Dust & Magnet: multivariate information visualization using a magnet metaphor

Ji Soo Yi[1]
Rachel Melton[2]
John Stasko[2]
Julie A. Jacko[1]

[1]*Laboratory for Human–Computer Interaction and Health Care Informatics, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, U.S.A.;* [2]*College of Computing/GVU Center, Georgia Institute of Technology, Atlanta, GA, U.S.A.*

**Correspondence: Ji Soo Yi, Laboratory for Human–Computer Interaction and Health Care Informatics, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, U.S.A.
Tel: +1 404 385 2545;
Fax: +1 404 385 6115.
E-mail: jyi@isye.gatech.edu**

Abstract

The use of multivariate information visualization techniques is intrinsically difficult because the multidimensional nature of data cannot be effectively presented and understood on real-world displays, which have limited dimensionalities. However, the necessity to use these techniques in daily life is increasing as the amount and complexity of data grows explosively in the information age. Thus, multivariate information visualization techniques that are easier to understand and more accessible are needed for the general population. In order to meet this need, the present paper proposes Dust & Magnet, a multivariate information visualization technique using a magnet metaphor and various interactive techniques. The intuitive magnet metaphor and subsequent interactions facilitate the ease of learning this multivariate information visualization technique. A visualization tool such as Dust & Magnet has the potential to increase the acceptance of and utility for multivariate information by a broader population of users who are not necessarily knowledgeable about multivariate information visualization techniques.
Information Visualization advance online publication, 23 June 2005;
doi:10.1057/palgrave.ivs.9500099

Introduction

Most of the data from engineering, science, and business are multivariate, containing more than three attributes. To analyze such data, statistical methodologies can be used. However, such methodologies tend to summarize and compress the amount of information, so decisions made based on statistical results can be misleading due to the loss of the overall context of data. Instead, transforming the raw data into a more understandable visual pattern with less compression or summarization can yield a more reliable and effective overall analysis and interpretation of the data.[1] This visual representation of data is called 'information visualization.' Various information visualization techniques have been developed to support individuals who analyze data. Among them, multivariate information visualization techniques specifically deal with multivariate data.

In spite of their usefulness, multivariate information visualization techniques are not widespread. This is not unexpected because the major sources of multivariate data are specialized domain areas, which involve high levels of complexity, such as bioinformatics.[2–4] Likewise, multivariate information visualization techniques that deal with these specific data sets tend to operate under the assumption that the users have sufficient levels of training in analyzing data and using the tools. However, many individuals, who are not necessarily trained in these techniques, face common life decisions that involve some form of multivariate data (e.g., investments decisions, college selections, home purchases, and health care

choices). Furthermore, in the current information age, the availability of information and data relevant to these decisions has virtually exploded. Multivariate information visualization techniques offer powerful insights for the decision-making process. However, the general (a.k.a. untrained) population cannot be expected to invest the time and practice necessary to effectively leverage the tools to their potential. There is a clear, growing need to make multivariate information visualization techniques more usable and useful to the general population.

Based on this motivation, the authors propose Dust & Magnet (DnM) as a new multivariate data visualization technique. The main goal of DnM is to visualize multivariate data in an easy-to-learn and easy-to-use way, enabling the general public to deal with multivariate data effectively and efficiently with more satisfactory outcomes.

To deliver the results of the study, this paper is organized in the following way. The first section of this paper is a brief introduction of existing multivariate information visualization techniques, along with their intrinsic problems, which will serve as the foundation of our discussion. Some possible approaches to resolve these intrinsic problems follow as well. Second, a scenario of using DnM with a simple data set is presented. Because of the interactive nature of DnM, the authors believe that the scenario-based explanation is the most suitable way to introduce DnM. Third, the underlying interaction techniques, additional features, and technical details of DnM are presented. Fourth, the results of a user evaluation are presented and interpreted. Finally, a more in-depth discussion about the strengths and weaknesses of DnM is presented.

