An Esri® White Paper • July 2010

2010 Methodology Statement: Esri® Data—Tapestry™ Segmentation for Block Groups and ZIP+4s

Esri, 380 New York St., Redlands, CA 92373-8100 USA
TEL 909-793-2853 • FAX 909-793-5953 • E-MAIL info@esri.com • WEB www.esri.com

2010 Methodology Statement: Esri Data—Tapestry Segmentation for Block Groups and ZIP+4s

An Esri White Paper

Contents **Page**

2010 Methodology Statement: Esri Data—Tapestry Segmentation for Block Groups and ZIP+4s

For the past 30 years, companies, agencies, and organizations have used segmentation to divide and group their markets to more precisely target their best customers, prospects, citizens, residents, members, and donors. Segmentation systems operate on the theory that people with similar tastes, lifestyles, and behaviors seek others with the same tastes—"like seeks like." These behaviors can be measured, predicted, and targeted. Segmentation explains customer diversity, describes lifestyle and lifestages, and incorporates a wide range of data.

Tapestry™ Segmentation represents the fourth generation of geodemographic market segmentation systems that began with the first mass release of machine-readable, small-area data from the 1970 Census. The availability of hundreds of variables for thousands of neighborhoods was both irresistible and daunting for marketers. What they needed was a structure—a way to create information from an overwhelming database. Market segments provide that structure, a system for classifying consumers using all the variables that can distinguish consumer behavior, from household characteristics like income and family type to personal traits such as age, education, or employment and even to housing choices.

Tapestry Segmentation classifies U.S. neighborhoods into 65 distinct market segments. Neighborhoods with the most similar characteristics are grouped together, while neighborhoods showing divergent characteristics are separated. Tapestry Segmentation combines the "who" of lifestyle demography with the "where" of local neighborhood geography to create a model of various lifestyle classifications or segments of actual neighborhoods with addresses—distinct behavioral market segments.

Statistical Methods

Cluster analysis is the generic approach used to create a market segmentation system. There are a number of different techniques or clustering methods that can be applied to identify and classify market types. Each technique has its strengths and weaknesses. Previous generations of Tapestry Segmentation have been built using a combination of techniques, such as the iterative partition K-means algorithm, to create the initial clusters or market segments, followed by application of Ward's hierarchical minimum-variance method to group the clusters. The combination has provided a complementary match of the strengths of each technique. Tapestry Segmentation combines the traditional with the latest data mining techniques to provide a robust and compelling segmentation of U.S. neighborhoods. Esri developed and incorporated these data mining techniques to complement and strengthen traditional methods to work with large geodemographic databases. Robust methods are less susceptible to extreme values, or outliers, and, therefore, crucial to small-area analysis. The traditional methodology of cluster analysis has a long track record in developing market segmentation systems. Complementary use of data mining techniques and implementation of robust methods enhance the effectiveness of traditional statistical methodology in developing Tapestry Segmentation.

For a broader view of consumer markets, cluster analysis was again used to develop the Tapestry Segmentation summary groups. Summary groups are ideal when users want to work with fewer than 65 segments. The 65 segments are combined into 12 LifeMode groups based on lifestyle and lifestage. The 11 Urbanization groups present an alternative way of combining the 65 segments based on the segments' geographic and physical features such as population density, size of city, location relative to a metropolitan area, and whether they are part of the economic and social center of a metropolitan area.

Data Used to Build Tapestry Segmentation

Cluster analysis techniques are essentially heuristic methods that rely on exploratory procedures to arrive at stable and optimal solutions. The key to developing a powerful market segmentation system lies in the selection of the variables used to classify consumers. U.S. consumer markets are multidimensional and diverse. Using a large, well-selected array of attributes captures this diversity with the most powerful data available. Data sources include Census 2000 data; Esri's Updated Demographics; Acxiom Corporation's InfoBase-X® consumer database; and consumer surveys, such as the Survey of the American Consumer from Gfk MRI, to capture the subtlety and vibrancy of the U.S. marketplace.

Selection of the variables used to identify consumer markets begins with data from Census 2000, the most accurate source of data on the American consumer. Census data is collected directly from the population, self-reported, then edited by the Census Bureau for consistency. Data includes household characteristics such as type single person or family, income, relationships (single or multigenerational), and owner/renter status; personal traits such as age, sex, education, employment, and marital status; and housing characteristics like home value or rent, type of housing (single family, apartment, or townhouse), seasonal status, and owner costs relative to income. In essence, any characteristic that is likely to differentiate consumer spending and preferences is assessed for use in identifying consumer markets.

The selection process draws on Esri's experience in working with the 1980 and 1990 censuses and includes a range of multivariate statistical methods, in addition to factor analysis, principal components analysis, correlation matrices, and graphic methods. Selecting the most relevant variables is critical to defining homogeneous market segments; however, determining the most effective measure of each variable is equally important. Is income best represented by a median, an average, or an interval? Would household or disposable income best measure actual buying power? In the end, selection was narrowed to more than 60 attributes to identify and cluster U.S. neighborhoods by market type.

