Geodemographic Segmentation: Do Birds of a Feather Flock Together?

Charlie Nelson
director foreseechange
April 2003

with Nadine Wake
QED Research

Introduction

Geodemographic segmentation is based on identifying demographically similar postcodes, or the smaller census collector districts. The theory is that “birds of a feather flock together”. Vendors of such systems typically divide the country into 30 to 60 segments of areas, which can then be used in target marketing. One vendor’s brochure describes geodemographic segmentation as enabling the pinpointing of who your target markets should be, where they are located and how to reach them. Such segmentation uses a statistical technique called cluster analysis and the segments are often called clusters. But that the clusters are not necessarily geographically close. A cluster can comprise quite widely dispersed areas that have a similar demographic profile in common.


Cluster-based marketing has gone mainstream and is now used by corporate, nonprofit, and political groups alike to target their audiences. Ad agencies search the clusters for hot buttons on the consumer psyche. Restaurants, banks, and retailers use cluster systems to identify locations in which to open new outlets or close poor performers. Direct mailers discover the best prospects from subscription lists by correlating addresses to cluster tastes and spending habits. Even health officials are using clusters to correct the bad habits of citizens, earmarking funds for prevention programs based on a community’s cluster profile.

Geodemographic segmentation works by grouping together small areas with similar demographic profiles. It does not group together similar people or households – because detailed data on each household is not available. So the validity of the segments depends on households within a small area having a similar demographic profile. Is this a reasonable assumption?

The Fallacy of Averages

A couple of years ago I was at a conference and a company was demonstrating to the audience their new software that classified people on the basis of their address. It was based on a cluster analysis of Australian Bureau of Statistics Household Census data and software that translated an address into a collectors district (typically 300 homes).
Once the collectors district was known, then the person at that address could be classified into one of 30-odd segments that were based on the cluster analysis of the census demographic data.

I asked for my address to be input and we were told that I was probably a “young single migrant”. At the time, I was in my late 40’s. I did migrate from England to Australia (at age nine) and I had been in a de facto relationship for about 15 years. I was not what a marketing manager would think of if his or her target was young single migrants. In my area (Brunswick, in inner Melbourne) there are very few people alike. There are a lot of young singles, many of whom have moved closer to town from their parents home in the outer suburbs. There are a lot of migrants, many of them arriving from Greece in the 1960’s. There are also a lot of retired Australians who worked at the now vanished factories in the area in the 1960’s and 1970’s.

In short, I live in a diverse area and classifying it as “young single migrants” is an example of the fallacy of averages. It is a very misleading description of the population.

Another example of this in Harry Dent’s book “The Roaring 2000’s”. Describing a new boomtown, Telluride in Colorado, using the Claritas Prizm model, he writes:

‘One hundred percent of the town is comprised of the “New Ecotobia” segment, the trendiest rural lifestyle category. These are true professional, environmentally conscious telecommuters, the epitome of the small-town living, big city incomes megatrend.’

Are there no teachers, farmers, storekeepers, cleaners, or gardeners in Telluride? Claritas describes the segment as a blend of educated urban exiles and older farm families but does not specify the mix. Obviously these two groups could be expected to have very different incomes and lifestyles to each other. Harry Dent has forgotten to mention the older farm families but even so, the point remains that the description of Telluride as 100% New Ecotobia is misleading. The fallacy of averages strikes again.

While it may be convenient to simplify the description of an area down to one “segment” in some cases, it may be less misleading to specify the mix of the main two or three demographic types.

A Case Study Based on 1996 Census Data for New South Wales

I have sought to further explore the “birds of a feather flock together” hypothesis by analysis of 1996 Census data for New South Wales (Basic Community Profile). There are 12,735 collectors districts and the population within each averages 528 people.

The first characteristic analysed is family type and relationship in household (composition of household for short). The main categories are couple family with children; couple family without children; one parent family; group household; lone person; visitors from within Australia and visitors from overseas; and others. It is
expected that these types of households would have very different lifestyles from each other.