Background
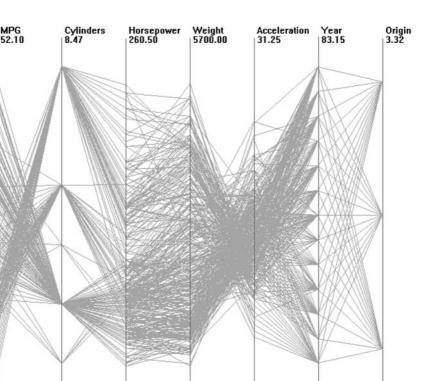
Multivariate information visualization techniques
A plethora of multivariate information visualization techniques have been invented.[5–11] Each has its own purposes and uses, so reviewing and comparing them on a feature-by-feature basis would be cumbersome. However, most of these techniques have the same underlying goal: to display complex, multidimensional information using a lower (e.g., two- or three-) dimensional space.

This common goal induces a problem called 'the curse of dimensionality reduction,'[12] which is called 'curse' because it is essentially an unavoidable problem in current multivariate information visualization. Depending on the visualization technique, dimensionality can be reduced in various ways. Some techniques compound several dimensions into a small number of dimensions (e.g., Principal Component Analysis[8] and Multidimensional Scaling (MDS)[13]). Other techniques do not preserve orthogonality between dimensions (e.g., Parallel Coordinates,[14] Star Coordinates,[11] Star Plot,[9] and Chernoff's Face[10]). And yet others display the relationship of a portion of dimensions (e.g., Scatter Plot Matrix,[8] Trellis plot,[5] and Worlds within Worlds[6]). These techniques

make it possible to portray multidimensional information on displays that can, pragmatically, only afford one-, two-, or three-dimensional illustrations of information. However, this dimensionality distortion confuses users because the distorted dimensions are not necessarily analogous to the properties of real-world phenomena. Training is therefore necessary for users to overcome this disconnection and to take advantage of these multivariate information technologies. Reviewing some exemplary multivariate information visualization techniques will provide a better understanding of this issue.

One of the most salient examples to demonstrate dimensionality distortion is MDS. MDS is a visualization technique employed to envision *n*-dimensional data on a two-dimensional space, while preserving the distance relationship of highly dimensional data points as much as possible. In other words, if two points are near to each other in the original *n*-dimensional space, MDS tries to preserve the proximity of the two points in the projected two-dimensional space. Thus, in MDS, an increase of dimensionality does not require more space, and multiple dimensions can be considered simultaneously. Even though MDS rather clearly shows clusters of data sets, it might be difficult to extract the meaning from the visualization because the original *n*-dimensional data are reduced to two dimensions. Without a proper level of training and background knowledge, it is generally hard to interpret the output of MDS.[13]

Another interesting technique is parallel coordinates. This technique contains a set of parallel lines, each representing a different variable. The values of the variables lie along these lines. Another set of lines, each representing a record of the data set, runs across the parallel lines crossing the parallel lines at the values of the variables for each record. Figure 1 is a visualization example of parallel coordinates using XmdvTool,[15] which is a public-domain tool for multivariate data visual exploration developed at Worcester Polytechnic Institute. The data set used for Figure 1 is a subset of a car data set from the Committee on Statistical Graphics of the American Statistical Association (ASA) Second Exposition of Statistical Graphics Technology.[16] As the figure shows, parallel coordinates is useful for detecting outliers and trends in a data set, although following individual cases is quite difficult, especially with large data sets. Having an interactive tool to assist with following the individual records or color coding particular cases aids with tracking specific data.[14,17,18] One might say that this technique has less dimensionality distortion because all the dimensions can be displayed without omission. However, this advantage is provided by removing orthogonality among dimensions, which makes presentation of data less intuitive.

Star coordinates is another exemplary technique that categorically falls between MDS and parallel coordinates. The star coordinates technique arranges multiple coordinate axes on a two-dimensional plane, made possible

Figure 1   Visualization of a car data set using parallel coordinates.

(as with parallel coordinates) by partially ignoring orthogonality (the term 'star' was derived from the radial shape of coordinate axes). Data points are plotted relative to these radial coordinate axes, so each point's location accounts for the entire set of multidimensional attributes. Ambiguities can arise when two points with different attributes are coincidently located at the same position.[11] Yet, the star coordinates technique can represent a greater number of dimensional relationships than the parallel coordinates technique, and with less dimensionality distortion than MDS.

