Product Plots: Visualizing Demographics in Purchasing Datasets for Marketing Analysis and Insight

Scott Emmons

Cyberinfrastructure for Network Science Center, Indiana University

& Bloomington High School North

1320 E. 10th St., Room 020, Bloomington, IN 47405

scott@scottemmons.com

Phone: 812.606.7839

Product Plots: Visualizing Demographics in Purchasing Datasets for Marketing Analysis and Insight

Abstract

Demographic *data*, which Ailawadi et al. (2001, p. 81) call "… both significant and intuitively appealing," has long been important to marketers. In the age of "big data," the difficulty for the marketer is making sense of demographic *datasets*. How is a marketer to interpret his or her large volume of data?

First, this paper draws on current marketing literature to create a conceptual framework for the use of demographic data in marketing campaigns. The framework contains four key areas: defining the *target*, or to whom to market; selecting the *reach*, or where to place an ad; identifying the *angle*, or how to appeal to the target audience; and assessing the *use*, or how users will interact with an advertisement.

Second, this paper introduces a novel information visualization approach called "product plots" to make demographic datasets informative to marketers. I develop an online, interactive visualization tool to plot products as points in demographic space and find that marketing insights can be gained from the absolute location of products in space, the relative location of products to each other, and the structure of the plots as a whole.

Finally, I encourage readers to explore the tool for themselves by visiting "http://www.scottemmons.com/productplots/".

Keywords: marketing, demographics, big data, information visualization, data visualization

DEMOGRAPHIC DATA, DEMOGRPAHIC DATASETS, AND THE MARKETER

Overview

Consumers are the driving force of business. Demographic data, or quantifiable information about consumers, has long been important to businesses.

Demographic attributes play a universal, underlying role in consumer behavior. They reflect the life stages and life cycles of consumers and influence consumers' final decision-making.

Marketers in particular use demographic data to make their campaigns more focused and efficient. This paper introduces a new conceptual framework to discuss the role of demographic information in marketing campaigns. The framework divides a marketing campaign into four key areas:

Target – to whom should the campaign appeal?

Reach – where should the message be placed for it to be seen?

Angle – how will the campaign appeal to the target audience?

Use – how will the audience interact with the advertisement?

Ailawadi et al (2001, p. 81) said that the relationship between demographic *information* and consumer attitudes and behaviors is "… both significant and intuitively appealing." Today, in the age of "big data," demographic information is becoming ever more accessible to businesses. As Erdem et al. (2001) noted, the implications of this data have profound potential to impact business.

The challenge lies in making sense of demographic *datasets*. How is the marketer to make sense of his or her large volume of data?

Through data visualization, this paper seeks to make demographic datasets both intuitively appealing and significant to marketers. Using a visual model I call “product plots”, I plot products as points in space based on their demographic attributes. In this way the visualization illustrates the structure of the dataset created by the demographics, making its comprehension both intuitive and useful.

The remainder of the paper is organized as follows. First, I use the conceptual framework introduced above to discuss other significant works related to the use of demographic data. Second, I describe the data used in this paper to construct the product plots. Third, I explain my methods and construct the product plots. Fourth, I analyze the characteristics of the resulting visualizations and give some specific examples of their potential application. Finally, I conclude by discussing the strengths and weaknesses of product plot analysis.

The Role of Demographic Data: A Brief Review

The current marketing literature accepts demographics as being universal and supplementary to marketing campaigns.

Universally, demographics influence final purchasing decisions as they both define a consumer’s current life stage (Du & Kamakura, 2006) and significantly correlate with his or her psychographic characteristics (Ailawadi et al., 2001). While these relationships are often intuitive, it can be difficult to identify precisely which demographic attributes are key (Macé & Neslin, 2004).

Concretely, the influence of demographics may be secondary to other decision-making factors. For example, the number of infants in a household influences the need for diapers, but factors such as brand loyalty and convenience of store location ultimately determine which product the consumer purchases.

Previous studies have evaluated the impact of demographic factors for marketers. This paper introduces a conceptual framework with four keys areas of a marketing campaign—the target, reach, angle, and use—to discuss the role of demographic data.

To identify the *target* of a campaign, the primary role of demographics is to profile consumers into specific segments for greater focus. One of the most significant methods of consumer profiling is dividing consumers into distinct life stages. Du and Kamakura identified thirteen distinct stages in consumer lifecycles, which range from a single or young couple with no child to a retired or old couple with adult dependents. These life stages are significant in consumer segmentation and annual household expenditures (Du & Kamakura, 2006).

To identify the *reach* of a campaign, the primary role of demographic information is to select the most appropriate medium for an advertisement. Marketers seek to place ads where they will be seen by the target consumer. For example, marketers use demographic information to identify and target specific readership segmentations in magazines, and magazine companies use demographic information to price their ads accordingly. This relationship was examined by Koschat and Putsis (2002).

To identify the *angle* of a campaign, the primary role of demographic information is to understand the purchasing decision tree of consumers. Demographic information has a significant, indirect effect on consumer psychographics. Consumers' price sensitivity (Ailawadi et al., 2001; Erdem et al., 2001), brand loyalty (Cooil, Keiningham, Aksoy, & Hsu, 2007), eye for product quality (Ailawadi et al., 2001), and creation of post-promotion dips in sales (Macé & Neslin, 2004) are all influenced by demographic attributes. Marketers can capitalize upon these relationships by appropriately tailoring campaigns to consumer disposition, such as providing sales and coupons to price-sensitive consumers.

