Curator

# Rethinking Museum Visitors: Using K-means Cluster Analysis to Explore a Museum's Audience

**Amanda Krantz, Randi Korn, and Margaret Menninger**

***Abstract*** Understanding visitors is a necessary and complex undertaking. In this article, we present K-means cluster analysis as one strategy that is particularly useful in unpacking the complex nature of museum visitors. Three questions organize the article and are as follows: 1) What is K-means cluster analysis? 2) How is K-means cluster analysis conducted? 3) Most importantly: What are the applications of K-means cluster analysis for museum practitioners? To answer these questions, we present five steps that are vital to conducting a K-means cluster analysis. We also present three cases studies to demonstrate differences among the results of three K-means cluster analyses and provide practical applications of the findings.

To help museums successfully achieve the impact they intend, museum practitioners, evaluators, and researchers need to know about visitors—including their attitudes, preferences, and previous experiences. While qualitative methodologies are often used to provide understanding of museum visitors, we find survey research and quantitative analysis to be equally insightful. In particular, we are using K-means cluster analysis, an exploratory statistical procedure that creates groups of like visitors from interval level data from a given data set, thereby providing a more descriptive understanding of visitors in a museum context.[1]

Our objective in writing this article is to help readers understand K-means cluster analysis and its application. To that end, we have intentionally avoided overly technical terms and have structured the article around three questions: 1) What is K-means cluster

*Amanda Krantz (krantz@randikorn.com) is a research associate, Randi Korn (korn@randikorn.com) is the founding director, and Margaret Menninger (menninger@sbcglobal.net) is a statistician at Randi Korn and Associates, Inc.*

analysis? 2) How is K-means cluster analysis conducted? 3) Most importantly: What are the uses of cluster analysis findings for museum practitioners? Throughout the article, we draw from our experience using K-means cluster analysis in studies for the Dallas Museum of Art, the Sports Legends Museum at Camden Yards, and the San Francisco Museum of Modern Art (Randi Korn and Associates, Inc. 2005; 2008; 2009). We use examples from these studies.[2]

## What is K-means Cluster Analysis?

There are several types of cluster analyses or clustering.[3] Regardless of the type, all clustering aims to make "natural" groups out of the given data. "Natural" is defined as "fitting and of practical use." In creating natural clusters, cluster analysis aims to form internally similar groups that differ from each other in distinct and meaningful ways (Kachigan 1991; SPSS, Inc. 2003; Tan, Steinbach, and Kumar 2006). To conceptualize this idea, it is helpful to think of the common adage, "Birds of a feather flock together."

K-means cluster analysis is a statistical algorithm that partitions visitors into a specified number of natural clusters. As an introduction to the algorithm and analysis, consider figure 1, which plots 20 visitors' ratings of two statements about sports on two 7-point rating scale variables.

In Figure 1, each small data point—whether it be a diamond, square, triangle, or circle—identifies visitors' ratings to two statements about sports. The four shapes of the data points identify visitors' cluster membership (Cluster 1, Cluster 2, Cluster 3, Cluster 4). Visitors in Cluster 1 are identified by diamonds. Visitors in Cluster 2 are identified by squares. Visitors in Cluster 3 are identified by triangles, and visitors in Cluster 4 are identified by circles.

At approximately the center of each cluster is an enlarged data point, which is the centroid. The centroid is the mean of all the data points in the cluster. Centroids are important to clustering. The distance—more technically, the Euclidean distance—of each data point to the centroid measures the proximity of the data point to the centroid, and proximity to the centroid determines cluster membership.[4]

Figure 1 is a basic visualization of four clusters' position on two variables. Nevertheless, it is useful in conceptualizing clusters, especially when thinking about the steps involved in K-means cluster analysis, which is discussed in detail in the next section.

## How is K-Means Cluster Analysis Conducted?

At this point you probably have questions, such as: What instrument do you use to elicit the data for cluster analysis? How do you decide how many clusters to specify for the analysis? Once the computer derives the clusters, how do you interpret them?