Chart 1 shows the distribution of the proportion of the population in each collectors district by composition of household. The box and whisker chart is interpreted as follows:

- The small box represents the median (half of the cd’s have a lower value and half have a higher value). For couple families with children, the median is 51.85% of the population.
- The larger box indicates the upper and lower quartiles. For couple families with children, the lower quartile (25% of cd’s have a lower value) is 38.57% of the population and the upper quartile (25% of cd’s have a higher value) is 61.59% of the population.
- The “whiskers” represent the 1 and 99 percentiles. For couple families with children, 1% of cd’s have a component of less than 3.35% and 1% have a component greater than 80.0%.
- The stars and circles represent outliers and extreme values and represent less than 1% of cd’s.

Chart 1

One of the implications of Chart 1 is that most cd’s have 50% or more of the population in couple family with children households and relatively few cd’s have less than 38% of the population in this type of household. Thus, every geodemographic segmentation scheme should specify what percentage of the cluster’s population is comprised of this type of household.

Also, there are virtually no cd’s where the majority of the population is in couple families without children, single parent families, group households, or lone
households. Thus, no segment should exist that is described only as one of these types of household (or a subset thereof).

Chart 2 presents a similar analysis based on age group.

**Chart 2**

![Distribution of Proportion of Population by Age Group](image)

Again, it can be observed that there are virtually no cd’s where any of these broad age groups is in the majority and so no segment should exist that is described only in terms of an age group.

These analyses have been based on very simple demographic measures. When income, ethnicity and other factors are included, and so further fractionate the population, it is clear that there is insufficient demographic clustering to justify segments such as “young singles” let alone “young single migrants”.

Of course, there could be groups of cd’s that contain a higher than average component of “young singles” but that group of people must still be in the minority and, therefore, not fully descriptive of the population. A higher than average concentration of “young singles” can be very valuable for targeted direct marketing, but it must be realised that most households reached will not be in the target market. This implies high wastage and a need to be conscious of the possible damage of badly targeted communication with the majority of recipients.

To further investigate this conclusion, we have developed a set of segments based on both household composition and age of the population in each cd (Table B22). In practice, a much larger set of variables would be used including income and ethnicity, for example.
The segments have been ranked on the basis of the proportion of the population living in couple families with children. Chart 3 shows the first three segments. There is a high degree of clustering in segments 1 and 2 in that more than 50% of the population are living in couple families with children. Segment 1 comprises primarily younger families while segment 2 is mainly made up of older families. Together these two segments make up over 56% of cd’s. Segment 3 does not have as much concentration in couple families with children and includes a relatively high level of empty nest couple only families.

**Chart 3**

![Segmentation on Age and HH Composition (1 of 3)](chart)

While more than half the population in segments 1 and 2 are in couple families with children, this is the type of household with the highest number of people on average. Taking this into account, less than half of the households in these clusters are couple families with children.

Thus, even for segments with the highest concentration, less than half of all households are of a particular type. In using such segments for gaining marketing efficiencies, it is important to consider the wastage and potential damage that can ensue. At best, a large proportion of households may not be interested in your offer and at worst, most may be offended because they are not eligible. As an example, a retail electricity company in Australia recently bundled discount home loans with their electricity offer in a mailing. As most households do not have a mortgage (less than 30% do) those who did not qualify for the deal may be offended and perceive that they were paying more for their electricity because of deals offered to other households.

In the next four segments (Chart 4), couple families with children still represent more of the population than other types of household, but they are areas with a high degree of mixing of household types.
Segments 8 and 9 are even more mixed (Chart 5) while segments 10 and 11 have high levels of visitors.

**Chart 4**

*Segmentation on Age and HH Composition (2 of 3)*

1996 NSW cd's

**Chart 5**

*Segmentation on Age and HH Composition (3 of 3)*

1996 NSW cd's
Chart 6 maps the segments for the Sydney area. Some segments are highly clustered geographically, while others are disjoint and likely to be more costly to reach.

Chart 6

These segments are quite effective at reaching parents with young children. For example, Segment 1 is 18.1% of cd’s but contains 29.0% of children aged 0 to 4. Segments 1, 2, and 3 collectively represent 67.8% of cd’s but contain 80.7% of children aged 5 to 14.

