

# Looking for Socio-Geographic Profiles from Survey Data: the Case of Home Ownership in the French Riviera

Giovanni Fusco  
University of Nice-  
Sophia Antipolis  
98 Bd E. Herriot  
Nice, France  
giovanni.fusco@unice.fr

Floriane Scarella  
University of Nice-  
Sophia Antipolis  
98 Bd E. Herriot  
Nice, France  
floriane.scarella@unice.fr

## Abstract

This paper presents a modelling application for interview/survey data analysis on home ownership within a metropolitan area. The goal of the application is to develop an exploratory methodological protocol to analyse quantitative and qualitative data covering several themes: preferences, behaviours, motivations and objective constraints of home ownership decisions, as well as spatial inscription of residential trajectories and of daily mobility habits. Its final result is the identification of coherent socio-geographic profiles capable of segmenting new home owners within the study area of the French Riviera. Home ownership decisions can thus be considered in the broader context of lifestyles.

Socio-geographic profiles are identified and characterized by a new five step clustering protocol using Bayesian Networks (BN). BN modelling can integrate both qualitative and quantitative data, and deal with uncertainty issues. Six profiles are identified for home buyers in the French Riviera and namely two urban profiles (one by choice, the other by constraint), a profile of urban periphery dwellers and two suburban profiles, having different metropolitan localization. Only one profile has no particular geographic localization. Beyond social position and present geographic localization, many other characteristics contribute to the socio-geographic profiles: biography, social inclusion, localisation and local environment of the previous dwelling, and, above all, the motivations and the logics underlying home purchase. One of the main results of the analysis was to show how the metropolitan context shapes several important aspects of home purchase decisions. The proposed protocol could be used as a heuristics for further discourse analysis.

*Keywords:* Home Ownership, Bayesian Networks, Bayesian Clustering, Socio-Geographic Profiles, French Riviera Metropolitan Area

## 1 Introduction

Construction of new sprawling residential subdivisions, urban densification, city centre decay or revival, are apparently different outcomes of housing policies within a city. Attempts to control future urban form through plans and regulation are, however, often thwarted by market behaviours which are difficult to master. Home ownership decisions seem to play an important role in such behaviours and even more within the context of integrated metropolitan areas, where city-centres, suburbs and rural villages constitute overlapping housing markets. It becomes thus important to study the complex interweaving of preferences, behaviours, motivations and objective constraints of home ownership decisions of households [4, 8, 14]. Going beyond the scopes of a socio-economic analysis, elements of the spatial context within which these decisions are taken will have to be fully taken into account [1,5].

The modelling application presented in this paper was developed within a wider research project federating sociologists (UMR LAMES / University of Aix-Marseille, VECT Mare Nostrum / University of Perpignan) and geographers (UMR ESPACE / University of Nice). The objective was to study the determinants and motivations of new home owners in metropolitan contexts within southern France. The research was exploratory: new questionings were formulated and tested over small population samples. Three fields were selected: the metropolitan area around Marseilles,

the metropolitan area of the French Riviera and the smaller regional metropolitan area around Perpignan. A semi-directive interview protocol was elaborated and used in order to question 160 households having bought their dwelling within the last 10 years (more precisely between 1997 and 2010).

