one other person present, and one person households. All of these categories were subdivided by two age classes (under 65, over 65) and seven income classes. Shafer (1978) divided those households with not-married heads into three marital statuses; single, widowed, and divorce-separated. These in turn were subdivided into as many as three age classes (under 30, 30-59, 60 and over).

From this, it can be seen that in empirical work there has tended to be a considerable mixing of household types. The one-family household, the one-family household with additional persons present, and the multi-family household tend to be lumped together if both spouses are present. Sometimes, one person households are singled out, sometimes they are grouped with single parents.

In the empirical work to be described in this paper, four different household types are allowed for; the primary nuclear family (PNF) household (i.e., where the head of household is also a head of family), the PNF household with additional persons, the one-person household, and the multiple person non-PNF household. Because of the paucity of literature on the occupancy patterns of the latter three types, no attempt is made to evaluate a priori why or how these types should be different from the PNF household. However, the empirical work does afford some interpretations as will be seen shortly.

In addition, for the head of household, 6 age (Under 25, 25-34, 35-44, 45-54, 55-64, and 65 and over), 2 sex, and 2 marital status (Married, Not Married) classes have been specified. 6 household size classes (1,2,3,4,5, or 6 or more persons) have been allowed for also although only certain size classes are permitted for each household type.[3] Allowing for types of household in the above, there are 360 different possible attribute classes of which one, for example, is a PNF household of size 2 headed by a married male under the age of 25.

1.3. Of Attributes, Groups, and Sample Size

Suppose that one has drawn a sample of households, selected and measured a certain occupancy pattern, and decided upon a number of household attributes to be considered. If one were to then proceed with some kind of grouping analysis and to find as a consequence that there were only 10 different occupancy groups, how would one interpret this? Are there in reality really only, or as many as, 10 groups?

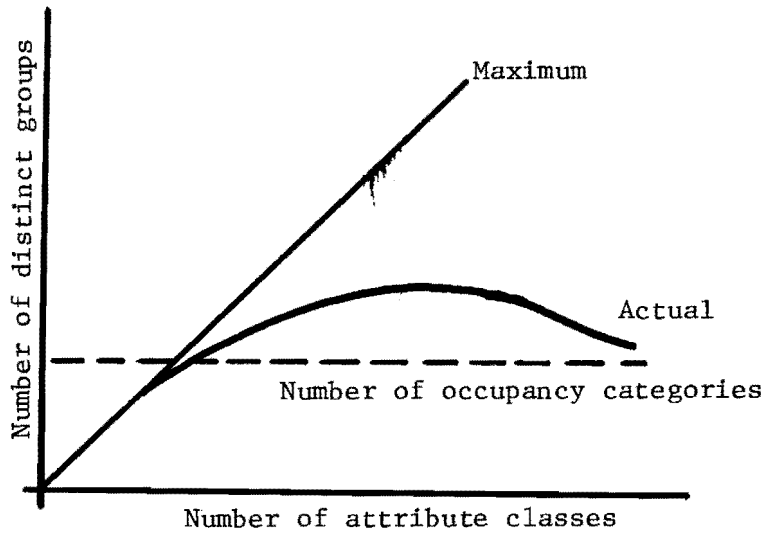
Presumably, there are a number of qualifications one would want to add in reference to a specific quantity such as 10. One might want to make some reference to the specific grouping method used as other methods could lead to different groups or different numbers of groups. Secondly, one would want to speak of these 10 groups in reference to the occupancy pattern being considered. Maybe there are 10 groups for tenure patterns but the number and composition of the groups could be different if one was considering a type-of-dwelling pattern. Thirdly, the delineation of these groups may be quite sensitive to the kinds and definitions of households attributes being considered.

---

[3] A PNF household must have at least two persons, A PNF household with additional persons at least three, a one-person household only one, and a multiple person non-PNF household at least two.

A fourth problem arises from the number of attribute classes being specified. Let us suppose that the only attribute being measured is the age of the head of household and that tenure patterns (owner-occupiers, renters) are being studied. If only two age classes were being considered (e.g., 'under 65' and '65 or older'), at most two groups will be possible. If the under 65's and 65 or older's had a similar proportional split between owner-occupiers and renters, the grouping method could result in only one group; the entire sample. Suppose however that three age classes had been used (e.g., under 30, 30-59, and 60 or over). A grouping method here might generate one, two, or three groups which are distinct in terms of tenure pattern. Thus, the maximum number of groups is limited by the number of attribute classes being considered although the actual number of groups derived may be substantially less.

One conjecture which suggests itself is that the actual number of distinct groups should in an idealized case form an inverted U-shaped relationship with respect to the number of attribute classes considered.



The rationale for this argument is that as the number of attribute classes becomes larger, it should be increasingly possible to identify groups which are homogeneous in their occupancy patterns.[4] In principle at least, if our understanding of occupancy patterns is good enough to enable an appropriate choice of attribute classes then we should be able to identify exactly the same number of distinct groups as occupancy categories being considered with each group assigned uniquely to a category. In reality, one might suspect that occupancy patterns will always be subject to some inexplicable variation and that even with a large number of attribute classes, the number of distinct groups will tend to exceed the number of occupancy categories. Thus, when the number of attribute classes is small, one might expect to find that an additional attribute class yields an additional distinct group. As the number of attribute classes becomes larger however, an additional attribute class may well lead to a consolidation and reduction in the number of distinct groups. These trends are described in an idealized form in the figure above.