However, when we more tightly define our target by including dimensions such as income and ethnicity, geodemographic segments are likely to provide less leverage.

**Population Mobility**

Population census data ages quite rapidly. Over a five year period, about 41% of the population shifts to another address within Australia. This is comprised of:

- 12% who move within their Statistical Local Area (or SLA, similar to local government area in some states);
- 15% who move to another SLA within their Statistical Division (SD, eg capital city or regional area);
- 7% who move to another SD within their state;
- 5% who move to another state.
Note that this data does not include people who moved overseas between one census and the next.

In October 2002, foreseechange conducted a telephone survey of 500 adults asking about likelihood of moving over the next five years. The proportion who said that they were very likely or likely to move was 39%, very similar to the historical figures.

The locations these people would move to are:

- overseas 4.4% of the total adult population;
- interstate 10.2%;
- within the state 23.8%

Census data in Australia is at least one year out of date when it becomes available and up to six years out of date before the next census data is available.

When using census data, at best 8% of people will have moved and at worst nearly half of the population will have moved. They will probably have moved out of the collectors district and may not have been replaced by an identical household.

This is a large potential source of error in geodemographic segmentation schemes.

The situation is worse in USA where the intercensal period is 10 years.

**Rapidly Changing Household Composition**

Another reason for geodemographic segmentation schemes becoming out of date is rapidly changing household composition in aggregate. In 1996, 53.6% of the population lived in a couple family with children. This type of living arrangement is projected to fall to 42.6% by 2011. Large growth is projected for lone households and couple only households.

**Using Other Data**

Demographic data is only loosely correlated with behaviour. Other important dimensions include attitudes. There is, of course, a correlation between attitudes and demographic measures such as age but the point remains that even if we clustered on the basis of the demographics of individual households, these clusters would only partially explain behaviour. Clustering on the basis of groups of 300 households only worsens the explanatory power of the segments.

Other data can be used to improve the explanatory power, and hence the effectiveness of using geodemographic segments.

Some vendors use the electoral roll or billing data to improve the explanatory power. For example, if we know that someone called Susan lives at a given address then we also know the probability distribution for the year that she was born. Some names are popular at particular times and not others and this can sometimes limit the likely age range.
Some market research companies conduct many interviews over the course of a year and can provide attitudinal and purchasing behaviour data, as well as demographic data, for small areas such as postcode areas.

In addition, people within an area face similar property prices, transport, shopping, and lifestyle options and this can be used to advantage if the defining characteristics can be discovered. Obviously, for example, we would expect most people living in Manly to share the characteristic of liking the beach.

**Conclusion**

Geodemographic segmentation can be a useful tool for effectively reaching certain types of household, but it is an imperfect tool. When using such segments for deciding which areas to target for distribution of advertising or where to locate stores, we recommend that you ask yourself and your vendor of the segments the following questions.

1. What types of household other than those you are targeting will also be reached and what level of wastage and damage could be done as a result? In other words, ask for the mix of households you will reach and avoid the fallacy of averages. Remember that more than half the households you reach will not be in your target segment.
2. Is there a reasonable relationship between the households you will reach and the behaviour you are trying to influence?
3. Is there a danger that the data on which the segments were based is too out of date to be useful?
4. Is there any other data that could be employed to improve the accuracy of your targeting?
5. Would you be better off developing a customised segmentation scheme that best meets your needs, rather than a generic scheme?

If you are correlating addresses with geodemographic segments to identify good prospects, remember the potential errors described above and that correlation does not imply causation. Use other data sources to validate your conclusions.

**More Information**

foreseechange ([www.foreseechange.com](http://www.foreseechange.com)) is a forecasting and futures research company that is involved in developing more insightful geodemographic segmentation schemes. Charlie Nelson’s email address is charlien@bigpond.com.

QED develops user-friendly, powerful, and customisable geodemographic mapping and analysis tools. Nadine Wake’s email address is nadine.wake@qed.com.au.