To identify the *use* of a campaign, the primary role of demographic information is to understand how the consumer will interact with advertisements. For example, demographics impact the search costs a consumer is willing to undertake before purchasing a product both online and offline (Ailawadi et al., 2001; Du & Kamakura, 2006); a consumer's responsiveness to the functionality of a website and online advertisements (Danaher, Mullarkey, & Essegaier, 2006); and a consumer's tendency to explore different product options (Ailawadi et al., 2001). Understanding how consumers use advertisements enables marketers to appropriately design marketing campaigns.

CREATING THE PRODUCT PLOTS

The Data Used

This paper uses data from the National Consumer Study (NCS), a cross-sectional consumer survey conducted by Experian Marketing Services. I had access to the Spring 2010 and Spring 2011 Adult 6-Month studies through a subscription to client software published by Experian called "Simmons OneView." I chose to use data from the NCS because it simultaneously covers demographic attributes and consumer purchasing decisions.

The survey is conducted in two stages. In the first, participants are interviewed over the phone to determine survey eligibility; in the second, participants fill out a self-administered questionnaire booklet.

The goal of the NCS is to "measure all American adults… with a uniform questionnaire and consistent data collection and data processing procedures." ("Simmons Survey Methodology," 2011, p. 1) The NCS uses a stratified sample design and weights data based on response rates in order to be nationally representative.

Simmons OneView provides access to both the original and weighted data. In this study I used the weighted, nationally representative data because it corrects for differences in sample size among different groups of respondents.

This study looks at two kinds of survey data. The first is demographic information about the respondents. I study information respondents gave about their age, gender, marital status, education level, employment status, and hours worked weekly; the number and age of the children in the household; the size and income of the household; and the presence of other adults in the household.

The second is information about consumer behavior, which includes purchasing decisions and media habits. This study examines 794 consumer products which respondents report that they currently use or have purchased recently, and it examines a total of 255 magazines and television programs they have read and watched.

Due to the sheer scope of the National Consumer Study I pruned the original dataset. The NCS contains questions about all kinds of lifestyle habits, which range from daily activities to attitudes and opinions. My goal was to retain a focus on commercial purchases. I included all survey questions about goods that can be purchased commercially, such as groceries and clothing, and excluded questions concerning the respondents' general lifestyle, such as location at a certain time of day and preferred means of transportation.

The original survey data can be exported from Simmons OneView in question-answer format. For example, one data point reads "Men's belts: do you wear?: yes." To aid in viewer comprehension, I parsed each data label to read only the product name; the previous data point is simply labeled "men's belts."

The greatest strength of the dataset is that it simultaneously covers each respondent's demographic information, purchasing decisions, and media behavior.

The greatest drawback of the dataset is that the questions for all products are not framed in the same way. An individual question may concern a past purchase, household ownership, or current use. To some degree this is inevitable as consumers use different products in different ways. However, it means that a respondent who uses a given product may not be its original owner purchaser.

Visualizing the Data

This analysis aims to use visualization to make demographic datasets intuitively appealing and useful to marketers. To accomplish this I plotted points in *demographic space*.

Specifically, I created two-dimensional space and defined each axis by a demographic attribute. See Figure 1.

[Insert Figure 1 from Tables and Figures, p. 17]

Because the space is defined by demographic attributes, each point in space is characterized by a pair of demographic values. Space in the upper right marks old men; the lower right, young men; the lower left, young women; and the upper left, old women.

The key to this analysis is that demographic data about consumers *quantitatively describes* the goods they purchase. Thus, products can be plotted as points in demographic space.

With perfect information this is an easy task. See Figure 2.

[Insert Figure 2 from Tables and Figures, p. 18]

However, I was limited in my analysis by the data to which I had access. Simmons OneView only provides binned responses to the demographic variables of age, years of schooling, number of hours worked, and household income.

To fit discontinuous binned data into plots with continuous space, I made the assumption that each value in a bin corresponds to the average of the bin's range. For example, I assumed that binned data like that of Table 1 (a fictitious example) was discrete data like that of Table 2.

[Insert Table 1 and Table 2 from Tables and Figures, p. 17]

I used the discrete data to calculate and plot the mean of the responses, which is 40.83 in this example.

Spatially, this discrete value assumption and mean calculation allows us to make the transformation shown in Figure 3.

[Insert Figure 3, from Tables and Figures p. 19]

The assumption that each value in the bin corresponds to the average of the bin's range overlooks nuances in the survey data and misrepresents it. However, I feel that the assumption's gains are worth its costs; namely, that an illustration of the concept of continuous demographic space is more valuable to the scope of this analysis than accurately representing binned data.

Online, Interactive Tool

I used the above method to calculate demographic values for each product in the dataset. Then, I used a JavaScript library called Data Driven Documents (D3) to create a program and visualize the results.

The program is an interactive tool that lets users choose which demographic variable to place on each axis. This functionality empowers users to ask and answer questions about the data for themselves. I encourage readers to explore the tool online at http://www.scottemmons.com/productplots/.

PRODUCT PLOTS

The Visualizations

[Insert Figure 4, from Tables and Figures p. 20]

[Insert Figure 5, from Tables and Figures p. 21]

[Insert Figure 6, from Tables and Figures p. 22]

[Insert Figure 7, from Tables and Figures p. 23]

[Insert Figure 8, from Tables and Figures p. 24]

Drawing Insight from the Images

Visualization, plotting the products as points in demographic space, makes demographic datasets intuitive and useful to marketers. Much can be learned from the absolute and relative locations of each product point.