A fifth problem arises from the size of the sample being considered. If the number of attribute classes and occupancy categories are large, a grouping method may fail to distinguish between two groups because of insufficient sample size. Thus, other things being equal, a large sample size may tend to increase the number of groups identified.

All of these problems serve to suggest that care be taken in the interpretation of any groupings found. In the empirical section, the same grouping method is applied to two different Toronto-area samples; one in 1971, the

---

[4] It is assumed here that the attribute classes chosen are relevant to the occupancy pattern being examined.

other in 1974. Although the groupings derived are broadly similar, there are some differences and these in part arise from these problems.

2. Issues in Grouping

There exists a substantial research literature and alternative grouping methods. Most often, this type of work is referred to as Cluster Analysis. There exist several good references and the reader is recommended to Anderberg (1973). The discussion in this section will center on just a few key issues in this kind of research. First, some attributes of this particular grouping problem will be discussed which restrict the kinds of clustering methods possible. Secondly, two approaches (aggregation and disaggregation) to cluster analysis will be assessed. Thirdly, a new method of clustering which we call 'split-recombination' (SR) analysis will be introduced.

2.1. Atributes of the Grouping Problem

There are two attributes to this grouping problem which restrict the kinds of cluster analysis methods which can be applied.

One of these attributes has to do with the existence of pre-defined groups. In cluster analysis, there are a number of techniques which attempt to allocate each of a set of sample observations to one of several pre-defined populations or groups. Here however, it is intended that the grouping method define its own groups by a partitioning of the sample. No direct attempt is made to pre-define groups.

The second attribute of interest is that the criterion variable (occupancy patterns in this case) is generally nominally-scaled. There are a number of cluster analysis methods which assume interval scaling and are therefore inappropriate. Further, some cluster analysis methods can employ a binary criterion variable. These latter methods may be appropriate when looking at tenure patterns (owner-occupiers versus renters) for example.

However, these methods are not relevant when occupancy can take on more than two forms as is the case in the empirical work in this paper.

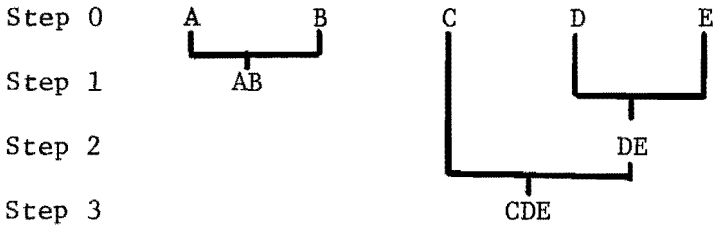
With a nominally-scaled criterion variable, one must generally rely on a contingency table measure to assess similarity or dissimilarity. In the empirical work for this paper, reliance was placed on a familar chi-square statistic. The advantage of this statistic is that confidence intervals can be calculated. The disadvantages of this particular statistic and the alternatives available are discussed in Anderberg (1973; pp. 75-83).

2.2 Two Approaches to Grouping

In most grouping problems, there are two alternative approaches to cluster analysis. One begins by disaggregating the sample into attribute classes. A class or group is then aggregated with either another class or group of classes into a new group based on some similarity measure. This aggregation can continue either until there is only one group left or until the similarity measure fails a certain critical value. The second approach (disaggregation) works in the opposite direction. In it, the entire sample is initially put into one group. This group is then split into two groups on the basis of some dis-similarity measure. These two groups and their successors are subsequently split each into two groups. This process continues until either the original attribute classes are obtained or until some critical dis-similarity value is failed.

A property common to both is the sequential nature of the algorithm. Consider the aggregation approach as an example. Suppose a sample contains five attribute classes; A,B,C,D,and E. An aggregation algorithm generally selects the strongest similarity value and suppose that this is for A and B. These two clases are put into a group, AB, in Step 1. Then, the similarity measures are re-calculated where necessary between all pairs of AB, C, D, and

E and the highest scoring pair combined in Step 2. Suppose that this continues until Step 4 when some critical similarity value is failed and the following results are obtained.



Because of the sequential nature of the algorithm, C in the above example has been grouped with D and E. This of course may be quite appropriate. Suppose however that among the initially similar pairs A and B, A and C, C and D, and D and E, that the highest similarity value (for A and B) had been very closely followed by that for A and C but the value for AB and C had been substantially lower. In that case, the appropriateness of grouping C with DE may be called into question.

A similar problem may well arise when using the disaggregation approach. In one sense, it is a partly inescapable consequence of any grouping method based on pairwise comparisons. However, the problem can be particularly serious in the aggregation approach when the number of attribute classes is large relative to the sample size.

To highlight this last argument, consider the case of the 2680 household CMHC sample used in this paper. As indicated earlier there are 360 possible attribute classes although sampled households are found in just 187 of these. Further, a large number of these attributes classes include only a few sampled households. A number of these less-frequented attribute classes furthermore tend to have all their sampled households in only one occupancy category (e.g. high density, rental). An aggregation method will usually begin by clustering

classes which have all their observations in the same one occupancy category. If there are enough of these, a group is created "which is large and has only one occupancy category occupied. This group then is likely to be significantly different from other attribute classes or groups even though its component classes individually are not.

Let us consider a small numerical example which illustrates the problem involved. Consider the following five attribute clases; A,B,C,D, and E.

