

# Cluster Analysis for Extension and Other Behavior Change Practitioners: Introduction<sup>1</sup>

Laura A. Warner<sup>2</sup>

## Introduction

Helping people adopt behaviors to improve social, economic, and environmental conditions is central to Extension's mission. Extension is tasked with consistently evolving and remaining relevant to meet the current needs of a broad range of audiences. This evolution can be seen in the shift from a one-way delivery of singular mass messages to a more refined, participatory approach where specific audience needs are met through targeted, tailored programming (Monaghan et al., 2014; Warner et al., 2019). Cluster analysis is a quantitative technique that can be used to identify audience subgroups so that tailored education and communications can be designed. The purpose of this publication is to describe cluster analysis and convey its value in supporting behavior change to (1) help readers understand how this technique is applied and (2) encourage others to consider using it. The target audience is Extension professionals or other social scientists working in any disciplinary area who want to understand how cluster analysis is used or are considering using it themselves. Companion publications to the current document include "Information and Terminology Needed for Cluster Analysis," "A Practical Example," and "Integrating the Results of Cluster Analysis into Meaningful Audience Engagement."

## Audience Segmentation and Targeted Extension Programming

The benefit of audience segmentation is that Extension professionals "can deliver the programming and messages that are most meaningful to an audience/clientele segment" (Monaghan et al., 2014, para. 1). In contrast to an approach where a single message is delivered broadly to everyone, audience segmentation allows for a meaningful division of a heterogeneous group into smaller subgroups that are likely to respond similarly to a message given the "likelihood that they will clump together in meaningful ways" (Andreasen, 2006, p. 105). This approach allows for more appropriate behavioral requests as well as an opportunity to provide the most relevant (i.e., targeted) programming (Gibson et al., 2021).

Audience segmentation is as much an art as a science. There is no one right answer on how best to reach an audience. In general, some combination of psychological (e.g., attitude), behavioral (e.g., likelihood of adoption), geographic (e.g., region of Florida), and/or sociodemographic (e.g., age) characteristics are used to segment an audience (Gibson et al., 2020; Monaghan et al., 2014). Sometimes it makes a lot of sense to select a simple characteristic as the factor with which an audience is subdivided. For example, dividing the residential audience by whether the household belongs to a homeowners' association is a simple but meaningful way to target individual needs. Other times, a framework

1. This document is AEC738, one of a series of the Department of Agricultural Education and Communication, UF/IFAS Extension. Original publication date November 2021. Visit the EDIS website at <https://edis.ifas.ufl.edu> for the currently supported version of this publication.

2. Laura A. Warner, associate professor and Extension specialist, Department of Agricultural Education and Communication, UF/IFAS Extension, Gainesville, FL 32611.

or model of behavior change can be used to subdivide a potential audience (Monaghan et al., 2014). For example, the Transtheoretical Model is used to group an audience according to their orientation to a behavior (e.g., whether target audience members are aware of the behavior, considering it, ready to act, or already engaged) (Shaw, 2009; Warner et al., 2014). More complex audience segmentation strategies can be used to identify more strategic education and communication.

For example, consider an Extension professional who is involved with the development of a new variety of Florida berry that tastes delicious and requires less water to grow compared to other varieties. The Extension professional might identify growers and consumers as two distinct audience segments to which they would deliver targeted programming (focused on growing and eating the new berries, respectively). In a more complex segmentation strategy, the Extension professional might segment the grower audience into multiple segments that would receive more specific education and communications: growers who are already growing the new variety might be given information on how to market this new fruit; growers who like to be among the first to grow new varieties might be encouraged to try growing the berries as part of a variety trial; and growers who hesitate to try something new might be encouraged to learn about the new variety as new data is generated and their peers start growing it more.

## Cluster Analysis

Many times, Extension professionals and other practitioners can segment their audiences simply by using available information (e.g., by homeowners' association membership, or by whether a household hires a landscape professional or maintains their yard themselves). However, in some cases it is impossible to know how certain characteristics will relate to one another, what the ideal number of subgroups should be, or who belongs to the most important segment for a given program. Cluster analysis can be used to answer questions like these, and a key benefit is this type of analysis reveals "natural groupings (or clusters) within a dataset that would otherwise not be apparent" (IBM, 2021c, para. 1).