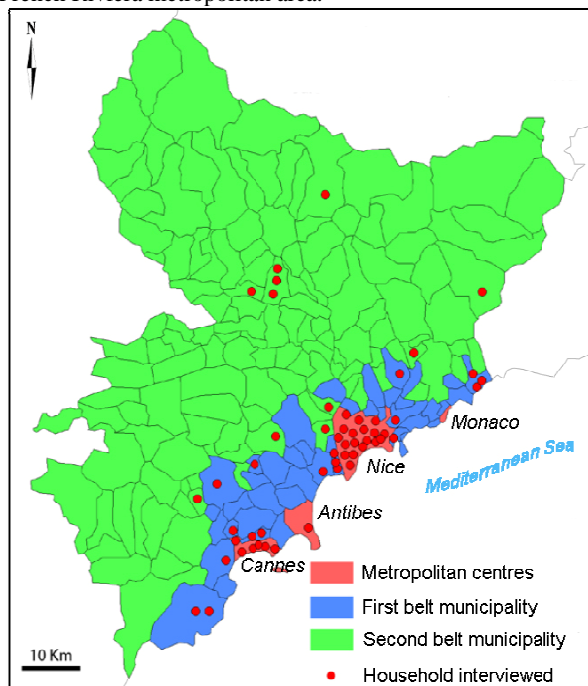
Within this context, we wanted to test new methodologies for numeric treatment of interview/survey data, combining sociological aspects revealed by interview (or open question survey) answers, and spatial information extracted from the residential trajectory of the households. The objective was to identify coherent socio-geographic profiles capable of segmenting new home owners within the study area. The coherence of the profiles would result from grouping recurrent logics within revealed preferences, behaviours, socio-economic conditions and spatial inscriptions of residential and daily mobility. Home ownership decisions can thus be considered in the broader context of lifestyles, going beyond contingent considerations of available housing units and/or transportation facilities. Looking for socio-geographic profiles is also a first attempt to generalize the logics within individual household interviews, although in an exploratory way, given the extremely limited size of the interviewed panel. This attempt was made through the application of new clustering protocols based on Bayesian artificial intelligence.

The next section will present the study area and the methodology employed for survey data segmentation. The following section will show the results obtained. A final section will resume these results and evaluate advantages and shortcomings of the modelling application.

## 2 Methodology

The new clustering methodology was applied only to the French Riviera survey data. The French Riviera is a polycentric metropolitan coastal region of over one million people, which can be roughly subdivided in three zones according to a centre-periphery gradient having important implications for the real estate market (figure 1): the coastal metropolitan centres (the cities of Nice, Cannes, Monaco and Antibes) concentrating most of the economic activity, the first belt of the close periphery including all the municipalities at less than 20 minutes from the main centres and covering most of the coastal area, the second belt of the further hinterland, including the municipalities under metropolitan influence in the alpine valleys beyond the coastal area. This geographic subdivision, as well as age class (less than 40, 40-60 and more than 60 years old) have been used to stratify the sample of 54 households interviewed, based on knowledge of localization and age of main home purchases within the study area thanks to the FILOCOM database of the French Ministry of Finance. Within this exploratory application, a simple quota sampling method was implemented. A few characteristics distinguish home ownership in the French Riviera compared to other French metropolitan areas (with the exception of Paris): the relative importance of apartments over houses within main homes (apartments make up 76% of main homes purchased and 85% of previous dwellings within the sample) and the importance of the financial effort for home ownership, both in terms of the total cost of the transaction and of the portion of household income devoted to housing. Overall satisfaction for housing conditions splits the household sample, with 44% affirming to live in their ideal (or almost ideal) dwelling and 56% showing different degrees of dissatisfaction.

Figure 1: Localization of the household sample within the French Riviera metropolitan area.



Interview questions cover several themes, namely the socio-economic characteristics of the household, the localization and the characteristics of both the present and the previous dwelling, the search and purchase of the present dwelling, the neighbourhood integration of the household and its future housing projects. Interviews were first recorded and interviewees' answers were subsequently coded and stocked in a small database (see table 1 at the end of the text) having the following characteristics:

- Data are mainly qualitative (yes/no, very satisfied/quite satisfied/dissatisfied, etc.).
- Quantitative data are discrete or have been discretized.
- The localization of the present and previous dwelling is characterized both at the metropolitan scale (metropolitan municipalities / first belt / second belt) and more locally in terms of morphological characteristics of the neighbourhood (city-centre, peripheral urban tissue, suburban tissue, rural village). One of the characteristics of the study area is the presence of suburban environments even within the metropolitan municipalities.
- The database is very small (54 records).

How can socio-geographic profiles be identified among new home owners in the French Riviera, in order to generalize the narratives of every single interview? We propose to take up this challenge by a new clustering application using Bayesian Networks (BN, [7, 9, 11, 12]). The application follows a pioneering protocol which has been lately applied in marketing studies [3]. BN modelling can integrate both qualitative and quantitative data, hard and soft modelling hypotheses and deal with uncertainty issues. BN allow two main kinds of applications. The first is causal knowledge discovery: BN can infer the most probable causal links among the variables from a database of observed records, even in the presence of missing and/or uncertain data. The second kind of application is knowledge inference through Bayesian simulation. Once the causal model is established, BN can perform probabilistic simulation with extreme freedom in the choice of input and output variables. Combining these two kinds of applications, BN can also become powerful clustering tools, as we will show through our application. More specifically, the clustering protocol is structured in several phases, coupling supervised and unsupervised learning from the database. The protocol was implemented using the BayesiaLab software [2].