The absolute location of each product in space makes intuitive its demographic attributes. For example, take the age vs. gender plot, shown in Figure 4. The x-axis is defined by the percent of a product's consumers that are male; products farthest left on the plot—such as mascara and tampons—are purchased solely by women, and products farthest right—such as chewing tobacco and cigars—are purchased mostly by men. In this way the "% male" demographic characteristic of each product is intuitive. Through visualization, demographic properties are seen.

In each plot, there is a large, highly dense, central cluster of products. These products are representative of the demographic attributes of the general population of consumers, such as "pretzels" and "mustard" (Figure 4). Surrounding the cluster are products that are defined by the plot's demographic attributes. The greater the magnitude of their distance from the central cluster, the more they are defined by a particular demographic attribute.

The relative location of each product in relation to other products makes the significance of demographic attributes intuitive. Knowing only that chocolate is roughly a 45% male product

(Figure 4) is not helpful without a frame of reference. Does being “45% male” make chocolate one of the most female-dominated products on the market? Other products on the plot give the answer. Because chocolate is only slightly left of the central cluster in Figure 4 and there are many products farther left than it, chocolate is only slightly dominated by female purchasers.

The visualization is useful to marketers in identifying the target, reach, angle, and use of marketing campaigns.

To identify the *target*, marketers can use the structure of the product space to identify attributes that distinguish their product from the population as a whole. For example, Figure 4 shows that hair spray and porcelain figurines are bought by older women.

Marketers can also use the product space to answer specific questions about their products’ demographic makeup. For example, a marketer may wish to see which products line up with the life stages studied by Du and Kamakura (2006). Four of these life stages are a small household couple with children less than seven years of age, a large household couple with children less than fifteen years of age, a large household couple with older children, and a small household couple with children less than fifteen years of age.

Figure 5, a plot showing the average age of children in the household versus the size of the household, illustrates how products relate to these variables. In this dataset there is a strong, direct, positive correlation between household size and age of children; few products are bought primarily by small households with old children or by large households with young children. This is explained by the strong correlation that the age average of children in a household shares with number of children and household size. Households with older children generally have more children and thus larger households.

To identify the *reach*, marketers can use the demographic plots to look for overlaps between the demographic segmentation of their product and the demographic segmentation of media outlets. For example, it is possible to overlay media, such as magazine titles and television programs, onto the product plots based the demographic attributes of their viewers. See Figure 6. In this way the demographic overlap between products and media outlets is visually apparent.

To identify the *angle*, marketers can use the plots to better understand their consumers. For example, prior studies have linked greater household size and lower household income to an increase in price sensitivity (Ailawadi et al., 2001; Erdem et al., 2001). Figure 7, a plot showing household income versus household size, shows how products relate to these variables. Consumers of products in the lower right of the graph, who live in a large household with a low annual income, are likely to show high price sensitivity. Marketing tactics such as coupons and sales can be especially effective to reach these consumers.

To identify the *use*, marketers can again use the demographic plots to understand their customers. For example, Du and Kamakura (2006) linked employment status and number of children to search costs. Consumers who work many hours and have many children have less time to browse products. Thus Figure 8, a plot showing number of children versus hours worked weekly, reveals products for which the consumer does not have much time to explore. Note that this data represents the average of all of a product's consumers. This average is why some products have 0.5 children and no product reaches 40 hours worked weekly.

As another example, Danaher et al. (2006) studied the relationship between the age of consumers and their visit duration on websites of various levels of functionality. It found the existence of an inflexion point at age 29; those above the age visited sites for less time as the

sites' functionality increased, while those below it visited sites with more functionality for longer periods of time. Any plot showing the age of consumers illustrates this property.

For the marketer, the plots can reveal opportunities to sell products across different industries to the same customer. While this requires greater knowledge about the lifestyle of consumers, the product plots reflect these trends.

For example, in Figure 1 video game consoles such as "Xbox 360" and "PlayStation" lie near "energy drinks." As it turns out, many video game players use energy drinks to stay awake during long, late-night gaming sessions. For this reason the energy drink companies Red Bull and Monster sponsor professional video gaming teams.

A final use of the product plots is to find opportunities for cross-promotion between products. For the manager of a store that sells many different products, the plots will provide intuition about the demographic relationships between those products. Products that cluster closely together can be advertised in the same way.

CONCLUSION

The Impact of This Work

This paper sought to make demographic *datasets* both intuitive and useful to marketers. In the age of big data, business leaders have tremendous amounts of demographic data. The challenge is to make sense of it.

Conceptually, this paper used visualization to make the demographic characteristics of products both intuitive and useful. By plotting products as points in demographic space, both the products' absolute and relative locations reveal information about the demographic attributes of their consumers.

Empirically, this paper examined metadata from Experian Marketing Services on a wide range of consumer products. I created a tool available at http://www.scottemmons.com/productplots/ to explore this data through information visualization.

Managerially, this paper introduced a conceptual framework and an interactive tool for use in the development of marketing campaigns. The absolute location of products in demographic space enables the marketer to focus the target, reach, angle, and use of marketing campaigns. The relative location of products in demographic space reveals the relationships between products, opening up opportunities for cross-promotion and increased marketing efficiency.

Strengths, Weaknesses, Next Steps

The strength of the product plots is their intuitive representation of products' underlying demographic attributes. The visual representation enables marketers to understand and explore the structure of their data.