Type of Occupancy	A	B	C	D	E
Low Density, Owner-Occupied	0	0	0	0	0
Low Density, Rented	0	0	0	0	1
High Density, Owner-Occuped	0	0	0	0	0
High Density, Rented	10	18	9	15	10

By most measures, all of these classes would be very similar. However, a grouping algorithm will generally combine A,B,C, and D first to create a group with 52 observations in the fourth occupancy category and none in the rest. This new group however has become different from E in terms of occupancy patterns. The one low density, rented household in E may be enough at this point to force one not to group E with the rest. In our empirical work, this has been found to occur quite frequently and is characterized by groupings which appear to be related more to the sample size of the attribute class than to intuitive ideas about the similarities among the included classes.

The same problem can of course occur in the disaggregation approach. If one had an ABCDE group and was considering splitting off E a similar result would be obtained. Although in principle the results could be the same, disaggregation tends to perform somewhat better in practice. There are two reasons for this. One is that disaggregation does not typically lead to an ABCDE group. Usually, there are other attribute classes similar

to E which are also included in the group so that E does not stand out alone. The second reason is that disaggregation as used here involves a binary split along an attribute value. Thus, for example, one might split the ABCDE group according to size with 2-person households in one group and all other sizes in another. Suppose attribute class E contains two-person households and thus is split off. Typically, however, at least one of the other four classes also contains two-person households. These too would have to be split off with E. This in turn would make it less likely that a split would occur because the difference between the two new groups would be muted. For these two reasons, disaggregation is here preferred to aggregation.

At the same time, a word of caution is in order. Suppose one employs a disaggregation approach and undertakes binary splitting until some critical dissimilarity value is failed. How should one interpret the groupings derived? A warning should be made about the nature of binary splitting in addition to all the issues raised in section 1.3. The groupings one derives are such that no additional binary splits of any group by an attribute value are possible which yield significantly different groups. The emphasis here is on the term binary. It may well be possible to further split a group by considering a combination of attribute values (e.g., two-person households with a head under 25 years of age) but there is no single attribute value along which a significant split can be generated.

2.3 The SR Method

The most popular example of a disaggregative clustering method is AID analysis.[5] In AID, a set of observations are split into a pair of groups.

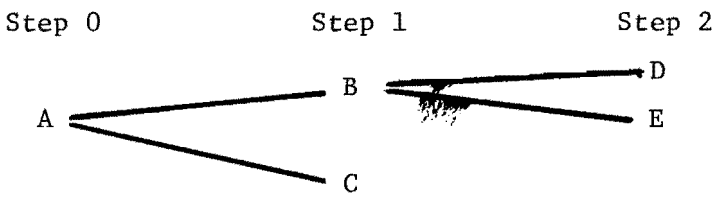
---

[5]Refer to Sonquist et al (1973).

Each group is then split into another pair. This splitting continues in much the same way as described in section 2.2. There is however one basic problem with this kind of disaggregation method. When a split occurs; one or both of the two new groups created may in fact be very similar to another already-existing group. In AID, there is no such possibility of a new group created being aggregated with similar groups.

At first glance, this might not seem to be such a problem. After all, one could use an AID approach to define an initial set of groups and then simply aggregate those groups with similar housing occupancy patterns. This is not entirely satisfactory however because it ignores the effect of intermediate aggregation on subsequent splitting. The best attribute value on which to split a group may change if we aggregate that group with other similar ones.

For this reason, we propose the Split-Recombination (SR) Method. This disaggregative method begins with one group. It splits that group into two based on the attribute value yielding the highest χ^2 value. It then calculates for each split group the maximum χ^2 value associated with a further split of that group for some other attribute value and splits whichever one yields the higher of these maximum χ^2 value. The following diagram is a typical sequence to this point.



The whole group (A) has been split into B and C and subsequently B has been split into D and E. There are thus three groups at the end of the second step (D,E, and C). A pair of steps ensues.

Step 3: The two newest groups (D and E) are each compared pairwise with all pre-existing groups (C). If any pairs have a sufficiently low χ^2, they are combined (e.g., CD or CE).

Step 4: For each group after Step 3, the best (maximum χ^2) binary split is calculated. However, only one group, that with the highest of these χ^2 values, is actually split (as in Step 1).

This pair of steps is repeated again and again until either no more recombinations or splits are feasible or until some arbitrary cutoff point is reached.

3. Empirical Findings Using the SHU File

In the Fall of 1974, Central Mortgage and Housing Corporation undertook a Survey of Housing Units (SHU) covering 23 urban areas in Canada. In the Toronto sub-sample, coverage was limited to the Municipality of Metropolitan Toronto, the Town of Mississauga, and the then Town of Port Credit (now part of Mississauga). This represents a somewhat smaller area than the Toronto CMA but contains its most populated areas. The sub-sample in each urban area was partly stratified and the interested reader is referred to CMHC (Updated)

As a result of this survey, CMHC produced a microdata file. This file contains the responses of individual households to many of the survey's questions. These cover (1) physical characteristics of the dwelling unit (ii) characteristics of the occupying household, (iii) payments for the upkeep of the dwelling, and (iv) for household heads who have moved since June 1971, the characteristics of the last dwelling occupied.

This file has however been 'screened' to preserve confidentiality by (i) removing all households with household incomes in excess of $100,000 and (ii) recoding ages of individuals, principals outstanding on mortgages, and market values of dwelling when these exceed pre-defined maximums. The way in which these data are used in this study makes (ii) unimportant.[6]

---

[6]Mortgage principles and market values are not used here and the oldest age group used is 65 and over.

However, screening effect (i) tends to make the sample somewhat biased towards lower income groups.[7] This latter effect is in addition to any bias introduced by the sampling stratification itself.