The final example technique for review is scatter plot matrix. A scatter plot matrix is composed of a collection of simple, two-dimensional scatter plots. A scatter plot shows correlations between two variables by plotting data points on two orthogonal axes. A scatter plot matrix shows the correlations between pairs of variables among multiple variables by integrating multiple scatter plots. Figure 2, which uses the same data and tool as Figure 1 does, shows an example of a scatter plot matrix. As shown in Figure 2, each scatter plot only visualizes two variables, so there is no severe dimensionality distortion. Therefore, it can be argued that a scatter plot matrix is easy to understand and does not introduce any dimensionality distortion. However, a scatter plot matrix does not facilitate the extraction of relationships between more than two variables. For this reason, Hand *et al*.[8] call the scatter plot matrix approach a multiple 'bivariate'

information visualization technique instead of a multivariate information visualization technique.

From these examples, it can be understood that multivariate information visualization techniques inherently have a dimensionality distortion problem. However, the severity of dimensionality distortion is not the same among the techniques described. Instead, a trade-off between the severity of dimensionality distortion and the number of dimensional relationships is also identified. Thus, trying to resolve the dimensionality distortion problem completely might be futile. Instead, it might be wiser to try to strike a balance by relieving the complexity and difficulties that are involved in multivariate information visualization. This is the question addressed in this paper.

Metaphor

One possible approach to make visualization techniques easier is by using a metaphor. Dix *et al*.[19] suggest that, 'Metaphors are used quite successfully to teach new concepts in terms of ones which are already understood' (p. 148). Given that multivariate information visualization is a relatively new concept, finding a proper metaphor for the visualization and delivering this new concept to general users is of primary importance.

However, a metaphor is not a panacea. At the initial stage of adopting a new concept, a metaphor can work well to bridge between a traditional object and the new
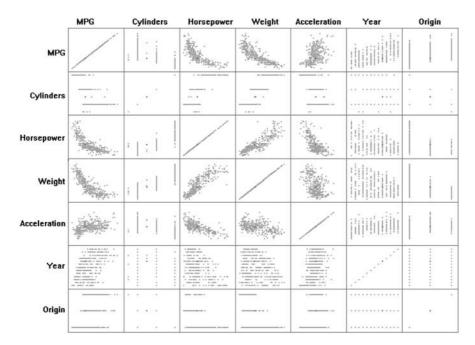
Figure 2   Visualization of a car data set using a scatter plot matrix.

concept, but it also can block exploration of new mechanisms and limit possible interfaces.[20] For example, the graphical user interface (GUI) metaphors of file, window, and desktop are considered passé by some researchers because these metaphors are bounded to real-world objects, so designers cannot leverage the metaphors to fully express all of the functionalities and interactions of a current user interface. Thus, using a metaphor based on a real-world object, which constrains possible interactions, is not appropriate. Therefore, a carefully chosen metaphor should be used.

Among several information visualization techniques using metaphors, one document visualization technique, named Visualization By Example (VIBE), was inspirational in the development of DnM. The VIBE system uses an interesting concept, called 'Point of Interest (POI),' to organize multiple documents. Each POI can represent keywords or sample documents. Documents are scored based on the similarity between the document and POI. The positions of documents are determined based on the scores. If a document is similar to a POI, the document has a higher score for the POI and is located closer to that POI than others with a lower score. If there are documents related to multiple POIs, the documents are located somewhere between those POIs. In other words, the distance between each document and each POI represents the similarity between the document and the POI. Olsen *et al.*[21] explained the metaphor of VIBE as organized stacks of documents in an office environment. As similar documents tend to be stacked at similar locations, similar documents in the VIBE system tend to be located at similar locations.

The concept of POI is evolved in WebVIBE, a descendant of VIBE.[22] Instead of relying on a document-stack metaphor, WebVIBE uses a magnet metaphor to represent POIs. As real-world magnets attract only ferrous (e.g., iron) particles among a mixture of iron and sand, the magnets on WebVIBE attract related documents to each magnet (POI). Because this phenomenon between magnets and ferrous particles is very familiar to the general population, this explanation of WebVIBE using a magnet metaphor is more intuitive than the explanation of VIBE using a POI and a document-stack metaphor.

The authors believe that a magnet metaphor can be used not only in document visualization, but also in multivariate information visualization. Specifically, applying a magnet metaphor in star coordinates is possible. As discussed, star coordinates uses multiple axes, in which each axis can be substituted with a magnet. As a data point is located along the axis according to its value, a data point can be located closer to or farther from a magnet according to its value. This substitution might change the way to understand the underlying logic of star coordinates. Originally, the concept of a vector summation of each data point's value on each dimension was used to explain the underlying logic of star coordinates.[11] Such an explanation can be too daunting for users who do not have a related mathematical background. However, using a magnet metaphor, the underlying logic of summation can be translated more easily. Each magnet represents a dimension (attribute or variable), and a data point with a high value in the given dimension is attracted to the magnet (axis end point) more strongly than a data point with a low value in the dimension. This

explanation leverages the familiarity of a magnet to deliver the complicated concept of multidimensionality.