From the neighborhood or block group level, Tapestry Segmentation profiles enable the comparison of consumer markets across the country by state, metropolitan area, county, place, census tract, ZIP Code™, and even congressional districts. However, direct mail campaigns frequently require a smaller target than a ZIP Code. To improve targeting capabilities and capture the diversity of the population within a block group, Tapestry Segmentation data is also provided at the ZIP+4 level.

Tapestry at the ZIP+4 Level

Because ZIP+4 is a postal designation represented by addresses, not boundaries, Esri built the ZIP+4 model using list-based household data. List data is different from census data. It represents an address list of households overlaid with information compiled from

a variety of different lists and sources, both public and private. Most lists include 120 to 130 million household records—more than the actual number of households in the country due to duplicate records. Addresses range from complete street address information to post office boxes, which confound the best geocoders.

The Census Bureau provides detailed analyses of its data collection, coverage, validation procedures, rates of imputation for incomplete items, and estimate errors. List providers are more circumspect regarding both their sources and validation procedures. However, census data cannot be provided for individuals or households due to Title 13, which protects the confidentiality of data collected. To provide market segments for addresses, list-based data is the only option.

To create segmentation at this level, Esri used InfoBase-X data from Acxiom Corporation. Acxiom compiles its lists and data from an unprecedented number of data sources including public real estate information, purchased data from catalogs, auto dealerships, consumer surveys, publications, product registrations, and telephone directories.

Esri aggregates household attributes from the InfoBase-X data to assign a Tapestry Segmentation code to each ZIP+4. By building the Tapestry Segmentation markets with census data, updating them with Esri data, and assigning the markets to ZIP+4 codes using an aggregate of household attributes, Esri has developed an effective way to use the vast amount of information from list compilers while maintaining the integrity of the data. Tapestry Segmentation ZIP+4 classifies each U.S. ZIP+4 by one of Tapestry's 65 distinctive market segments. Excluding post office box and business ZIP+4s, approximately 24 million residential ZIP+4s are currently available in Tapestry Segmentation ZIP+4.

Verification Procedures

Verification procedures follow the creation of the segments to ensure their stability and validity. Replicating the segments with independent samples checks stability. Validity is checked through characteristics that are not used to generate the segments. Linking Tapestry Segmentation to the latest consumer survey data is the critical test. A market segmentation system must be able to distinguish consumer behavior—spending patterns and lifestyle choices—as expected. Esri verifies the efficacy of its Tapestry Segmentation markets against the consumer surveys from GfK MRI, which include nearly 6,000 product and service brands in 550 categories, along with readership of hundreds of magazines and newspapers, Internet usage, TV viewership by channel and program, radio listening, and use of Yellow Pages. When creating Tapestry Segmentation at the ZIP+4 level, Esri employs the same rigorous verification tests that are used to check the block group-level markets. The validity check provides the answer to the most important question: Does it work? It works.

For more information about Tapestry Segmentation, visit www.esri.com/tapestry or call 1-800-447-9778.

Esri's Data Development Team

Led by chief demographer Lynn Wombold, Esri's data development team has a 30-year history of excellence in market intelligence. The combined expertise of the team's economists, statisticians, demographers, geographers, and analysts totals nearly a century

of data and segmentation development experience. In addition to Tapestry Segmentation, the team also develops Updated Demographics, Consumer Spending, Market Potential and Retail MarketPlace datasets, which are now industry benchmarks.

For more information about Esri® Data, visit www.esri.com/data.

About Esri

Since 1969, Esri has been helping organizations map and model our world. Esri's GIS software tools and methodologies enable these organizations to effectively analyze and manage their geographic information and make better decisions. They are supported by our experienced and knowledgeable staff and extensive network of business partners and international distributors.

A full-service GIS company, Esri supports the implementation of GIS technology on desktops, servers, online services, and mobile devices. These GIS solutions are flexible, customizable, and easy to use.

Our Focus

Esri software is used by hundreds of thousands of organizations that apply GIS to solve problems and make our world a better place to live. We pay close attention to our users to ensure they have the best tools possible to accomplish their missions. A comprehensive suite of training options offered worldwide helps our users fully leverage their GIS applications.

Esri is a socially conscious business, actively supporting organizations involved in education, conservation, sustainable development, and humanitarian affairs.

Contact Esri

1-800-GIS-XPRT (1-800-447-9778)
Phone: 909-793-2853
Fax: 909-793-5953
info@esri.com
www.esri.com

Offices worldwide
www.esri.com/locations

380 New York Street
Redlands, CA 92373-8100 USA

Printed in USA