We answer these questions and others as we guide you through the steps of the analysis. To exemplify each step, we use concrete examples from studies we have conducted and provide a rationale for these steps.

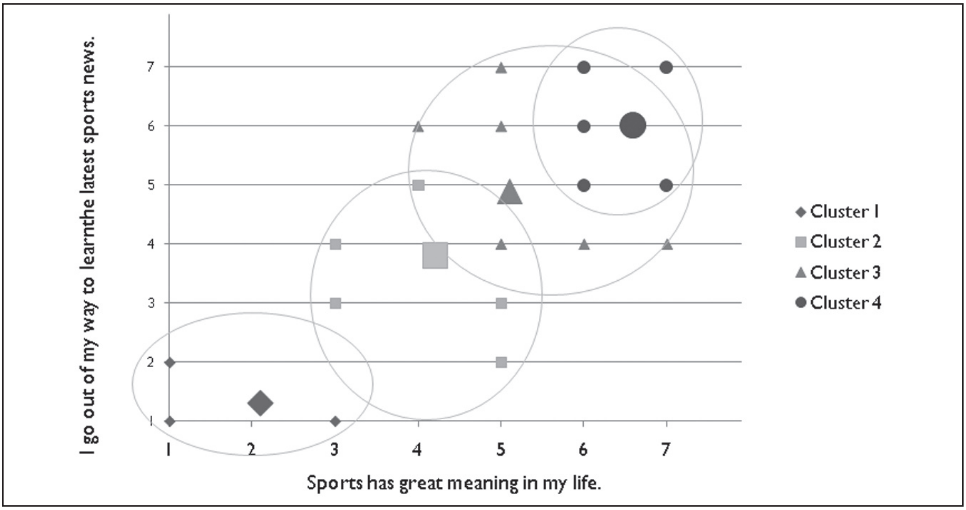


Figure 1. Scatterplot of visitors' ratings of two statements about sports.

**Step 1: Create a series of rating statements as variables by which to cluster visitors**—Rating statements are useful, since they go beyond demographics, investigating visitors' thoughts, attitudes, and beliefs. Further, rating statements provide interval data that can be analyzed using K-means cluster analysis.

In writing rating statements and determining associated rating scales, it is important to consider what you want to know and the question you want to answer. For instance, the Dallas Museum of Art wanted to understand visitors' level of engagement with art (Randi Korn and Associates, Inc. 2005), while the San Francisco Museum of Art wanted to understand engagement with art among adults visiting with children between 4 and 11 years (Randi Korn and Associates, Inc. 2009). The statements we designed for each museum are different, since each set of statements consisted of responses to the particular questions each museum was asking. In another study, the Sport Legends Museum at Camden Yards wanted to understand what types of sports fans are visiting the Museum (Randi Korn and Associates Inc. 2008). Table 1 shows the statements we designed to explore types of sports fans.

Crafting the statements and rating scales is crucial to the resulting clusters, since the statements are the variables by which the clusters are formed. The statements must directly link to your research question. If not, it is impossible to create natural clusters that are of any practical use. So, for example, when drafting the statements for the Sports Legends Museum at Camden Yards, it was necessary to investigate what attitudes and behaviors characterize different types of sports fan. In conducting such investigations, we recommend using research from the field, input from museum staff, and interviews with museum visitors to inform the statements. Further, it is imperative that you pretest these statements and the rating scale with museum visitors.[5]

**Step 2: Administer the questionnaire**—Once you have pretested your rating scale statements and made any necessary changes in the questionnaire, administer the question-

Table 1. Statements and rating scale drafted for the Sports Legends Museum at Camden Yards (Randi Korn and Associates, Inc. 2008).

| What kind of sports fan are you? | Does not describe me (1) | | | | | | Describes me very well (7) |
|---|---|---|---|---|---|---|---|
| I prefer to watch sports by going to games. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| When my team is losing, I usually feel bad. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I prefer to watch sports on TV. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I watch sports to cheer the entire team's effort. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I watch sports to see the athletes I like. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Sports has great meaning in my life. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I like learning about the history of my favorite teams. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I regularly attend sporting events. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I regularly participate in sports. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I go out of my way to learn the latest sports news. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

naire to a randomly selected sample of visitors from the population you are studying. It is important for respondents to rate *all* of the statements. The K-means cluster program on SPSS and other statistical packages can handle large samples efficiently. Most of the cluster studies at Randi Korn and Associates have sample sizes in excess of 300 respondents.