Fortunately, a magnet metaphor avoids the common problem of metaphors discussed earlier. While a magnet is a real-world object, it does not necessarily have a strictly defined interaction with the exception of its unique physical characteristic – attraction. Thus, it is possible to say that a magnet metaphor is open to other possible interactions.

Animation and interaction

Many information visualization techniques emphasize the importance of an interactive user interface (e.g., Dynamic Queries,[23] Star Coordinates,[11] VIBE,[21] Worlds within Worlds,[6] and Table Lens[7]). Because interactive and dynamic visualization can provide more opportunities to convey various aspects of complex data compared with a static visualization, interaction with the data and the visualization tool is naturally emphasized in the domain of multivariate information visualization.

However, in many cases, interactive user interfaces come into play after the initial visualization of given data is finished. They do not show how the visualization or output is constructed. For example, both star coordinates and the VIBE system support an interactive user interface, but they only allow users to manipulate the data output after it is constructed. In other words, when a data set is imported to the software of Star Coordinates, the data points are instantly shown according to the values of data points and orientations of axes. Then, the user can change the directions and sizes of axes, which in turn changes the locations of the data points. The drawback of this approach is that users do not have a chance to see how the initial output is constructed, which may be useful in helping individuals better understand the nature of the data, itself, or even the functioning of the visualization tool or technique. The same thing happens in the VIBE system.

One more interesting point is that even though VIBE and WebVIBE utilize a magnet metaphor, they operate differently from how a real-world magnet does. As shown in the left side of Figure 3, a data point, which has a value of 2 for attribute 'A' and 1 for attribute 'B,' is located between the POI 'A' and POI 'B.' Even though the data point is twice as close to POI 'A' than POI 'B,' the data point stays between the two POIs. The physical phenom-
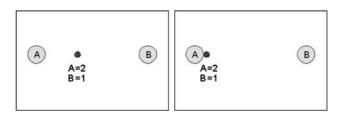
enon that real-world magnets show differs from what the left side of Figure 3 describes. If there is a slight difference in the magnitudes of attraction between two attractors, or magnets, the ferrous particle moves toward a strong attractor, or a stronger magnet. Thus, the right side of Figure 3 might depict a more realistic or intuitive conception.

However, the right side of Figure 3 has an important problem in this context: if every data point ends up being attracted and attached to one POI, how can a user see differences among data points?

Both of these problems can be resolved by animated visualization as Figure 4 depicts. There are two data points: one (red; the one on top) has a value of 3 for attribute 'A' and 1 for attribute 'B'; the other (blue; the one on bottom) has 2 for attribute 'A' and 1 for attribute 'B.' Instead of showing the stable and static visuals, animated visualization shows that the two data points are moving toward POI 'A' because both of the data points have bigger values for attribute 'A' than attribute 'B.' However, the speeds of the two data points are different. The red (top) data point approaches POI 'A' faster than the blue (bottom) data point. The animation of the two data points shows that they have different values for the attributes. This dynamic visual is more similar to the real-world phenomenon of magnetic attraction, so it can be more intuitively understood than the output of the VIBE system. Additionally, this can give users another chance to understand the underlying logic of this visualization.

An interactive user interface also can play an important role here. Both data points end up attached to POI 'A' in the end if the animation is not stoppable. If a user can stop the animation at the proper time, the user can preserve a screenshot that shows a distinction between the two data points. Additionally, if a user can move around a POI or magnet and observes how data points follow the POI or magnet, then the user can have more chances to build intuition about the data set. It is akin to playing with a magnet and seeing how iron particles are
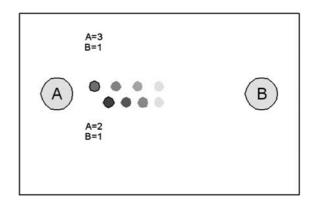

Figure 3   The attraction as is described for POIs of VIBE system (left) and a more realistic attraction (right).