### 2.1 Step 1 – Unsupervised learning of associations

The first step is the search of probabilistically strong links among the 46 survey variables through unsupervised learning of a BN, a classical knowledge discovery application. At this stage, no previous knowledge is entered by the modeller over the BN structure. Despite the small size of the database, it was possible to produce a network containing fairly robust links from a probabilistic point of view. The first BN thus obtained is made up of 26 nodes, linked by 26 edges. After having eliminated the weakest edges (identified through an edge strength analysis), the same 22 variables and 22 edges (figure 2) are obtained through the application of different BN learning algorithms.







### 3.5 Native city dwellers, optimizing their urban environment within the household life cycle

The fifth profile covers almost 13% of the sample, too. It concerns mainly city dwellers (both within city centres or village centres) born in the French Riviera. Moving to city centres from peripheral/suburban localisations is sometimes also observed. In acceding to home ownership these city dwellers try to preserve and optimize their urban location, in order to adapt to changes in the household life-cycle (both singles and couples with children make up the profile) and to improve the comfort and the immediate environment of the dwelling. These households are relatively wealthy; they often opt for recent apartments with limited use of loans. Their spatial and modal habits are relatively complex: car use for commuting, use of several modes for leisure activities both in the neighbourhood and further away. Like households in the previous profile, they feel that they live in a privileged context within the French Riviera (although it's an urban instead of a suburban context) and have a strong feeling of local belonging and show low probability of future residential mobility (0.28).

### 3.6 Young single active people buying old dwellings in the city centres due to financial constraints

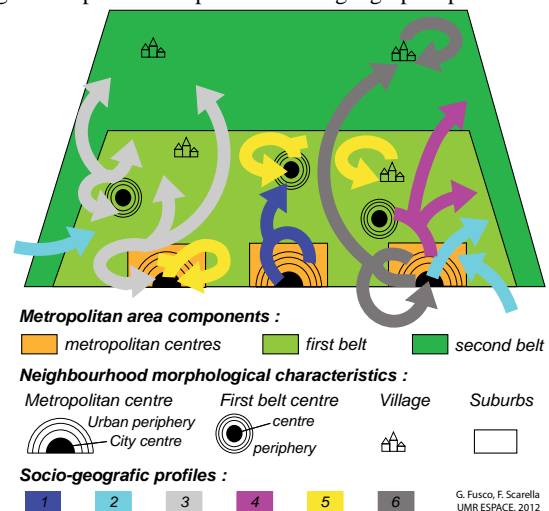
The last profile covers almost 13% of the sample, too. Like profile 2 it is characterised by many variables. It concerns mainly young single active people without children whose purchase motivations are linked to work constraints (with probability 0.72). Less probable, but not negligible, are couples with children and retirees within the profile. The common point for these households is nevertheless their limited income. They are thus forced to buy old dwellings in not coveted city neighbourhoods or villages, as well as to take out heavy and long-term loans. They normally possess a single motor vehicle (even if they don't have private parking) but also use transit and walking on a daily base. Contrary to the two previous profiles they feel they live in disadvantaged residential environments.

### 3.7 Inscribing profiles in geographic and social space

Bayesian analysis of interview data could thus reveal two urban profiles (one by choice, the other by constraint), a profile of urban periphery dwellers and two suburban profiles, having different metropolitan localizations. Only profile 3 has no particular geographic localization for the present dwelling. Spatial inscription of socio-geographic profiles becomes even more interesting when the whole residential trajectory of households is taken into account. Going beyond a simple map of profile distribution in space, Figure 5 schematizes and visualizes these trajectories. As far as social position is concerned, we have a middle-class (mainly employees) profile, two upper-middle-class profiles, a profile with more modest income and two profiles with more diversified socio-economic content. Beyond social position and present/past geographic localization, many other characteristics contribute to the socio-geographic profiles: biography, social inclusion,

spatial inscription of daily and leisure mobility, and, above all, the motivations and the logics underlying home purchase. One of the main results of the analysis was to show how the metropolitan context shapes several important aspects of home purchase decisions. Of course, these profiles don't make up an exhaustive segmentation of households within the French Riviera metropolitan area (tenants were not considered), nor of all home owners (only most recent home buyers were interviewed). Moreover, the extremely limited sample possibly missed other non negligible profiles within home buyers.