**Step 3: Decide or experiment with how many clusters visitors should be grouped into**—The K-means algorithm requires you to specify in advance the number of clusters to be derived. This can be tricky. We always try more than one cluster solution. We generally ask for two, three, and four cluster solutions, and then review the results to identify which cluster solution is natural, fitting, and practical for the museum. In Step 5, we will discuss how to decide whether clusters are natural.

**Step 4: The statistical program outputs the clusters**—If, for example, we specify three clusters, the statistical program begins the algorithm by identifying three initial centroids, often at random.[6] Remember, the centroid is the mean value on all clustering variables of the cluster's members. Next, the statistical program runs a number of passes, or iterations. With each iteration, the program reassigns each case to the cluster with the closest centroid (based on Euclidean distance), and then recalculates the centroids based on the updated cluster membership. Iterations continue until cluster membership stabilizes with a solution that *minimizes* the variability within each cluster and *maximizes* the variability between each cluster (Kachigan 1991; SPSS, Inc. 2003; Tan, Steinbach, and Kumar 2006). Again, this is a step that you do not physically do, but rather, the statistical program you are using does.[7]

**Step 5: Determine whether the clusters are natural**—This is one of the most important steps of the cluster analysis because the researchers' interpretation and judgment is required to determine whether the clusters are natural. In this way, cluster analysis is exploratory and similar to qualitative analysis.

In considering whether the clusters are natural, you should consider the number of clusters defined as well as experiment with more or fewer clusters. Choosing the appropriate number of clusters is admittedly difficult and comes through experience and experimentation with the data (Dubes 1987; Koehly et al. 2001). For instance, it is important to consider whether the cluster membership of any one cluster is too small or too large. A cluster with just 10 members is probably too small to be practical, so the researcher should specify fewer clusters and re-run the analysis. On the other hand, if one cluster is overly large and dominant, the researcher might want to re-run the analysis with more clusters.

In our experience, defining three to five clusters brings out the nuances but still creates practical groupings of visitors. However, we acknowledge that every data set is unique, and depending on the research question, there are uses for various numbers of clusters (Kachigan 1991).

Once the clusters emerge, we take a close look at each cluster by comparing their demographic characteristics, visit characteristics, psychographic characteristics, and any other variables on the questionnaire.

K-means cluster analysis does not always produce useful clusters. For example, you might find that a three-cluster solution simply reflects three age groups (three clusters: young visitors, middle-age visitors, and older visitors, for instance). That particular cluster solution doesn't add anything new to your understanding of visitors, since you have probably already examined your data for age differences. In general, however, we have found that carefully designed rating statements and scales yield data that produce useful clusters, meaning that the clusters enhance our understanding of visitors in interesting and distinctive new ways.

## K-means Cluster Analysis Is Not for Everyone

While we have found that K-means cluster analysis is useful for audience research in museums, some statisticians do not think K-means cluster analysis is rigorous enough. In particular, the random assignment of the cores of clusters is problematic to some, and the researchers' determination of natural clusters is problematic to others. Moreover, cluster results may not be robust. Adding cases to an existing data set or using an entirely new data set may yield a cluster solution that is quite different.

Nevertheless, K-means is a well-accepted method in social science research, often used in data mining and analysis of social networks, particularly because it is exploratory (Huang 1998; Tan, Steinbach, and Kumar 2006). So, like other exploratory methods, if you find an interesting result, you should view it as a jumping off point for additional, confirmatory research.

## What are the Uses of Cluster Analysis Findings for Museum Practitioners?

Before explaining how museum practitioners apply cluster analysis findings to their work, we present findings from studies at different museums. In reading the findings, we encourage you to consider the types of information presented as well as how museums could use the information.