Figure 4   The animated visualization of attraction. (Notice that the red (top) dot is moving toward A faster than the blue (bottom) dot.)

attracted to the magnet. Actually, this is how most individuals learn about magnets and their physical characteristics. This sparkling moment in our childhood should be more fully utilized.

Dust & Magnet

Overview
DnM is a new way to represent and manipulate multi-variate information. As the name implies, it vigorously uses a magnet metaphor. Most of us have had experience playing with magnets when we were children. A magnet can be used to separate ferrous particles from a mixture of ferrous and non-ferrous particles. The functionality of this interesting tool (and toy) is quite familiar, so to use the magnet metaphor in this case yields the possibility of leveraging its familiarity to enable users to understand new concepts of our multivariate information visualization technique, DnM. As a child plays with a mixture of particles and magnets, DnM lets a user play with data and magnets. Similarly, interesting information is often hidden within complex data, and DnM can help users pull out this interesting information selectively. Figure 5 shows a full screen shot of DnM.

DnM consists of three different views. In Figure 5, the largest view on the left side of the display (labeled 'Dust & Magnet') is the main view, on which most of the user interactions occur. The top right frame is the control view, which contains the options and parameters that users can adjust. The bottom right frame is the detail view, which contains detailed information about the selected data point(s). As shown in the picture, each data point (called a 'Dust' particle) in the main view is represented as a small black circle. Each variable (called a 'Magnet') in the main view is represented as a black square with a yellow label. The authors intentionally represent the Dust and Magnets with circles and squares, respectively, in order to make them distinguishable.

Scenario: choosing a cereal
Due to the interactive nature of DnM, it is difficult to exhaustively explain its features in a narrative way. Instead, an illustrative scenario will be introduced and some features of DnM will be explained within it. Consider the following scenario:

The data set for this scenario consists of a full list of 77 cereals each consisting of 12 attributes including brand
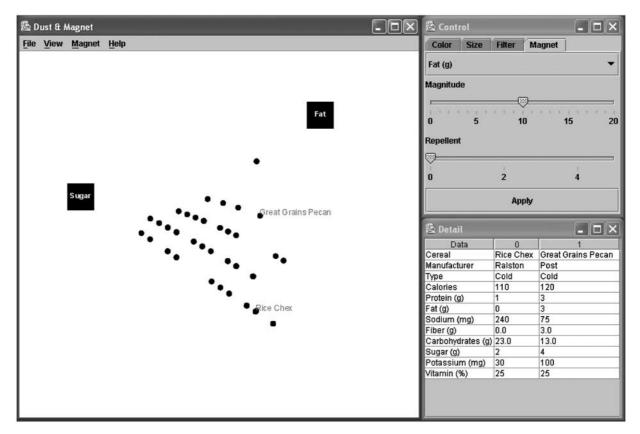

Figure 5 A full screenshot of DnM. On the left is the main window, called 'Dust & Magnet,' where most of the activity takes place; the top right window, called 'Control,' is where the user can adjust the values of various variables; and the bottom right window, called 'Detail,' is where the user can view the details of selected Dust particles.

name, manufacturer, type (cold or hot), calories, protein, fat, sodium, fiber, carbohydrates, sugars, potassium, and vitamins. Many people want their cereal to be low in fat and sugar, but high in protein and vitamins. Reviewing the contents of 77 cereals, however, is not an easy task. The following describes how to use DnM to aid in this complex multivariate decision-making task.

After importing the cereal data into DnM, all the pieces of Dust are located at the center of the main view and occlude each other. When the user selects the Magnet menu, all of the quantitative or ordinal attributes of the data are listed as possible Magnets. By clicking 'Magnet (Sugar (g))' under the Magnet menu, the user places the 'Sugar' Magnet on the main view. The user can then drag the Magnet using a mouse. As the Magnet is being dragged, the Dust particles with higher sugar levels are attracted toward the 'Sugar' Magnet faster than the others with lower sugar levels. In other words, based on the different sugar levels, the pieces of Dust start to separate. It is necessary to note that when the user drags a Magnet across the Dust particles, the particles do not stick to the Magnet, as true iron particles attach to a magnet. Dragging the 'Sugar' Magnet results in the representation shown in Figure 6.