A cluster is a group of relatively similar cases or observations (King, 2015). Cluster analysis is an objective data reduction tool, and the ultimate purpose is to classify data into meaningful subgroups (a.k.a. clusters or segments) using a set of independent variables (Burns & Burns, 2008; King, 2015). This exploratory technique seeks to maximize the similarity of observations within the resulting clusters while also maximizing the dissimilarity between resulting

clusters (Yim & Ramdeen, 2015). Cluster analysis can be used to assign cases (i.e., people) or variables to subgroups, and this publication focuses on the former. Throughout this EDIS series, the term *case* applies to people (i.e., respondents or observations). Statistical software such as SPSS can be used to carry out these types of analyses.

Cluster analysis has been used extensively in biology and ecological sciences, but these types of analyses have not been used to their full potential in the social sciences (Yim & Ramdeen, 2015). Accordingly, social science researchers, such as those working within Extension education and human dimensions of natural resources and agricultural education contexts, may be missing out on a valuable research tool.

Cluster analyses can be classified in two major categories: hierarchical and nonhierarchical.

- Hierarchical cluster analysis starts with each case considered as its own one-case cluster and merges them until all are combined (IBM, 2021a). For example, for a health-related program, you might have 250 Extension clients you would like to group by measures of dietary quality (e.g., reported servings of vegetables per week) and exercise (reported days per week of achieving 20 minutes of exercise). You would start with 250 clusters (i.e., each Extension is their own cluster initially), and your data analysis software would combine the most similar clusters systematically until all of the 250 cases were assigned to a single cluster (i.e., one cluster containing all 250 Extension clients). Then, you would review the output from the analysis and decide which solution (i.e., number of clusters) is the most appropriate.
- Nonhierarchical cluster analysis is used to assign individual cases to the most appropriate cluster when the number of clusters is known beforehand (IBM, 2021b). For example, you may want to assign individuals to one of three clusters based on their social media use and level of interest in nutrition. Nonhierarchical cluster analysis can be used to assign individuals to those three clusters so the resulting clusters are as different from one another as possible, while maximizing similarities among individuals within the groups.

Sometimes, these two types of cluster analysis are combined. The different methods of cluster analysis and corresponding terminology will be described in detail in the next publication in this series.

## Examples of Cluster Analysis from UF/IFAS Researchers

The following are four examples of cluster analysis used within UF/IFAS. Consider the implications for Extension programming and communications given each set of subgroups identified below.

- Ali et al. (2018) segmented about 1,600 US residents using cluster analysis with respondents' value of eight benefits they received from their yard/landscape (aesthetics, environmental, food, habitat, health and comfort, monetary, social, and well-being benefits). The researchers found three subgroups based on how much respondents valued each of the benefit types. After assigning respondents to the groups based on how much value individuals placed on the benefits, the authors compared engagement in water conservation practices (e.g., using a rain sensor to turn off irrigation when it is not needed), and household and neighborhood characteristics. The following describes the three subgroups:
  - *Water-saving reward-seekers* had the highest overall value of landscape benefits and were the most engaged in water-saving practices. Members of this group were more likely to belong to a homeowners' association that prescribed penalties and rewards for unattractive and attractive landscape aesthetics, respectively.
  - *Moderate landscape appreciators* moderately valued landscape benefits and somewhat engaged in water conservation practices. Members of this group were also likely to belong to a homeowners' association that penalized residents for unattractive landscape aesthetics, but these individuals indicated their communities did not offer rewards for attractive landscapes.
  - *Necessary irrigators* had the lowest value of landscape benefits and were the least engaged in water conservation practices.

Ali et al. (2018) initially hypothesized the people in the subgroup that valued landscape aesthetics the most would be least likely to conserve water, but the cluster analysis revealed the opposite was the case. This example of cluster analysis provided some interesting implications for promoting water conservation. For example, one of several key differences between the moderate landscape appreciators and the water-saving reward-seekers was that the latter, more conservation-minded group received recognition from their homeowners' association for maintaining an aesthetically pleasing landscape.

- Khachatryan et al. (2019) surveyed 3,000 people from three states and used cluster analysis with six characteristics (knowledge about irrigation, motivations behind adopting water-saving technologies, whether individuals thought about the results of their action in the near or longterm, and whether participants were environmentally conscious or not) to generate audience segments pertaining to purchasing smart irrigation systems. The authors reported identifying four subgroups. After the subgroups were identified, the authors measured characteristics of each, such as likelihood of purchasing smart irrigation and demographic information.
  - *Proactive consumers* were not environmentally conscious, had the most knowledge about irrigation, and were thoughtful about future impact of their actions.
  - *Price-sensitive environmentalists* were environmentally conscious, had the least knowledge about irrigation, and were thoughtful about future impact of their actions.
  - *High-end professionals* were unconcerned about the future impacts of their decisions and were not particularly environmentally conscious.
  - *Content retirees* were not particularly thoughtful about future impacts of their actions, were moderately knowledgeable about irrigation, and were not particularly environmentally conscious.