Figure 5: Spatial inscription of socio-geographic profiles.



## 4 Conclusive Remarks

The results obtained from the proposed analysis protocol show the interest of cross analyzing variables on conditions, terms and motivations of home purchase with geographic variables inscribing the residential trajectory within the metropolitan space. The origins and the destinations of these residential movements, taking into account the local characteristics of the residential environment and the localization within the metropolitan area, are indeed essential to understand the underlying logics of home purchase. These results couldn't be obtained from more traditional multivariate clustering techniques [6, 15], as done in [1] and [5]. Our five step Bayesian protocol can more particularly be compared to factor analysis followed by hierarchical clustering. These techniques need quantitative data structures, whereas interviews constitute a rich corpus of data with few numeric variables. Qualitative information is therefore central. Only textual statistics [10] could outmatch BN clustering in taking into account qualitative information. The BN protocol has however the advantage of a "controlled" reduction of model variables, of easy concept discovery through variable associations and of careful uncertainty appraisal in clustering.

Generalization of analysis results is clearly limited by interview sample size and sample method. As far as sample size is concerned, given the limited number of values of the original variables (2 to 4), a sample ten times as big as the one we used could already give a satisfactory estimation for the

parameters of the Bayesian networks employed. An appropriate sampling frame would then be necessary to obtain unbiased stratification. The proposed application is thus more a prototype of Bayesian statistical modelling than a complete analysis of socio-geographic profiles within French Riviera home buyers. It shows nevertheless the possibility of identifying profiles which are probabilistically quite robust, even from a very limited database. The methodology also shows the interest of a human/computer interaction in analysing qualitative data from interviews. This approach could be used as a heuristics in order to determine typical profiles for a more detailed discourse analysis [13] of interview recordings.

Other potential developments can be foreseen for scenario building through probabilistic simulation in the BN. We could thus infer probable residential choices of growing households or of first time home buyers, within a given socio-geographic profile and under the assumption of short-term stability of profiles. Time series analysis could be used in order to prospect the future share of different profiles. The main limit that has to be overcome in order to develop such applications is the availability of large databases of interview recordings and/or detailed open answer surveys. A future perspective is thus to apply the proposed methodology to larger databases containing a geographic dimension, like household mobility surveys.

## References

- [1] Hans S. Andersen. Explanations for preferences in Denmark for surroundings and location of the home, *ENHR Conference*, Prague, 28.6-1.7.2009.
- [2] Bayesia. *BayesiaLab Documentation*, version 4.6, Bayesia, Laval, 2009.
- [3] Bayesia – Repères. Bayesian Networks: a new tool for consumer segmentation. In *Skim Conference*, Barcelona, May 28th 2008.
- [4] Antonio da Cunha et al. *Mobilité résidentielle, aspirations des ménages et transformations de l'habitat : L'agglomération lausannoise*. Observatoire Ville et Développement Durable, Université de Lausanne, 2004.
- [5] Eva Debusschere. Segmentation of the Flemish housing market, *ENHR Conference*, Prague, 28.6-1.7.2009.
- [6] Joseph F. Hair et al. *Multivariate Data Analysis*. Prentice Hall, 5th edition, 1998.
- [7] Finn V. Jensen. *Bayesian Networks and Decision Graphs*. Springer, New York, 2001.
- [8] Tom Kayko. Expressions of housing consumer preferences: Proposition for a research agenda, *Housing, Theory and Society*, vol. 23, n. 2, p. 92-108 2006.
- [9] Kevin B. Korb, Ann E. Nicholson. *Bayesian Artificial Intelligence*. Chapman & Hall, Boca Raton, FA, 2004.
- [10] Ludovic Lebart, André Salem, Lisette Berry. *Exploring Textual Data*. Springer, Berlin, 1998.
- [11] Judea Pearl. *Causality. Models, reasoning and inference*. Cambridge University Press, Cambridge, 2000.
- [12] Olivier Pourret, Patrick Naim, Bruce Marcot, editors. *Bayesian Networks: A Practical Guide to Applications*. John Wiley & Sons, New York, 2008.
- [13] Jan Renkema. *Introduction to Discourse Studies*. John Benjamins, 2004.
- [14] Lionel Rougé. Inégale mobilité et urbanité par défaut des périurbains modestes toulousains. *EspaceTemps.net*, 25.04.2007.
- [15] Lena Sanders. *L'analyse statistique des données en géographie*. Alidade/RECLUS, Montpellier, 1989.