The Toronto area subsample contains 2704 households after screening. For each of these households, the type of dwelling occupied (low density versus high density as defined earlier) can be ascertained. The tenure categories allowed for in SHU are

(1) Owned or being bought as a condominium by a member of this household.

(2) Owned or being bought by a member of this household.

(3) Rented for money by a member of this household.

(4) Other.

Categories (1) and (2) were combined under the 'owner-occupier' label. Category (3) is of course the 'renter' group. Households in category (4) which include those living in dwelling units owned by others but not paying rent are excluded in the analysis which follows. Without category (4), the Toronto area subsample is reduced to 2680 households.

The SR method was applied directly to this reduced sample of 2680 households. No attempt was made to treat the bias problems introduced by stratification and screening. It proved to be important to keep track of the small cell counts occurring in a number of the finely defined attribute classes. This proved to be administratively difficult when working with anything but the raw sample. At the same time, the potential effect of sample bias on grouping is recognized.

Let us begin by describing some of the initial steps taken in the SR

---

[7] In all, 5 households of the 2709 sampled in the Toronto area have been removed from the microdata file for this reason.

method. Some interesting interpretations are possible of the groups initially formed. Subsequently, we shall look at the groupings which result when the SR method is allowed to continue until a critical χ^2 value is failed.[8]

3.1. Initial Groupings Using the SR Method.

Consider the possibilities for an initial split of the 2680 household sample. There are 17 distinct splits possible with the specified attribute classes. These are.

(1) age=15-24 vs. other
(2) age=25-34 vs. other
(3) age=35-44 vs. other
(4) age=45-54 vs. other
(5) age=55-64 vs. other
(6) age=65 or older vs. other
(7) married vs. not married
(8) males vs. female
(9) size = 1 vs. other[9]
(10) size = 2 vs. other
(11) size = 3 vs. other
(12) size = 4 vs. other
(13) size=5 vs. other
(14) size=6 vs. other
(15) type=PNF vs. other[10]
(16) type=PNFP vs other[10]
(17) type=NPNF vs other[10]

The split which yields the highest χ^2 value (χ^2 = 394.) is (7); households headed by married persons (group A.1) versus those with not married heads (group A.2). The distribution of households by occupancy type before and after the split is shown below.

	All Households	Head's Marital Status: Married (A.1)	Other (A.2)
Low Density			
Own	1255	1076	179
Rent	163	112	51
High Density			
Own	126	97	29
Rent	1136	547	589
Total	2680	1832	848
		χ^2=394.	df=3

[8] A 95% confidence level is assumed. The critical χ^2 value depends on the degrees of freedom which in turn is the product of the number of nonempty rows (in the contingency table) minus one and the number of columns (always 2 in this case) minus one.

[9] Size = 1 also of course implies the one-person household type.

[10] PNF=Primary Nuclear Family, PNFP=PNF plus Additional Persons; NPNF=Multiple Person Non-PNF household.

The effect of this split is quite substantial. The group containing married heads is much more likely to be residing in owner-occupied dwellings (both low and high density). Conversely, the not married head group is much more likely to occupy high density rental units. Both groups have about the same tendency (6% of all households in each case) to occupy low density rental accommodation.

Thus splitting households by the marital status of heads yields the most distinctive patterns of housing occupancy. The alternatives to an initial split by marital status of head appear to be relatively inferior. The next best splits of the original sample and their associated χ^2 values are as follows; by sex of head (8) with a χ^2 value of 237., by youngest age of head group (1) with a value of 160., and by PNF (15) with a value of 153. It is noted in addition that several of these are reasonably well correlated with the marital status split. Most of the female heads and most of the non-PNF households for example are headed by persons who are not married.[11]

The next step in the SR method is to split whichever of either the married head, or not married head, group that can be split with the highest χ^2 value. In either case there are now 16 possible splits; one less than before because a marital status split is no longer possible. For group A.1, the best split is according to the oldest age group (6) with a χ^2 of 51.1. For the married head group (A.2) however, the best split is according to households of size two (10) with a χ^2 of 155. The latter split is thus

---

[11] In SHU, the 'head' of household is the person making the greatest financial contribution to the household's operation. There are very few married women in the sample who qualify as heads under this definition.

carried out and the following results are obtained.[12]

	All Households	Not Married Heads (A.1)	Married Heads Size = 2 (B.1)	Married Heads Size ≠ 2 (B.2)
Low Density	Hhlds (%)	Hhlds (%)	Hhlds (%)	Hhlds (%)
Own	1255 (47)	179 (21)	275 (43)	801 (67)
Rent	163 (6)	51 (6)	23 (4)	89 (7)
High Density				
Own	126 (5)	29 (3)	35 (5)	62 (5)
Rent	1136 (42)	589 (70)	304 (48)	243 (20)
Total	2680 (100)	848 (100)	637 (100)	1195 (100)
			$\chi^2 = 155.$	df = 3

The figures in the parentheses are the percentages of households in a group in each category.

The effect of this split on occupancy patterns is quite visible. Two person households with married heads (Group B.1) are less likely to reside in low density units (either as owner-occupiers or renters) than are other-sized households with married heads (b.2). At the same time both groups are equally likely to be high density owners so the difference is made up in the high density renter category.