In order to find out more about other factors, such as protein, fat, and vitamins, the user can select all of the related Magnets under the Magnet menu. To look for items that have a high level of protein and vitamins, the user can place the 'Protein' and 'Vitamin' Magnets toward the top of the main view. If the user also seeks low sugar and fat content, the 'Sugar' and 'Fat' Magnets
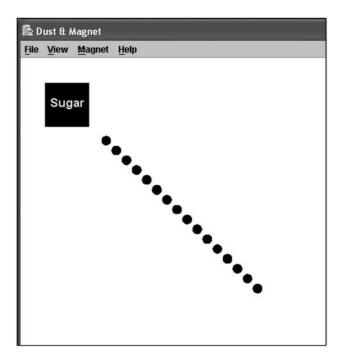
can be placed toward the bottom and then the user can drag one of the Magnets in order to attract the Dust. Even though only one of the Magnets is moved, all of the Magnets attract the Dust. In other words, all of the Dust particles are attracted toward all of the Magnets simultaneously according to the values of the attributes of the Dust. Some of the pieces of Dust are attracted to the bottom area. Others are attracted to the top area, and still others remain around the center area. The cereals near the top meet the user's hypothetical criteria of high protein/vitamin and low sugar/fat. In order to determine the brand names of the strong candidates, the user can click on the different pieces of Dust on the top area while holding down the 'CTRL' key on the keyboard. The selected Dust particles are then marked with brand names in red (see Figure 7).

From the visualization displayed in Figure 7, one is able to see that 'Special K' has relatively high levels of protein, while 'Product 19' and 'Total Whole Grain' possess high levels of vitamins. 'Total Whole Grain' is displayed slightly below 'Product 19.' 'Total Whole Grain' contains some fat in it, unlike 'Product 19' and 'Special-K,' which have none, as shown in the detail view in Figure 7. Using DnM, one can decide quickly which cereal to choose based on those attributes deemed important.

In addition, like many other information visualization techniques, DnM allows users to encode extra information by allowing the user to change the size or color of Dust particles. For example, if the user also decides to compare the calories, instead of doing the arrangement and attraction all over again, one can encode the amount of calories per serving to the size of the Dust as in Figure 8. The items with the fewest calories (50, in this case) have a diameter of 3 pixels, while those with the most calories (160, in this case) have a diameter of 20 pixels. Interestingly, the size encoding shows that 'Special K,' 'Product 19,' and 'Total Whole Grain' have average levels of calories per serving. Instead, 'All-Bran with Extra Fiber,' which is a newly emerged candidate, might be a better cereal choice for this user because it has a relatively small number of calories as represented by the size of the Dust.

In order to see which manufacturers make the cereals, the user can encode the manufacturer information by the color of the Dust particles. On the Color tab, the user can select 'Manufacturer' from the drop down menu, which then produces a list of the possible values. Default colors are automatically assigned for the list, so the user can simply click the Apply button. Or, in order to change any of the colors, the user clicks on the box for that variable and selects the desired color on a color palette that pops up. Figure 9 shows the result. The color-coding shows that Kellogg's (green: 'Product 19,' 'Special K,' and 'All-Bran with Extra Fiber') or General Mills (blue: 'Total Whole Grain') manufactures the strong candidates that we already identified.

Now suppose that the user believes that 'All-Bran with Extra Fiber' has an unappealing taste. DnM allows the

Figure 6 A screenshot of a possible result after dragging the Sugar Magnet on the main view.

user to change criteria easily. This time, the user persists with the belief that some sugar is necessary, but too much sugar is still unacceptable. In order to apply this criterion, the user moves the Sugar Magnet from the bottom of the main view to the top. Then, the user goes to the Magnet tab and selects Sugar from the drop-down menu where a slider labeled 'Magnitude' is located. Dragging the slider to the left can decrease the magnitude or influence of the 'Sugar' Magnet as shown in Figure 10. The size of the 'Sugar' Magnet also becomes smaller than that of the other Magnets as the magnitude of attraction becomes lower, which visually represents that the attraction of the 'Sugar' Magnet is weaker than those of the other Magnets. Thus, the user can identify easily which Magnets have a stronger pull and which have a weaker pull. The resulting view, in this case, shows that 'All-Bran with Extra Fiber' has moved elsewhere. 'Total Raisin Bran' is probably the best choice if both nutrition and taste are considered.