However, as with all visualization of big data, a view of the big picture comes at the cost of underlying details. The greatest weakness of the product plots is that the demographic attributes of each point are the *average* of the attributes of each consumer, and there is no way to tell more detailed consumer segmentation from the average. For example, all the viewer may know about a product is that it has an average age of 35. Whether the product is bought primarily by those in their mid-thirties, or equally by those in their twenties and forties, is unknown.

It is important to keep in mind that demographics serve as a guide for understanding consumers. While they provide useful insights, additional information is required for a comprehensive understanding of consumers.

In this paper, I chose only to visualize two dimensions of demographic data at a time because it is intuitive to comprehend and can accommodate a variety of different demographic attributes. Opportunities for further research are to develop visualizations that show more than two demographic variables at once. These visualizations could comprehensively answer a specific question. For example, a single visualization capturing the marital status, age, household size, number of children, and age of children could illustrate all the different household lifecycles identified by Du and Kamakura (2006).

Acknowledgments

This work would not have been possible without Dr. Katy Börner, Victor H. Yngve Profesor of Information Science at Indiana University, who provided continual guidance and support throughout this project.

I would also like to thank Dr. Ray Burke, Professor of Marketing and E.W. Kelley Chair of Business Administration at Indiana University, for marketing insights. Christina Sheley, Head of Business/SPEA Information Commons, provided expertise with Simmons OneView. Adam Simpson, Software Developer at the Cyberinfrastructure for Network Science Center, offered programming advice. Ted Polley, a former research assistant at the Cyberinfrastructure for Network Science Center and current Social Sciences Librarian at IUPUI University Library, gave suggestions and feedback. Don Adams, Teacher at Bloomington High School North, and Julie Adams, Chief Financial Officer and Vice President at Oliver Winery, read and reviewed drafts of this work.

Finally, thanks go to Experian Marketing Services for use of the National Consumer Study data. More information about the study can be found at: http://www.experian.com/simmons-research/national-consumer-studies.html.

REFERENCES

Ailawadi, K. L., Neslin, S. A., & Gedenk, K. (2001). Pursuing the Value-Conscious Consumer: Store Brands Versus National Brand Promotions. *Journal of Marketing*, *65*(1), 71–89. doi:10.1509/jmkg.65.1.71.18132

Cooil, B., Keiningham, T. L., Aksoy, L., & Hsu, M. (2007). A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing*, *71*(1), 67–83. doi:10.1509/jmkg.71.1.67

Danaher, P. J., Mullarkey, G. W., & Essegaier, S. (2006). Factors Affecting Web Site Visit Duration: A Cross-Domain Analysis. *Journal of Marketing Research*, *43*(2), 182–194. doi:10.1509/jmkr.43.2.182

Du, R. Y., & Kamakura, W. A. (2006). Household Life Cycles and Lifestyles in the United States. *Journal of Marketing Research*, *43*(1), 121–132. doi:10.1509/jmkr.43.1.121

Erdem, T., Mayhew, G., & Sun, B. (2001). Understanding Reference-Price Shoppers: A Within- and Cross-Category Analysis. *Journal of Marketing Research*, *38*(4), 445–457. doi:10.1509/jmkr.38.4.445.18909

Koschat, M. A., & Putsis, W. P. (2002). Audience Characteristics and Bundling: A Hedonic Analysis of Magazine Advertising Rates. *Journal of Marketing Research*, *39*(2), 262–273. doi:10.1509/jmkr.39.2.262.19083

Macé, S., & Neslin, S. A. (2004). The Determinants of Pre- and Postpromotion Dips in Sales of Frequently Purchased Goods. *Journal of Marketing Research*, *41*(3), 339–350. doi:10.1509/jmkr.41.3.339.35992

Simmons Survey Methodology: Spring 2011 NHCS Adult Study 06-month. (2011, June). Experian Marketing Services.

Tables and Figures

Product	Age: 30-34	Age: 35-39	Age: 40-44	Age: 45-49
Beer	10	15	20	15

Table 1: An example with fictitious data of the original, binned data format.

Product	Age: 32.5	Age: 37.5	Age: 42.5	Age: 47.5
Beer	10	15	20	15

Table 2: The data format resulting from the assumption that each value is the average of its bin's range.

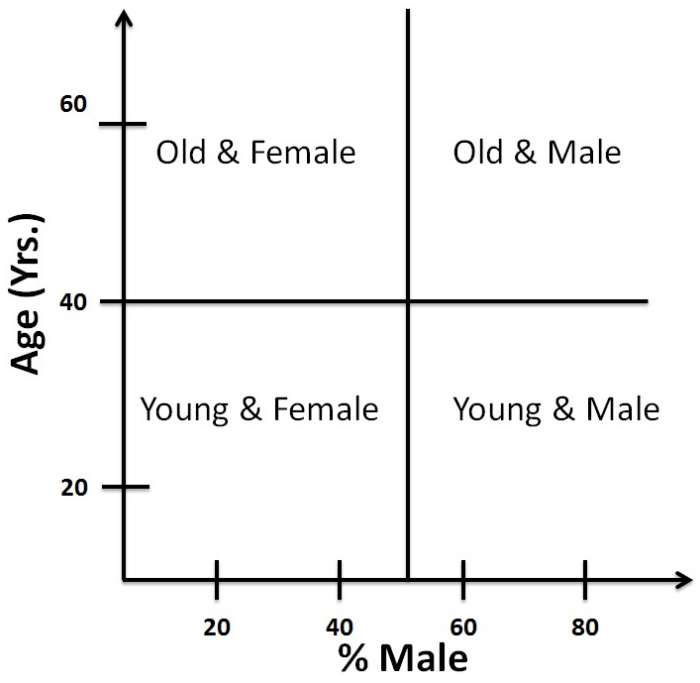

Figure 1: Two-dimensional space defined by the demographic attributes of age and gender.