The third step in the SR method is to see whether either of the two new groups (B.1 or B.2) should be combined with the pre-existing group (A.1). The appropriate χ^2 values for recombination are 94.3 and 517. respectively. Both of these values are so high that one would reject at a 95% confidence level the hypotheses that the occupancy patterns are independent of groupings (B.1 vs. A.1 and B.2 vs. A.1). Thus at the end of the third step, one is here left with three distinct groups.

---

[12]Note that the χ^2 value is computed only for the contingency table formed by the two right columns of this table. In other words, the choice of split for the married heads group for example is not affected by the existence of the not married group.

From this point on, the SR method involves successive repetitions of the second and third steps. In Steps 4 and 6, households of size 3 and 1 respectively were split off from the B.2 group. In Step 5, no recombination was found to be acceptable. However, in Step 7, the one-person households with a married head were recombined with the non-married households (group A.1).[13] The groupings at that point were as follows:

Group Label	Group Description	Low Density		High Density		Total
		Own	Rent	Own	Rent	
		(Percent of Households in Group)				
E.1	Not Married; Married, Size = 1	21	6	3	70	100
B.1	Married; Size = 2	43	4	5	48	100
C.1	Married; Size = 3	53	7	8	33	100
D.1	Married; Size $\geq$ 4	75	8	4	13	100
ALL	All Households	47	6	5	42	100

It is noteworthy that to this point, only two kinds of attributes have been used to define groups; marital status of head and number of persons in the household. Age and sex of head and household type (other than one person households) have not been introduced. At the same time, the comment made earlier is repeated here that marital status, household type, and sex of head may be quite inter-related. However, the results to this point could be used to argue that the lack of attention given non-family households in previous studies is warranted in that family households tend to dominate three of the four groups so far defined.

---

[13] The Chi-square value for recombination in this case was 1.7, well below the critical value.

In concluding this section, it is instructive to ask how much each additional split and recombination have contributed to our understanding of occupancy patterns. In other words, how much information is present in the last table in addition to that in the preceding table or the one preceding that. One measure is of course the Goodman-Kruskol asymmetric λ (with occupancy category as the dependent variable). In the last table $\lambda = .315$, in the preceding one $\lambda = .308$, and in the one preceding that $\lambda = .287$. Thus, the (A.1, A.2) groupings, and the (A.1, B.1, B.2) groupings to a lesser extent, contribute much to the prediction of a household's occupancy category while the additional refinements offered by B.1, C.1, D.1, and E.1 contribute very little over and above this.

3.2 Final Groupings Using the SR Method

Before arriving at a point where no additional splits created significant new groups or where recombinations were feasible, there were a total of 33 splits and 20 recombinations. In the end, this left 14 distinct groups. These groups are described below in Table 1. Attribute classes or contiguous clusters of classes which contain 20 or more sample households are identified as 'major' included classes for each group.

Table 1: Household Groups By Number of Sample Households Included (Size), By Major Included Attribute Classes, and By Occupancy Category Distribution; Toronto Study Area, 1974 Survey of Housing Units.

Group	Size	Major Included Attribute Classes	Low Density Own	Low Density Rent	High Density Own	High Density Rent	Total
			(Percent of Total Households)				
A	84	Female, 15-34, Size=1.	0	0	0	100	100
B	239	Male, 15-54, Size=1. Not Married, 25-44, Size=2. Male, Married, 15-24, Size=3.	7	5	6	82	100
C	66	Not Married, 15-24, Size=2-3.	0	18	0	82	100
D	61	Married, 15-24, Size=2.	12	16	2	71	100
E	437	Female, 35+, Size=1. Males, 55+, Size=1. Married, 25-34, Size=2.	28	2	4	67	100
F	78	Married, 35+, Size=2. Female, Not Married, 45+, Size=2.	46	0	1	53	100
G	232	Married, 25-34, Size=3.	43	10	5	42	100
H	457	Male, Married, 35+, Size=2. Male, Not Married, 45+, Size=2.	54	2	6	38	100
I	47	None.	30	36	9	26	100
J	209	Married, 35+, Family, Size=3.	64	5	8	24	100
K	236	Married, 25-34, Size=4+.	66	10	3	21	100
L	86	Male, Married, 35+, PNF, Size=6+.	70	21	2	7	100
M	471	Married, 35+, Size=4-5. Married, 35+, PNFP, Size=6+.	81	4	5	10	100
N	1	None.	0	0	100	0	100
ALL	2680	All Households.	47	6	5	42	100

NOTES: (1) Specified attributes are listed in the following order; sex, marital status, and age of head, household type, and household size.

(2) Where an attribute is not specified, that group includes all values for that attribute. For example, group K includes both male and female heads.

(3) 'Family' includes both Primary Nuclear Family (PNF) and PNF plus Additional Persons (PNFP) households.

(4) An NPNF household is a multiple person non-PNF household.

One of those groups, group N, is something of an anomaly. It contains just one household from the sample. Three things happened which led to the creation of this group. One is that this household just happened to be a high density owner; one of the two infrequently occupied categories. If it had been a low density owner or a high density renter for example, it would likely have been recombined with group M or A respectively. Another is that the group from which this household was split off just happened to have that one household with a particular attribute characteristic. In this case, this household was the only one in a previous group with a married head. The third thing which happened was that the χ^2 value for splitting although significant was getting close to a critical value. This kind of anomalous splitting, if it occurs, can be expected in the later steps when the chi-square values are becoming fairly small.

The remaining thirteen groups however do have an intuitive interpretability.[14] These are perhaps best seen by looking at the broad associations between groups and household size. Each group contains only a small range of household sizes. Also, as might be expected, there tends to be systematic relationship between size and the propensity to live in low density accommodation. Finally households of the same size are found in different groups depending on other circumstances. Let us consider individual household sizes.