From this scenario, DnM demonstrates its potential effectiveness in representing multivariate information in
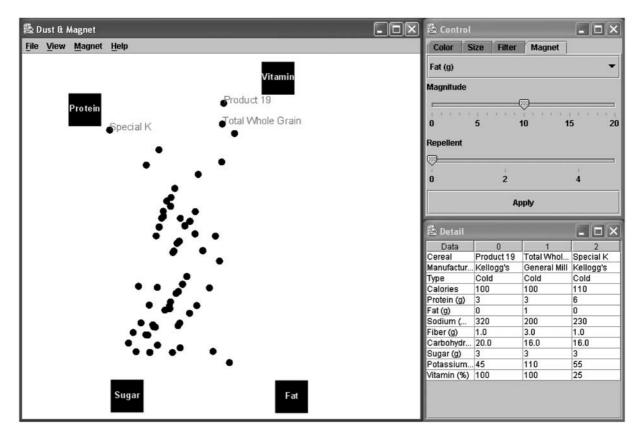
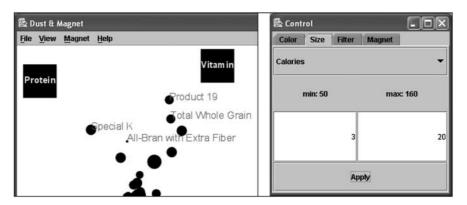**Figure 7**   A screenshot of the view after manipulating the four Magnets.

Figure 8   The resulting view after changing the size of Dust to reflect the calorie information.
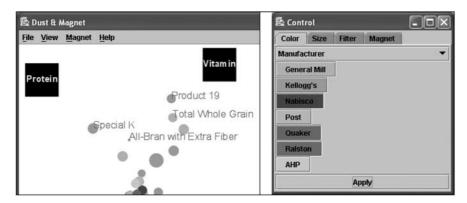
Figure 9   The resulting view after changing the color of the Dust to reflect the manufacturer information.
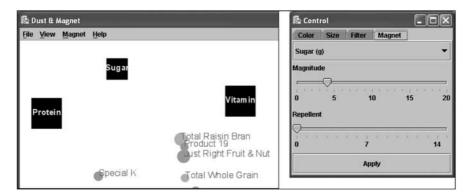

Figure 10   The resulting view after changing the magnitude of the 'Sugar' Magnet and moving it near the top of the view.

a user-friendly and exploratory manner. Users can sort through a data set using a single criterion or multiple criteria without losing the context of the whole data set. After sorting, detailed information can be reviewed by simply clicking on a particle of Dust and viewing its information in the detail view. In addition, the user can select multiple particles in order to compare the details. Encoding certain variables using the size and the color of the Dust is also effective to reveal additional information and trends. Each criterion can have a different weight (or magnitude, in magnet terms) by changing the size of the Magnet. Most importantly, the authors believe that the scenario demonstrates that those various tasks can be done easily and naturally using the intuitive metaphor of a magnet.

Primary interactions

From the scenario, some features of DnM are demonstrated briefly. In this section, two primary interactions ('attraction' and 'adjustment') of DnM are discussed in depth. The attraction between Dust and Magnets is the primary interaction of DnM. Whenever a user drags a Magnet, each Magnet attracts Dust. The more relevant a Dust particle, the faster the Dust particle is attracted to each Magnet. Because the levels of attraction are different according to various factors, Dust is clustered and sorted,

so the layout of Dust might become a more meaningful pattern. However, the attraction itself is not enough to make DnM beneficial. Like other information visualization techniques, occlusion among data points happens, so proper ways to avoid occlusion (e.g., 'Shake Dust') have been developed and will be discussed later in this paper. Additionally, DnM has a high degree of freedom, which makes DnM more interactive, although this also makes regenerating the same presentation difficult. Thus, ways to 'adjust' these problems compose the other primary interaction technique.

Attraction   In order to simplify the explanation of attraction, suppose only one Dust particle and one Magnet are located on the main view, and the magnet is dragged to generate attraction between the two. The magnitude of attraction is determined by the following four factors:

- the assigned attribute of the Magnet,
- the value of the matched attribute of the Dust particle,
- the magnitude (strength) of the Magnet,
- the repellent threshold of the Magnet.

The first two factors should be considered together. Each Magnet is assigned to an attribute. Among all of the attributes of the Dust, only those values of the matched

attribute of the Dust affect the attraction. For example, if the Magnet is assigned to attribute 'A,' only the value of attribute 'A' of a Dust particle affects the magnitude of attraction. Taking advantage of these factors, a user can attract Dust particles selectively.