Product	% Male	Average Age (Yrs.)
Beer	60	33
Hair Spray	19	42
Dentures	47	63

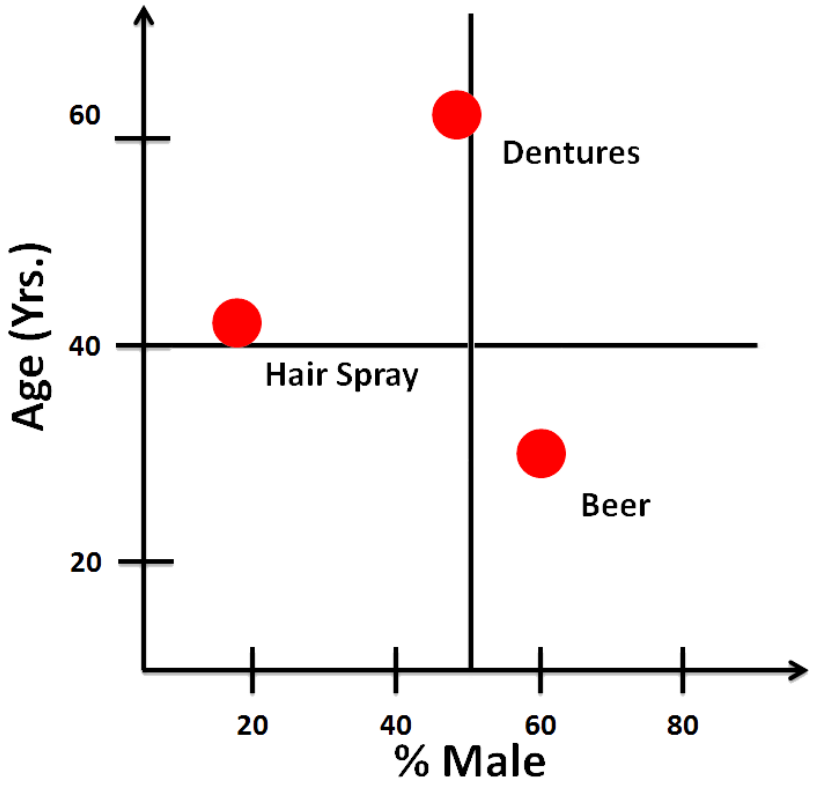

Figure 2: An example of demographic data that quantitatively describes a product and its mapping in space. Note that this data is a fictitious example.

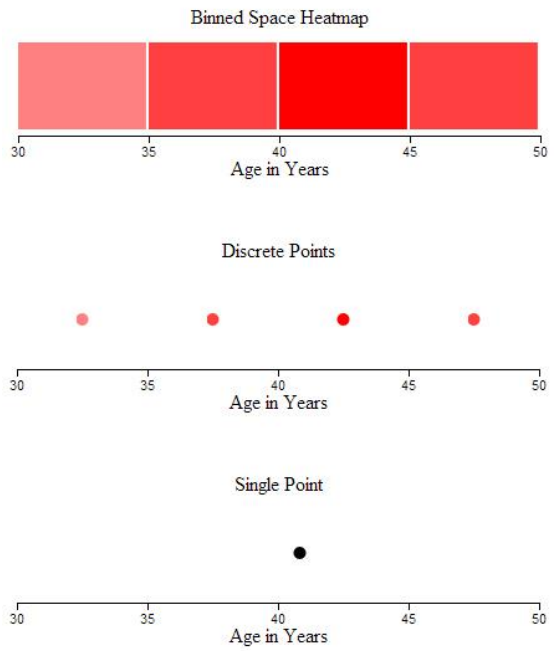

Figure 3: A spatial illustration of the discrete value assumption and mean calculation.