> One-Person Households. Younger females are found in group A (highly likely to be high density renters), younger males in group B (less but still very likely to be high density renters), and older males and females in group E (with approximately 30% in low density accommodation). These differentials may be attributable to income and wealth differentials between age groups and sexes. For older persons, it is also partly explicable in terms of the circumstances (e.g., widowhood) under which the person came to reside in a one-person household.

---

[14] Group I which has no major attribute classes present is a combination of a lot of small strata basically consisting of households (often NPNF) with 3-4 persons headed by a non-family person.

Two-Person Households. Two-person households with married heads aged 15-24 are found in group D and those with married heads aged 25-34 in group E. Group D does not have a high level of owner occupancy (only 14%) but does have a moderate level of low density occupancy (28%). Group E by contrast has about the same level of low density occupancy (30%) but a much higher level of ownership (32%). Thus the difference between the two age groups reflects a change in tenure more than in type of dwelling and many be attributable to income and wealth differentials. Two person households with not-married heads aged 15-24 are found in group C while for those with not-married heads aged 25-44 are found in group B. Again, one finds approximately the same level of low density occupancy (18% in C versus 12% in B) but a higher incidence of ownership for the older age group. The levels of ownership here are substantially below those for married persons. For older age groups, the split is not by marital status but by sex of head alone. Older male-headed households are found in group H while older female-headed ones are in F. The difference between the two is a lower level of low density occupancy and ownership in F visa-vis H. This may well be attributable to income and wealth differentials between the sexes as well as the possibly smaller frequency of a second wage earner in female-headed households.

Three-Person Households. Households of size 3 with a married head tend to fall into one of three groups depending on the age of the head; group B for 15-24 year olds, group G for the 25-34's, and group J for the over 34's. This pattern shows a steady shift with age towards both low density occupancy and ownership. This is somewhat different than was found above for two-person households where there was a plateau in low density occupancy in the 15-34 age bracket. Further, it is noted that among the over-34's, size 3 households with married heads are much more likely to be owners (of both low and high density units) or low density occupants (either owners or renters) than are similar households of size 2. Three-person households with not-married heads appear mainly in the 15-24 age group which is included in group C with corresponding two-person households. Other age groups are not well represented in this household size in the SHU sample.

Four-Person Households or Larger. There are few households in the sample of size 4 or larger which have either a not-married head or a head who is under 25 years of age. Households with married heads aged 25-34 years are grouped into K if they have 4 or more persons. This group has a higher level of low density occupancy than do similar two or three person households (groups G and E) and a lower rate of high density occupancy (both rented and owned). Households with married heads over 34, and having 4 to 5 persons are grouped into M which has the highest rate (81%) of low density ownership. For larger households with 6 or more persons however, the relevant group depends on whether the household is a PNF or a PNFP. If there is a Primary Nuclear Family only in the household, the grouping changes to L which has a higher level of low density

occupancy but a lower level of ownership.[15] If however there is a PNF with additional persons present, the household remains in group M. The question of causality alluded to earlier becomes clear here. Do PNFP households have a high rate of low density ownership because they have these additional persons or do they have additional persons present because they are owner-occupiers?

On the basis of the above, several patterns can be identified in these groupings. First, there appear to be few differences in occupancy patterns among age groups once over the age of 34 if size and marital status are controlled for. Secondly, the size of household is very important in shaping occupancy patterns when age and marital status of the head are controlled for. In general, household size has a positive effect on low density occupancy and an inverted-U shaped effect on ownership levels. Finally, marital status also has an important effect with married-head households being much more likely to be low density occupiers and owners. However, the relatively infrequent occurrence of larger households with not-married heads in the sample restricts our conclusion here basically to one and two person households.

4. Temporal Stability of Household Groups

One way of evaluating the groups derived in Section 3 is to ask how stable they are. Would samples of households from other urban areas show the same kinds of groupings? Would a sample of households in the same area at another point in time yield similar groupings? In this section, only the latter question is explored.

The 1971 Census of Canada Public Use Sample for the Toronto CMA provides

---

[15] This corresponds to an argument raised by David (1962) which asserts that larger families typically have more competing demands on the household's income and are therefore less able to purchase housing. Instead, such families resort to renting the low-density housing they need. That larger families may also tend to be poorer to begin with makes this pattern even more pronounced.

a data set that is in many respects comparable with the 1974 SHU file.[16] It will be used here in beginning to answer the above question. It might be argued that a 1971 sample should generate a set of groupings similar to that in 1974 since many households will not have relocated during this period. While admitting this, it is noteworthy that the SHU survey found that fully 40% of all household heads had changed their place of residence between the Census and SHU survey dates. In addition, these two data sets are the only comparable ones available for the Toronto area.

Although the SHU and PUS Files share many similarities, there are some notable differences. In the 1971 Census, a male spouse if present is generally taken to be the head of household while in SHU the head is the principal financial contributor to the household's upkeep. In the SHU file however, there are few cases of a married female head when a spouse is present so this difference may not be important in practice. In terms of grouping algorithms however, an important difference between the two samples in that the SHU sample is much larger with 7735 households.[17] From the discussion in Section 1.3, one might therefore expect more groups to be defined with the PUS sample than was the case with the SHU sample. Also, between 1971 and 1974 in Canada, there has been a substantial increase in condominium ownership. This relatively-new form of ownership may have caused occupancy patterns to change.