**Visitors to the Dallas Museum of Art (DMA)**—In studying visitors' level of engagement with art, we identified four clusters of visitors to the Dallas Museum of Art. The DMA labeled them as follows: Curious Participants (32 percent of visitors), Committed Enthusiasts (26 percent), Tentative Observers (23 percent), and Discerning Independents (19 percent) (Randi Korn and Associates 2005).

Curious Participants are the largest cluster of visitors. Curious Participants are reasonably comfortable looking at art and want to connect with works of art in a variety of ways, including performances and readings. Visitors in this group have some difficulty with art terminology and are not particularly confident explaining it to others in spite of their reactions to art, which may be more emotional than cerebral.

Committed Enthusiasts are the second largest cluster of visitors. Committed Enthusiasts are confident, enthusiastic, highly knowledgeable, and emotionally connected to works of art. They are comfortable looking at art and talking about it. These visitors are sponges for knowledge about art and seek information of all types and formats.

Tentative Observers are the second smallest cluster of visitors. Tentative Observers are neither very knowledgeable about art, nor emotionally connected to art. They are uncomfortable talking with others about art, or explaining art to others. They are interested in obtaining straightforward, basic information about works of art.

Discerning Independents are the smallest cluster of visitors. Discerning Independents are confident, highly knowledgeable and emotionally connected to works of art. They are comfortable looking at art and talking about it. Discerning Independents want to develop their own interpretations of art and are less interested in others' explanations or views.

**Family Visitors to the San Francisco Museum of Modern Art**—In studying the level of engagement with art among adults visiting with children between 4 and 11 years, we identified three visitor clusters: Enthusiasts (50 percent of visitors), Art Lovers (27 percent), and Socials (23 percent) (Randi Korn and Associates 2009).

Enthusiasts are the largest cluster of visitors. Enthusiasts are fans of art and art museums. They place the highest value on "having an educational experience for their children," and they place the lowest value on "having a spiritual experience looking at art."

Art Lovers are the second largest cluster of visitors. Like Enthusiasts, Art Lovers enjoy art and art museums. However, Art Lovers value looking at art for their own visual pleasure, and they do not value making art with children.

Socials are the smallest cluster of visitors. Compared to the other two clusters, socials are less enthusiastic about art and art museums. Socials least value having spiritual experiences while looking at art and most value spending time with family and friends.

**Visitors to the Sports Legends Museum at Camden Yards**—In studying what kinds of sports fans are visitors to the Sports Legends Museum at Camden Yards, we identified four clusters: Middle-Road Fans (35 percent of visitors), TV Enthusiasts (32 percent), Active Enthusiasts (23 percent), and Indifferent Companions (11 percent) (Randi Korn and Associates 2008).

Middle-Road Fans comprise the largest cluster. While interested in sports, they are not emotional, die-hard fans. Middle-Road Fans are interested in watching sports to cheer the entire team's effort, and they prefer going to games rather than watching sports on TV. Middle-Road Fans are moderately attentive to the latest sporting news, and they admit being somewhat unhappy when their favorite teams lose.

TV Enthusiasts comprise the second largest cluster. They are highly engaged by sports and are team-connected and athlete-connected. Although TV Enthusiasts regularly attend sporting events, TV Enthusiasts are the only ones who respond favorably to the statement "I prefer to watch sports on TV." Like Active Enthusiasts, TV Enthusiasts feel that sports have great meaning in their lives and they go out of their way to learn the latest sporting news.

Active Enthusiasts comprise the second smallest cluster. They are highly committed to sports, have a powerful emotional connection to their favorite teams, and go out of their way to learn the latest sporting news. Of the four clusters, Active Enthusiasts have the strongest preference for watching sports by going to games and are least interested in watching sports on TV. They are also far more likely than are members of the other three clusters to regularly participate in sports.

Indifferent Companions comprise the smallest cluster of visitors. Indifferent Companions do not relate to sports and would not call themselves sports fans. Indifferent Companions have little interest in sporting news, they do not feel that sports has meaning in their lives, and they do not regularly participate in sports.