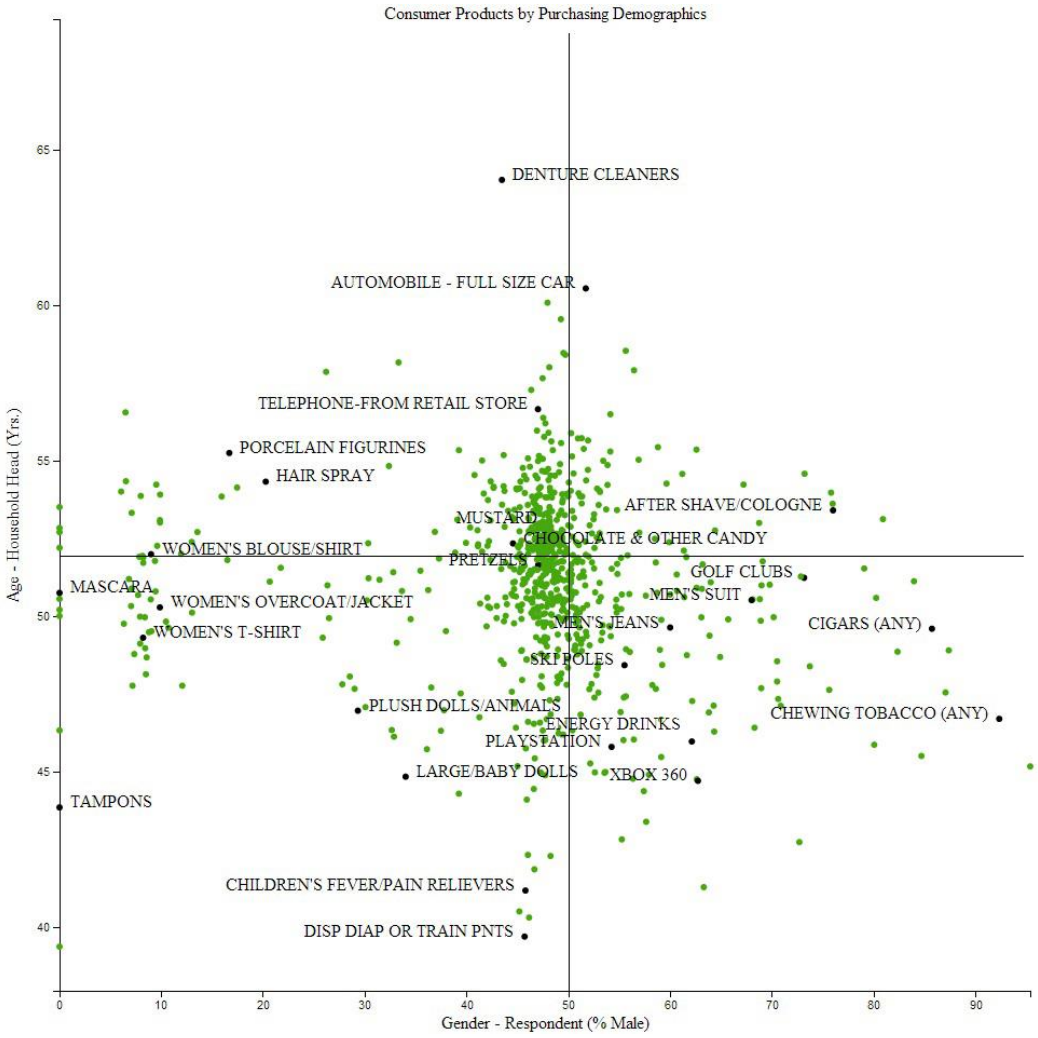

Figure 4: A mapping of consumer products based on the demographic attributes of age and gender. Data from Experian Marketing Services' Simmons National Consumer Study.

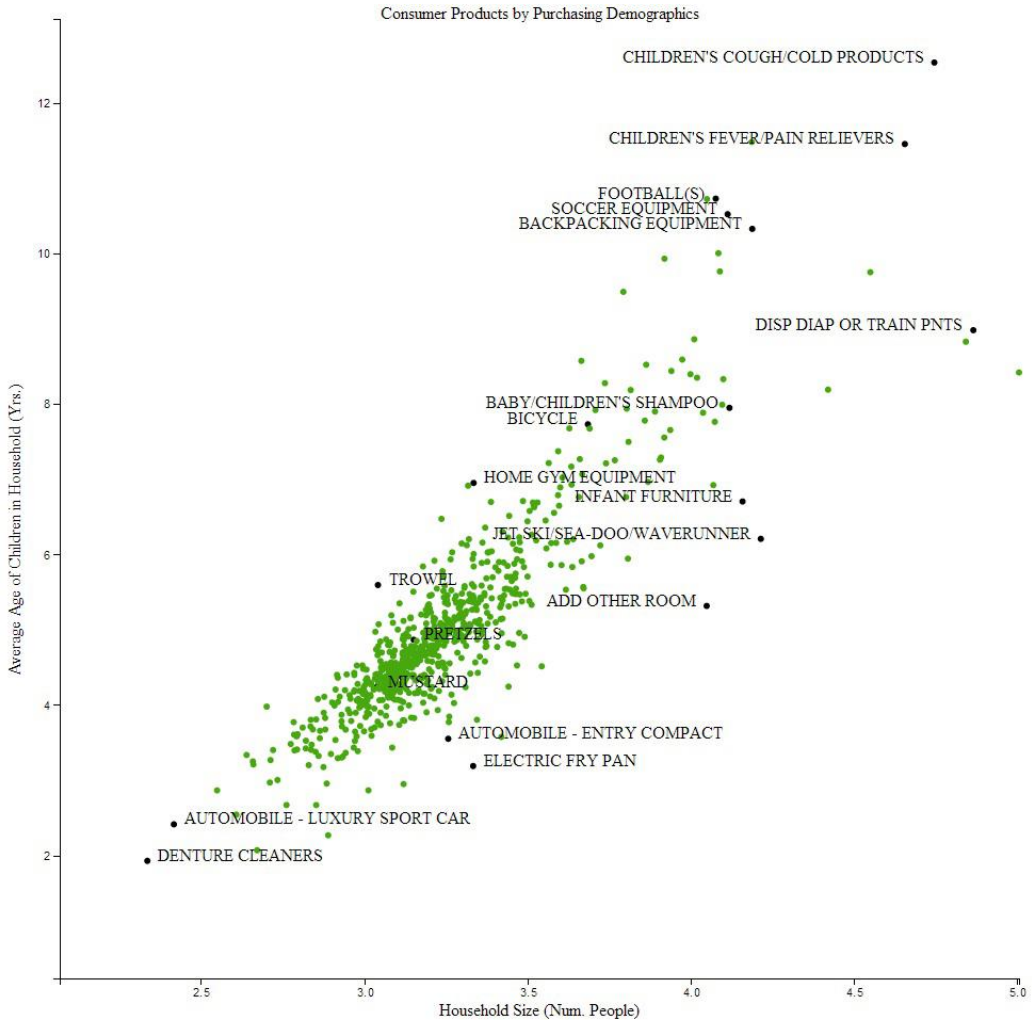

Figure 5: A mapping of consumer products based on the demographic attributes of average age of children and household size. Data from Experian Marketing Services' Simmons National Consumer Study.