---

[16] The Public Use Sample household file is used here. This is a simple 1% random sample of all households in the Toronto CMA.

[17] The actual PUS sample is slightly larger than this but includes households which are normally resident in the CMA but are residing abroad temporarily. These have been excluded here.

In applying the SR method to this 1971 sample, a total of 49 splits and 31 recombinations were carried out. This left 19 distinct groups as described in Table 2. Again, the 'major' included attribute classes (those containing 20 or more households) are identified. There are two relatively small groups, O and S, which do not have any major included attribute classes.

At first glance, the groupings in Tables 1 and 2 appear to be bewilderingly different. With the 1971 sample for example, there is an emphasis on household type (PNF, PNFP, and NANF) which is not found in the 1974 groupings. However there are also some broad similarities between the two groupings. This can be seen by comparing specific attribut class groupings using Table 3. In this table, the most frequented attribute classes in each sample are identified together with their group affiliations. All excluded attribute classes individually contain less than 20 sampled households. The following detailed comparisons can be made.

Table 2: Household Groups by Number of Sample Households Included (Size), by Major Included Attribute Classes, and by Occupancy Category Distribution; Toronto CMA, 1971 Public Use Sample.

Group	Size	Major Included Attribute Classes	Low Density Own	Low Density Rent	High Density Own	High Density Rent	Total
A	63	Female, Not Married, 35-44, Size = 1.	5	2	0	94	100
B	185	Female, Not Married, 15-34, NPNF, Size = 1-3.	0	10	0	90	100
C	617	Male, Married, 15-24, PNF, Size=2-3. Male, Not Married, 15-44, Size = 1. Male, Not Married, 15-34, Size=2-3.	7	16	1	76	100
D	766	Male, Married, 25-34, PNF, Size=2. Male, Not Married, 35-44, NPNF, Size = 2. Not Married, 45-64, Size = 1. Female, Not Married, 35-44, 2-3, PNF.	24	9	1	66	100
E	298	Female, Not Married, 65+, Size = 1.	35	2	3	61	100
F	144	Male, Married, 15-24, PNF, Size = 4+. Male, Married, 15-24, PNFP, Size=3+.	19	32	0	49	100
G	345	Male, Married, 25-34, Family, Size=3.	35	14	2	48	100
H	154	Male, Not Married, 65+, Size = 1. Female, Not Married, 45-54, Size=2.	46	7	0	47	100
I	606	Male, Married, 35-44, PNF, Size=2. Male, Married, 25-34, Family, Size=4. Male, Not Married, 45-64, NPNF, Size=2.	48	15	1	35	100
J	1062	Male, Married, 45+, PNF, Size=2. Male, Not Married, 65+, NPNF, Size=2. Female, Not Married, 65+, Size=2.	60	3	3	33	100
K	301	Male, Married, 35-44, Family, Size=3. Female, Not Married, 55-64, Size=2. Female, Not Married, 65+, Size=3.	61	10	3	26	100
L	194	Male, Married, 25-34, Family, Size=5.	59	19	2	21	100

Table 2 continued...

M	895	Male, Married, 45+, Family, Size=3. Male, Married, 35-44, Family, Size=5.	76	7	3	15	100
N	426	Male, Married, 35-44, PNF, Size=4.	72	12	2	14	100
O	28	None.	29	57	0	14	100
P	803	Male, Married, 45+, Family, Size=4-5.	84	6	2	8	100
Q	545	Male, Married, 25+, PNF, Size = 6+. Male, Married, 25-34, PNFP, Size=6+.	73	21	0	6	100
R	305	Male, Married, 35+, PNFP, Size=6+.	85	11	2	2	100
S	6	None.	50	33	17	0	100
ALL	7735	All Households.	53	11	2	35	100

NOTE: See Note to Table 1.

Table 3: Comparison of 1971 and 1974 Groupings for Selected Strata.

	1974 SHU Sample							1971 PUS					
	Size of Household							Size of Household					
Age of Head	1	2	3	4	5	6+		1	2	3	4	5	6+
15-24	-	D	B	-	-	-		-	C	C	F	-	-
25-34	-	E	G	K	K	-		-	D	G	I	L	Q
35-44	-	H	J	M	M	L	Males Married	-	I	K	N	M	Q
45-54	-	H	J	M	M	L	PNF	-	J	M	P	P	Q
55-64	-	H	J	M	-	-		-	J	M	P	P	Q
65+	-	H	J	M	-	-		-	J	M	P	P	-
15-24	-	-	-	-	-	-		-	-	F	-	-	-
25-34	-	-	G	K	K	K		-	-	G	I	L	Q
35-44	-	-	J	M	M	M	Males Married	-	-	K	N	M	R
45-54	-	-	J	M	M	M	PNFP	-	-	M	P	P	R
55-64	-	-	J	-	-	-		-	-	M	P	P	R
65+	-	-	J	-	-	-		-	-	M	P	-	-
15-24	B	C	-	-	-	-		C	C	C	-	-	-
25-34	B	B	-	-	-	-		C	C	C	-	-	-
35-44	B	B	-	-	-	-	Males Not Married	C	D	-	-	-	-
45-54	B	H	-	-	-	-	NPNF	D	I	-	-	-	-
55-64	E	H	-	-	-	-		D	I	-	-	-	-
65+	E	H	-	-	-	-		H	J	K	-	-	-
15-24	A	C	-	-	-	-		B	B	B	-	-	-
25-34	A	B	-	-	-	-		B	B	B	-	-	-
35-44	E	B	-	-	-	-	Females Not Married NPNF	A	D	-	-	-	-
45-54	E	F	-	-	-	-		D	H	-	-	-	-
55-64	E	F	-	-	-	-		D	K	-	-	-	-
65+	E	F	-	-	-	-		E	J	K	-	-	-

Note: '-' indicates few or no sampled households or else a household size which is inconsistent with household type.