## Applications of Cluster Analysis Findings

In talking with museum staff and hearing their reactions, we have found that cluster analysis findings help museum staff come to an understanding of their visitors. The task is challenging, given that human diversity is so complex. Cluster analysis is useful since it allows for the nuances of visitors to emerge, yet it also groups similar visitors according to a specific research question.

To clarify, let's further consider the findings from the Dallas Museum of Art (DMA). In studying visitors' level of engagement with art at the DMA, we noticed that certain clusters of visitors responded similarly to particular statements. As seen in table 2, which displays the mean rating by each cluster, Curious Participants and Discerning Independents rated the statements "I enjoy talking with others about the art we are looking at" and "I like to know about the materials and techniques used by the artist" in similar ways. Discerning Independents and Committed Enthusiasts rated the statements "I feel comfortable looking at most types of art" and "Art affects me emotionally" similarly.

Table 2 also shows how the clusters are most different. There is great variation in

Table 2. Ratings of art viewing preferences by cluster (Randi Korn and Associates, Inc. 2005)

| **7-point rating scale: Does not describe me (1) Describes me very well (7)** | **Cluster** | | | |
|---|---|---|---|---|
| | **Tentative observers ($n$ = 256)** | **Curious participants ($n$ = 352)** | **Discerning independents ($n$ = 211)** | **Committed enthusiasts ($n$ = 284)** |
| | **mean** | **mean** | **mean** | **mean** |
| I like to view a work of art on my own, without explanations or interpretations.[1] | 3.9 | 5.2 | 5.9 | 4.3 |
| I am comfortable explaining the meaning of a work of art to a friend.[2] | 2.5 | 4.6 | 5.2 | 5.8 |
| Some terms used in art museums are difficult for me to understand.[3] | 3.4 | 5.2 | 2.3 | 1.8 |
| I enjoy talking with others about the art we are looking at.[4] | 4.0 | 5.8 | 5.7 | 6.4 |
| I like to know about the materials and techniques used by the artist.[5] | 4.3 | 5.6 | 5.6 | 6.1 |
| I feel comfortable looking at most types of art.[6] | 5.3 | 6.2 | 6.6 | 6.5 |
| Art affects me emotionally.[7] | 3.7 | 5.4 | 5.9 | 6.1 |
| I like to be told a straight-forward insight to help me know what the work of art is about.[8] | 5.4 | 5.8 | 2.6 | 6.0 |
| I like to know the story portrayed in a work of art.[9] | 5.5 | 6.2 | 4.7 | 6.4 |
| I like to connect with works of art through music, dance, dramatic performances, readings.[10] | 3.0 | 5.5 | 4.3 | 5.1 |

[1]$F = 55.549$, $p = .000$; [2]$F = 201.624$, $p = .000$; [3]$F = 375.627$, $p = .000$; [4]$F = 148.790$, $p = .000$; [5]$F = 70.538$, $p = .000$; [6]$F = 70.387$, $p = .000$; [7]$F = 141.133$, $p = .000$; [8]$F = 331.465$, $p = .000$; [9]$F = 100.361$, $p = .000$; [10]$F = 103.414$, $p = .000$.

the ratings for the statements "I am comfortable explaining the meaning of a work of art to a friend" and "I like to be told a straightforward insight to help me know what the work of art is about." Another distinguishing statement was: "I like to view art on my own without explanations." This statement was instrumental in differentiating the two clusters of visitors with a high level of interest in art: Committed Enthusiasts and Discerning Independents. These two clusters are both at the top level of engagement with art, but one group doesn't want to be told what to think, while the other group is a sponge for information.[8]

Understanding how clusters are most similar and dissimilar is extremely useful for marketing, programming, and even exhibition design. For instance, if a museum wanted to reach a broad range of visitors, it may look to find the similarities among clusters. For example, data indicate that the DMA might reach a broad range of visitors by offering a program on art technique, since three clusters—Curious Participants, Discerning Inde-

Table 3. Ratings of sports identity (Randi Korn and Associates, Inc. 2008).