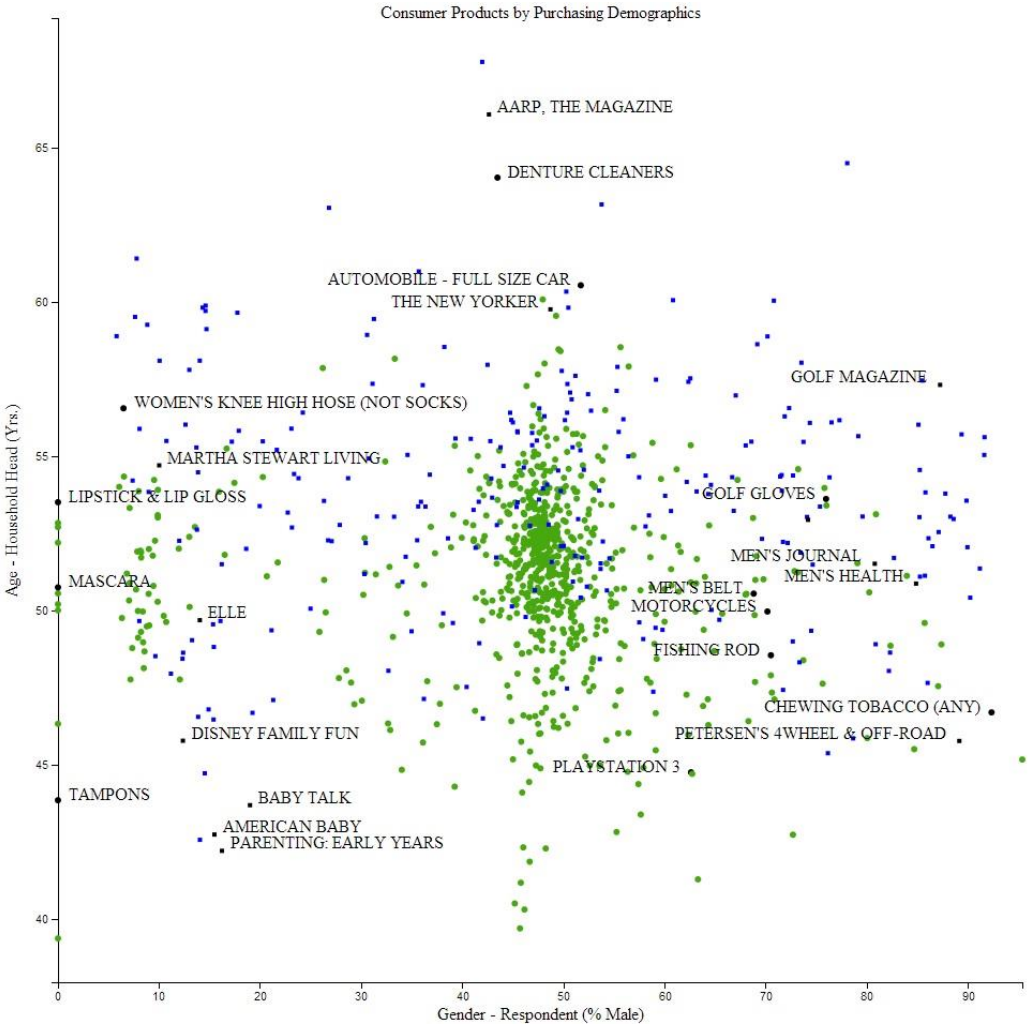

Figure 6: Magazine and television media overlaid onto Figure 4, the age-gender mapping of consumer products. Data from Experian Marketing Services' Simmons National Consumer Study.