This cluster analysis provided excellent information for promoting smart irrigation technologies to save water. For example, the proactive consumer subgroup was the most likely to purchase smart irrigation technologies. The researchers suggested this subgroup would be responsive to information emphasizing both short- and long-term benefits of smart irrigation. The content retirees were the least likely to purchase smart irrigation technologies, and the researchers suggested the other three subgroups might be better targets for future campaigns.

- Warner et al. (2016) used cluster analysis as an approach to encouraging residential landscape water conservation. The authors conducted a cluster analysis using engagement in 18 water conservation practices and identified three meaningful subgroups among around 1,100 Florida residents.
  - The *water-savvy conservationists* highly valued water resources and were already conserving as much as possible with little room for improvement.
  - The *water-considerate majority* were highly concerned about water resources but were less engaged in water

conservation practices (i.e., they had significant capacity for conservation).

- The *unconcerned water users* were unconcerned about protecting water resources and were unengaged in doing so.

This study provided information useful for promoting water conservation. For example, the water-savvy conservationists were not a great target for promoting water conservation because they had little room for improvement. However, this subgroup could potentially be engaged in other types of activities to protect water resources, such as volunteering for a cleanup event. The authors suggested the water-considerate majority was an excellent target for promoting water conservation because they value water resources, are motivated to change, and had substantial room for improvement in their conservation practices. The unconcerned water users were considered low priority for conservation programs. Even though they had significant capacity for improved conservation, the lack of motivation for protecting water resources meant they would be less likely to change their behaviors. The authors suggested education for this group might focus on raising awareness of water issues and improving attitudes to hopefully shift members into the water-considerate majority group with greater concern for water resources.

## Applications to Extension Programming

The examples provided above are drawn from water conservation and landscaping, but Extension professionals and practitioners could consider using cluster analysis to segment audiences in any discipline. Extension professionals and practitioners who want to change behaviors are encouraged to:

- Explore the literature to see how cluster analysis has been used within their discipline.
- Think about current audience segmentation strategies and assess how the audience is currently segmented.
- Consider, if the audience is currently segmented, whether there is a possibility cluster analysis could reveal a new approach that could improve programming efficacy?
- Consider, if the audience is not currently segmented, whether cluster analysis could possibly reveal a strategic approach for targeted programming?
- Think about potential partners with whom to collect new data or access existing data and conduct cluster analysis to reveal subgroups.

As described above, audience segmentation is a way to deliver the most relevant programming to stakeholders (Monaghan et al., 2014), and cluster analysis is an objective, mathematical approach to segmenting an audience (King, 2015). To fully integrate the results of a cluster analysis, Extension professionals and other practitioners can examine cluster information to develop communication materials and engage the segments appropriately. It may be valuable to examine descriptive information about the resulting clusters after individuals are assigned to segments. For example, water conservation may be used to assign individuals to audience segments (i.e., as the independent variables used to conduct cluster analysis), and the resulting subgroups may have different educational preferences. In this example there were three resulting audience segments. The more conservation-oriented segment was more likely to engage in active Extension education, such as volunteer activities, but there were no differences in preferences for passive education, such as reading Extension factsheets (Warner et al., 2017). Findings such as this can provide insight into appropriate Extension planning and delivery specific to audience subgroups, and Extension professionals can take information such as this to target the audience segments appropriately. There may be important differences in communication preferences (e.g., preferred social media channels), geographical location (e.g., region or county), or topical content needs (e.g., general watershed awareness versus environmental advocacy topics), among others, that can be used to guide programming. For more information about applying cluster analysis results, please see “Integrating the Results of Cluster Analysis into Meaningful Audience Engagement.”