(1) Households with Six or More Persons. In the SHU and PUS samples, PNF households with six or more persons are treated as groups separate from other household sizes. Further, the level of low density occupancy is high although this is achieved in both cases by considerable low density rentals. Among PNFP households, there is a difference between the 1974 and 1971 groups. In the SHU group, 6+ person households are lumped with 4 and 5 person households while in the PUS group they are treated as distinct entities. As argued earlier however, one expects with larger samples such as PUS to be able to better identify distinct groups.

(2) Households with Four or Five Persons. In the SHU groupings, four and five person households are split mainly into two groups; K for those with a head under 35 and M for the rest. In the 1971 groupings, there are two differences. The age of head split now occurs at age 45 and below this age there are distinct groups for four and five person households. This again may have arisen because of the larger size of the 1971 sample.

(3) Three Person Households. In both the SHU and PUS groupings, family households show an increasing tendency toward ownership and low density occupancy with increasing age of head up to some plateau. As with 4-5 person households, this plateau is reached around age 35 in the SHU sample and age 45 in PUS.

(4) Two-Person Households. Among PNF households, the difference between the 1971 and 1974 groupings is primarily in the plateau age of head. Again, it appears as 35 in 1974 and 45 in 1971. For NPNF households, the two sets of groupings appear to be more dis-similar. In the SHU groupings there are two plateaus. Younger heads cluster into the similar groups B and C while heads over 44 are grouped into F (for females) or H (for males). In the 1971 groupings however, the 15-34 year old heads are grouped together while the remaining age groups show an increasing tendency toward low-density occupancy and household headship with age.

(5) One-Person Households. As with two-person NPNF households, the 1974 groupings tend to provide for fewer groups. For males, there are two groups, those under 55 and the rest. The same is true for females with the split however at age 35. In the 1971 groupings for males, 15-44, 45-64, and 65+ age groups are identified. For females in 1971, the age groupings used are 15-34, 35-44, 45-64, and 65+.

Thus, one can summarize the differences between the two sample groupings as follows. First, in the 1974 sample, there is a tendency to put households with heads over 34 in the same groups. In the 1971 sample, the tendency is more often to break at age 45. Whether this reflects changes in occupancy patterns between 1971 and 1974 or a temporal instability in the groupings is

unknown. One could argue that the rapid increase in the supply of low-cost condominium apartments and townhouses between 1971 and 1974 may well have contributed to a lowering of the ages at which heads became owners and low-density occupants. Secondly, there is a tendency for the 1971 groupings to disaggregate groups identified in the 1974 sample. As argued earlier, this may partly be because the 1971 sample is large enough to identify statistical differences that would be treated as chance discrepancies in the 1974 sample.

5. Conclusion

This paper has surveyed some of the conceptual problems in identifying groupings of households with similar occupancy patterns. It also describes a new approach, the SR Method, designed to derive a set of distinct groupings which are in one sense internally homogeneous. Applying this method to a sample of 2680 households in the Toronto area in 1974, 14 distinct household groups were identified in terms of type of dwelling and tenure patterns. The same method applied to a 1971 Toronto CMA sample of 7735 households yielded 19 groups. These two sets of groupings are broadly similar although some differences between them have been highlighted.

The purpose of this paper has been to explore the feasibility of deriving homogeneous groups of households. It is concluded that while tools of analysis such as the SR Method do provide some clues, there are a number of limitations to grouping, both conceptual and empirical which need to be borne in mind. These relate to the occupancy patterns being measured, the household attributes being considered, the sample size, the grouping method and criterion being used, and the dissimilarities possible in groupings performed with different samples of households.

6. References

Anderberg, M.R. (1973). Cluster Analysis for Applications. Academic Press.

Beresford, J.C. and A.M. Rivlin (1966). Privacy, Poverty, and Old Age. Demography, 3. 247-258.

Central Mortgage and Housing Corporation (undated). 1974 Survey of Housing Units. Background Information and Statistical Notes.

David, M.H. (1962). Family Composition and Consumption. North Holland.

Doling, J. (1976). The Family Life Cycle and Housing Choice. Urban Studies, 13. 55-58.

Kain, J. and J. Quigley (1975). Housing Markets and Racial Discrimination. NBER.

Li, M. (1977). A Logit Model of Homeownership. Econometrica, 45. 1081-1097.

Maisel, S.J. (1966). Rates of Ownership, Mobility, and Purchase. Essays in Urban Land Economics. University of California Press. 76-108.

Shafer, R. (1978). Metropolitan Form and Demographic Change. Urban Studies, 15. 23-33.

Smith, L. (1971). Housing in Canada: Market Performance and Policy Performance. Research Monography No. 2. CMHC.

Sonquist, J.A., et al. (1973). Searching for Structure, Revised Edition. University of Michigan.

Struyk, R.J. and S.A. Marshall (1974). The Determinants of Household Home Ownership. Urban Studies, 11. 289-299.

Struyk, R.J. and S.A. Marshall (1975). Income and Urban Home Ownership. Review of Economics and Statistics, 57. 19-26.