| 7-point rating scale: Does not describe me (1) Describes me very well (7) | Cluster | | | |
|---|---|---|---|---|
| | Indifferent companions ($n$ = 32) | Middle-road fans ($n$ =106) | TV enthusiasts ($n$ = 95) | Active enthusiasts ($n$ = 69) |
| | mean | mean | mean | mean |
| I like learning about the history of my favorite teams.[1] | 3.4 | 5.6 | 6.4 | 6.3 |
| I watch sports to cheer the entire team's effort.[2] | 3.5 | 5.1 | 6.3 | 5.5 |
| Sports has great meaning in my life.[3] | 2.3 | 4.8 | 6.2 | 6.4 |
| I prefer to watch sports by going to games.[4] | 3.0 | 4.9 | 4.5 | 6.2 |
| I go out of my way to learn the latest sports news.[5] | 1.8 | 4.4 | 6.1 | 6.2 |
| I regularly attend sporting events.[6] | 3.6 | 4.1 | 6.2 | 5.6 |
| When my team is losing I usually feel bad.[7] | 2.4 | 4.5 | 5.4 | 6.3 |
| I watch sports to see the athletes I like.[8] | 3.9 | 4.2 | 5.7 | 4.5 |
| I regularly participate in sports.[9] | 2.4 | 3.1 | 4.7 | 6.2 |
| I prefer to watch sports on TV.[10] | 3.9 | 3.9 | 5.2 | 3.1 |

[1]$F = 67.767$, $p = .000$; [2]$F = 37.935$, $p = .000$; [3]$F = 132.993$, $p = .000$; [4]$F = 38.812$, $p = .000$; [5]$F = 127.375$, $p = .000$; [6]$F = 43.827$, $p = .000$; [7]$F = 73.904$, $p = .000$; [8]$F = 20.088$, $p = .000$; [9]$F = 66.671$; $p = .000$; [10]$F = 31.109$, $p = .000$

pendents, and Committed Enthusiasts— agreed relatively strongly with the statement "I like to know about the materials and techniques used by the artist." A museum might also want to know how the clusters are different, since differences indicate aspects that will prove least successful in reaching the broadest audience. For instance, the variation in responses to the statement "I like to be told a straightforward insight to help me know what the work of art is about" indicates that interpretation that only provides straightforward insights is not a successful strategy for reaching all clusters.

While making the museum accessible to everyone is important, museums may also consider focusing on one particular cluster in terms of marketing, programming, and event planning. Selecting one cluster to serve in a program, for example, is highly effective. Museum practitioners can better focus that program if they know the characteristics of a cluster. For instance, the Sports Legends Museum at Camden Yards may consider offering talks with baseball players to target TV Enthusiasts. We know that TV Enthusiasts "watch sports to see the athletes they like," "like learning about the history of their favorite teams," and stay up-to-date on the latest sports news (Randi Korn and Associates 2008). (See table 3).

Even when a museum chooses to focus its resources and programming on one cluster, such a decision does not preclude other clusters from attending and enjoying the experience, since visitors will not even know about these high-level decisions. Clusters can inform resource allocation. In this way, museums can use their resources to achieve the most impact by serving one segment of the audience very well.

## Conclusion

We find K-means cluster analysis to be a very useful analysis tool for museums that want to explore their visitors from a particular vantage point. We also use the concept of clustering to demonstrate that museum visitors are complicated and nuanced. When people walk into a museum, their dispositions do not change — they are who they are. Museum environments, though, can bring out different elements of visitors' inherent personalities and behaviors. We use K-means cluster analysis to explore museum-motivated ideas in order to provide an enhanced understanding of and insight into people.

## Acknowledgments

We would like to thank the Dallas Museum of Art, the San Francisco Museum of Modern Art, and the Sports Legends Museum at Camden Yards for giving us permission to share the results of the studies discussed in this article.