## Summary

Cluster analysis techniques are underused but may offer significant potential for guiding audience segmentation activities. These techniques can reveal natural groupings within the broader potential audience that may be otherwise unapparent. This first publication in the *Cluster Analysis for Extension and Other Behavior Change Practitioners* series introduced cluster analysis as a technique for segmenting an audience. The other publications in this series include “Cluster Analysis for Extension and Other Behavior Change Practitioners: Information and Terminology Needed for Cluster Analysis,” “Cluster Analysis for Extension and Other Behavior Change Practitioners: A Practical Example,” and *Cluster Analysis for Extension and Other Behavior Change Practitioners: Integrating the Results of Cluster Analysis into Meaningful Audience Engagement.*”



## References

- Ali, A. D., Warner, L. A., & Kumar Chaudhary, A. (2018). Using perceived landscape benefits to subgroup Extension clientstopromote urban landscape water conservation. *EDIS*, 2018(4). <https://doi.org/10.32473/edis-wc291-2018>
- Andreasen, A. R. (2006). *Social marketing in the 21st century*. Thousand Oaks, California: Sage Publications.
- Burns, R. B., & Burns, R. A. (2008). Cluster analysis. In R. B. Burns & R. A. Burns (Eds.), *Business research methods and statistics using SPSS* (pp. 552–567). London: Sage.
- Gibson, K. E., Fortner, A. R., Lamm, A. J., & Warner, L. A. (2021). Managing demand-side water conservation in the United States: An audience segmentation approach. *Water*, 13(21), 2992. <https://doi.org/10.3390/w13212992>
- Gibson, K. E., Lamm, A. J., & Lamm, K. W. (2020). Identifying audience needs to effectively communicate about the cost of implementing sustainable farming practices. *Journal of Applied Communications*, 104(3). <https://doi.org/10.4148/1051-0834.2334>
- IBM Corporation. (2021a). Hierarchical cluster analysis. <https://www.ibm.com/docs/en/spss-statistics/28.0.0?topic=features-hierarchical-cluster-analysis>
- IBM Corporation. (2021b). K-means cluster analysis. <https://www.ibm.com/docs/en/spss-statistics/28.0.0?topic=features-k-means-cluster-analysis>
- IBM Corporation. (2021c). Twostep cluster analysis. <https://www.ibm.com/docs/en/spss-statistics/23.0.0?topic=option-twostep-cluster-analysis>
- Khachatryan, H., Rihn, A., Warwick, C. R., & Dukes, M. (2019). Who is interested in purchasing smart irrigation systems? *EDIS*, 2019(5), 7. <https://doi.org/10.32473/edis-fe1069-2019>
- King, R. S. (2015). *Cluster analysis and data mining*. Dulles, VA: Mercury Learning and Information.
- Kumar Chaudhary, A., & Warner, L. A. (2018). Understanding good irrigation and fertilization behaviors among households using landscape design features. *EDIS*, 2018(1), 4. <https://doi.org/10.32473/edis-wc292-2018>
- Monaghan, P., Warner, L., Telg, R., & Irani, T. (2014). Improving Extension program development using audience segmentation. *EDIS*, 2014(6). <http://edis.ifas.ufl.edu/wc188>
- Shaw, B. R. (2009). Using temporally oriented social science models and audience segmentation to influence environmental behaviors. In L. Kahlor & P. Stout (Eds.), *Communicating Science* (pp. 109–130). <https://doi.org/10.4324/9780203867631>
- Warner, L., Galindo-Gonzalez, S., & Gutter, M. S. (2014). Building impactful Extension programs by understanding how people change. *EDIS*, 2014(6). <http://edis.ifas.ufl.edu/wc189>
- Warner, L. A., Israel, G. D., & Diaz, J. M. (2019). Identifying and meeting the needs of Extension's target audiences. *EDIS*, 2019(3). <https://doi.org/10.32473/edis-wc336-2019>
- Warner, L. A., Kumar Chaudhary, A., Rumble, J. N., Lamm, A. J., & Momol, E. (2017). Using audience segmentation to tailor residential irrigation water conservation programs. *Journal of Agricultural Education*, 58(1), 313–333. <https://doi.org/10.5032/jae.2017.01313>
- Warner, L. A., Lamm, A. J., Rumble, J. N., Martin, E., & Cantrell, R. (2016). Classifying residents who use landscape irrigation: Implications for encouraging water conservation behavior. *Environmental Management*, 58(2), 238–253. <https://doi.org/10.1007/s00267-016-0706-2>
- Yim, O., & Ramdeen, K. T. (2015). Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data. *The Quantitative Methods for Psychology*, 11(1), 8–21. <https://doi.org/10.20982/tqmp.11.1.p008>