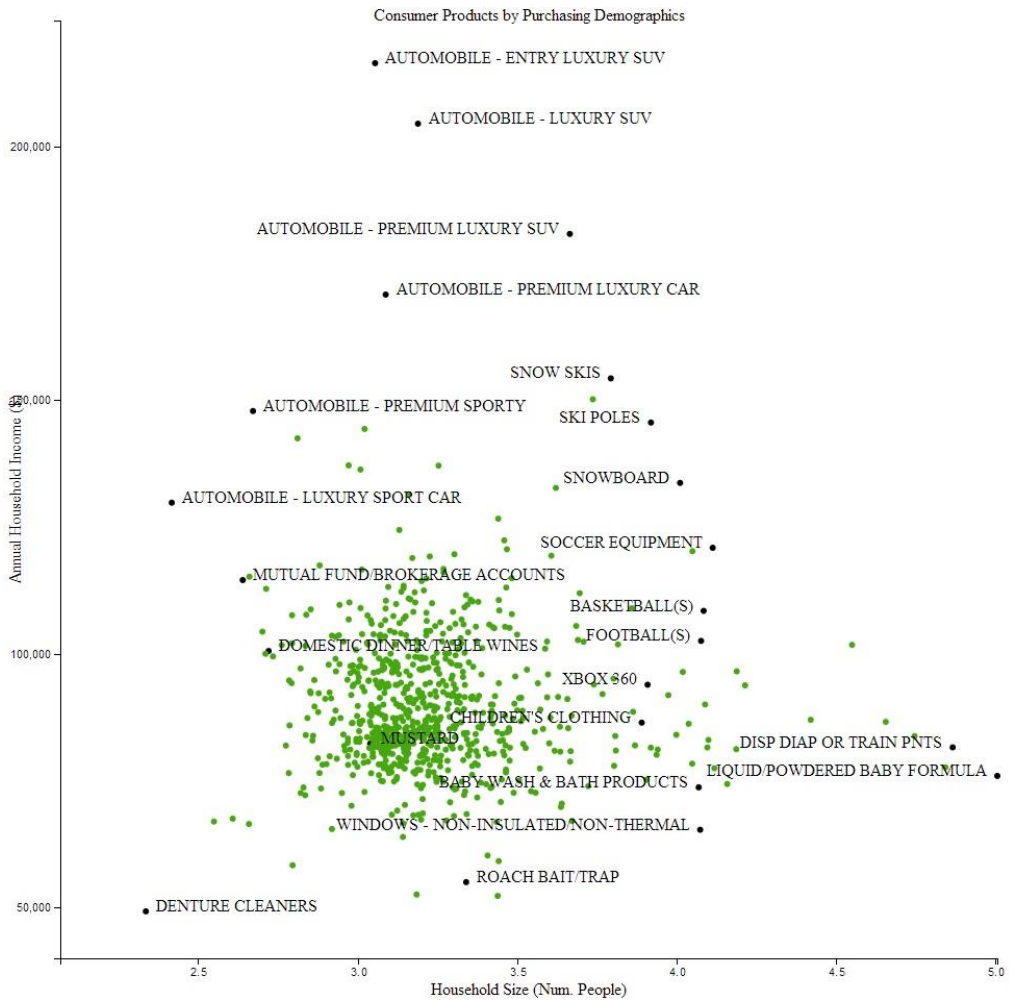

Figure 7: A mapping of consumer products based on the demographic attributes of household income and household size. Data from Experian Marketing Services' Simmons National Consumer Study.

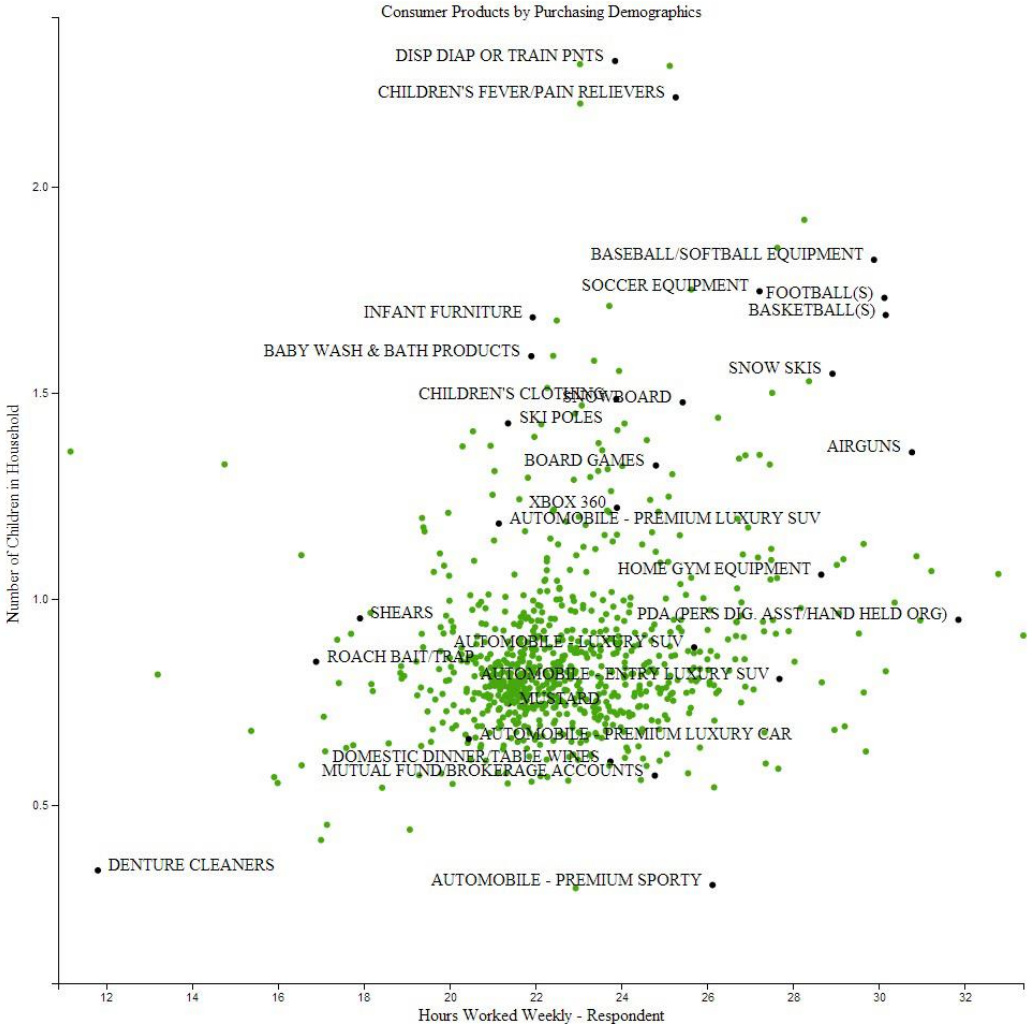

Figure 8: A mapping of consumer products based on the demographic attributes of number of children and hours worked weekly. Data from Experian Marketing Services' Simmons National Consumer Study.