Table 1: List of variables coded from interview transcriptions.

VARIABLE CODE	Variable Definition	Number of Values
AGE	Age of the interviewee,	3
ORIGINE	Geographical origin of the interviewee,	4
CSP	Socio-professional status of the interviewee,	4
REVENUS	Household income of the interviewee,	3
NB_ACTIFS	Number of active people in the household,	3
SITUATION_FAMILIALE	Household typology of the interviewee,	3
ZONE_LGM	Metropolitan zone of the present dwelling,	3
HABITAT	Morphological characteristics of the neighbourhood of the present dwelling,	4
TYPE_LGM	Typology of the present dwelling,	2
PIECES	Number of rooms of the present dwelling,	3
JARDIN	Presence of a private garden in the present dwelling,	2
PARKING	Presence of a private parking in the present dwelling,	2
PISCINE	Presence of a private swimming pool in the present dwelling,	2
PARKING_COLL	Presence of a common parking within the subdivision/condominium of the present dwelling,	2
JARDIN_COLL	Presence of a common garden within the subdivision/condominium of the present dwelling,	2
PISCINE_COLL	Presence of a common swimming pool within the subdivision/condominium of the present dwelling,	2
PERIODE_LGM	Period of construction of the present dwelling,	4
AUTRE_LGM	Ownership of another dwelling,	3
ZONE_LGM_ANT	Metropolitan zone of the previous dwelling,	4
TYPE_LGM_ANT	Typology of the previous dwelling,	2
HABITAT_LGM_ANT	Morphological characteristics of the neighbourhood of the previous dwelling,	4
TAILLE_LGM_ANT	Size difference of previous and present dwelling,	3
STATUT_LGM_ANT	Occupancy status of the previous dwelling,	3
MOTIF	Main reason for the purchase of the present dwelling,	4
CRITERE	Priority criterion in the choice of the present dwelling,	4
PART_DEPENSES_LGM	Portion of monthly household income devoted to housing,	3
PERIODE_ACHAT	Purchase period of the present dwelling,	4
TYPE_ACHAT	Typology of the purchase of the present dwelling,	3
TEMPS_RECHERCHE	Search time for the present dwelling,	3
DIFFICULTES_RECH	Difficulties encountered within the search for the present dwelling,	4
COUT	Total cost for the purchase of the present dwelling including construction/renovation,	3
TRAVAUX	% of construction/renovation costs for the present dwelling,	3
CREDIT	Portion of loan financing for the purchase of the present dwelling,	4
ESPACE_QUOTIDIEN	Space frequented daily by the household,	3
ESPACE_LOISIRS	Space frequented regularly by the household for leisure activities,	3
MODE_NAVETTE	Transportation mode used for daily commuting,	4
MODE_LOISIRS	Transportation mode used for leisure activities,	3
VEHICULES	Number of motor vehicles owned by the household,	3
TEMPS_NAVETTE	Travel time for daily commuting for the interviewee,	4
PERCEPTION_QUARTIER	Social perception of the neighbourhood of the present dwelling by the interviewee,	2
APPARTENANCE	Feeling of belonging to a geographic space by the interviewee,	4
REL_VOISINAGE	Relationship of the household with the neighbours,	3
MOBILITE_FUTURE	Project of future residential mobility,	3
TRAVAUX_FUTURS	Project of future renovation works in the present dwelling,	3
ENGAGEMENT	Engagement of the household in the social life of the local community,	2
LGM_IDEAL	Correspondence between the present dwelling and the (almost) ideal dwelling of the interviewee,	2