## Notes

1. Please note that our intention in conducting K-means cluster analysis is simply to understand a particular museum's visitors. We are not using findings to understand a larger or different population such as non-museum visitors.
2. All three studies are cited as unpublished manuscripts because they are the property of the museums that contracted the study. However, thanks to each museum, all three reports are available via the Internet. Audience Research: Levels of Engagement with Art[SM], A Two-Year Study, 2003-2005 is accessed at the Dallas Museum of Art Web site, www.dallasmuseumofart.org/AboutUs/LOEA. Audience Research: Exploring Family Visitors to Art Museums, for SFMOMA, and Audience Research: Sports Legends Museum at Camden Yards 2008 Visitor Survey are available at www.informalscience.org. Please refer to the reports if you desire further information regarding the methodology of each study, including sample sizes and response rates.
3. Other clustering methods include Hierarchical clustering and Two-step clustering. We use K-means clustering because it works efficiently with large data sets of interval or continuous data (SAS Institute 2004; SPSS, Inc. 2003).
4. Euclidean distance is the most commonly used distance metric; in fact, it is the

only distance metric option available in the K-means clustering algorithms in the SPSS and SAS statistical software packages (SAS Institute 2004; SPSS, Inc. 2003).

5. We cannot stress enough the importance of developing quality rating statements and scales that are pretested. In addition to examining statements and scales for clarity and content, we check the pretest results to see that the statements yield data with sufficient inter-item variability. If there is little variability in the data, you will not find substantial or meaningful distinctions among clusters. This is particularly important when conducting research on museum visitors—an audience that is relatively homogeneous.
6. The researcher may also specify the initial centroid values based on previous research. A word of caution: The solution reached by the K-means algorithm depends on the starting points, and different initial centroid values may lead to different cluster solutions. If there is any concern that the cluster solution reflects a local rather than global optimum, the researcher can run the K-means algorithm for a given number of clusters several times using different initial centroid values and then choose the solution with the smallest sum of the squared errors (the error is the distance from each case to the nearest centroid). Both SPSS and SAS programs will save the "distance to center" for each case in the data file. Ultimately, the validity of any cluster solution must be scrutinized by ascertaining whether or not the clusters differ from each other in distinct, meaningful ways that tell you something interesting and useful about the museum's visitors. This issue is discussed in more detail in the "Applications of Cluster Analysis Findings" section of the report.
7. Always check the iteration history on the program output to confirm that iterations converged with a stable solution. If there is no convergence, increase the maximum number of iterations and re-run the analysis.
8. This distinction is most interesting because the DMA's LOEA hypothesis states that there should be three levels of engagement with art, implying that there should be only three distinct clusters.

## References

Dubes, R. 1987. How many clusters are best?: An experiment. *Pattern Recognition* 20(6): 645–663.

Huang, Z. 1998. Extensions to the K-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery* 2: 283–304.

Kachigan, S. K. 1991. *Multivariate Statistical Analysis: A Conceptual Introduction*. New York: Radius Press.

Koehly, L., P. Arabie, E. Bradlow, W. Hutchinson. 2001. How do I choose the optimal number of clusters in cluster analysis? *The Journal of Consumer Psychology* 10(1/2): 102–104.

Randi Korn and Associates, Inc. 2009. Audience Research: Exploring Family Visitors to Art Museums. San Francisco, CA: San Francisco Museum of Modern Art. Accessed at www.informalscience.org.

Randi Korn and Associates, Inc. 2008. Audience Research: Sports Legends Museum at Camden Yards 2008 Visitor Survey. Baltimore, MD: Sports Legends Museum at Camden Yards. Accessed at www.informalscience.org.

Randi Korn and Associates, Inc. 2005. Audience Research: Levels of Engagement with Art[SM], A Two-Year Study, 2003–2005. Dallas, TX: Dallas Museum of Art. Accessed at www.dallasmuseumofart.org/AboutUs/LOEA.

SAS Institute. 2004. SAS/STAT 9.1 Users Guide. Cary, NC: SAS Publishing.

SPSS, Inc. 2003. SPSS for Windows 12.01 Computer software. The Apache Software Foundation.

Tan, P., M. Steinbach, and V. Kumar. 2006. *Introduction to Data Mining*. Boston: Pearson Addison-Wesley.