However, the value of the matched attribute of the Dust particle cannot solely determine the level of attraction. Two more pieces of information should be considered: the type of the attribute and the range of values of the attribute. The range of values can vary according to the attribute. For example, suppose that human age can be between 0 and 120 years and human weight can be between 0 and 500 lb (about 226 kg). Thus, the numerical value of 100 in the attribute of human age differs in meaning from 100 in the attribute of human weight. The former corresponds to a relatively high value, corresponding to a high level of attraction. However, the latter is relatively low, so it should produce a low level of attraction. In order to normalize this difference of scales, min–max normalization is used. Min–max normalization maps the minimum value to 0 and the maximum value to 1. If the min–max normalization is applied in the example of human age and weight, the numerical value of 100 can be mapped to 0.83 and 0.20, respectively. The other information that should be considered is the type of the matched attribute, which is one of three attribute types: nominal, ordinal, and quantitative. Nominal data cannot be quantified, meaning that the level of attraction for the Magnet cannot be calculated. Thus, nominal attributes are not listed as Magnet options. However, nominal data are still useful to provide additional information about data, such as name, title, and description. The other two types of data, quantitative and ordinal, are quantifiable, making these types of attributes viable Magnet options. Strictly speaking, ordinal attributes should not be listed as Magnets as well because values of ordinal data do not carry precise numerical values. However, sequential integers are assigned to ordinal data based on the order of the values, and it turns out that the attraction based on these assigned sequential integers can be effective to show clusters and trends. Thus, depending on the type of the matched attribute, the value of each Dust particle has a different meaning. If it is nominal, it does not affect the magnitude of attraction at all. If it is quantitative or ordinal, it affects the magnitude of attraction, but the way to calculate the level of attraction in the quantitative type is slightly different from that of the ordinal type. More details about the algorithm used will be explained when the repellent feature is introduced.

The magnitude of a Magnet is another factor. Figure 10 shows how to change the magnitude of a Magnet. As shown in the figure, the magnitude of a Magnet can be adjusted using a slide bar, and the size of a Magnet in the main view also changes according to the slide bar, which makes the change of magnitude more visually obvious to users. The default magnitude of Magnets is 10, and it can be adjusted between 0 and 20. If the magnitude of a

Magnet is zero, it means the Magnet does not affect any Dust. However, the size of the Magnet never becomes zero, which would make the Magnet invisible. Instead, the Magnet becomes small and grays out to show that the Magnet does not have attraction.

A user can also adjust the repellent threshold. Using this factor, a Magnet not only attracts Dust, but can also repel Dust. Suppose the assigned attribute of a Magnet is quantitative. If the repellent threshold is zero (the default), there is only attraction and no repulsion. If the threshold is adjusted to a value (x), Dust particles whose value of the attribute is less than the repellent threshold ($<x$) are repelled and Dust particles with higher values than the threshold ($>x$) are attracted. For example, in the left screen shot of Figure 11, the repellent threshold is set to 2, so Dust particles that have a value of less than 2 for the Protein attribute would be repelled from the 'Protein' Magnet. In contrast, Dust particles that have a value greater than 2 in the 'Protein' attribute would be attracted to the 'Protein' Magnet. In the case that the matched attribute is ordinal, a user can select values to repel by selecting checkboxes. Dust particles that have the selected target value for the selected attribute (i.e., the attribute values with the checks) are repelled. For example, in the right screenshot of Figure 11, 'Quaker' and 'Ralston' are selected, so Dust particles that have 'Quaker' or 'Ralston' as their manufacturer attribute would be repelled from the 'Manufacturer' Magnet, but other Dust particles would be attracted to the Magnet. This factor is very useful to separate Dust particles effectively because the attraction and repulsion of Dust particles are totally opposite. Fortunately, the possibility of occlusion can also decrease.

Thus, the magnitude of attraction between a Dust particle and a Magnet can be described by the following equations.

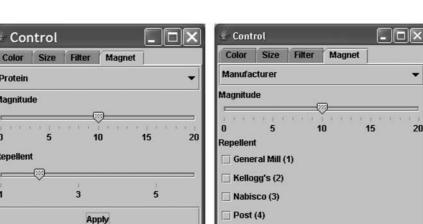In the case that the jth attribute is quantitative,

$$attraction(M_j, D_i) =$$
$$\frac{MM_j(DV_i^j - RT_j)}{\max(\{DV_k^j\}_{k=1,\ldots,d}) - \min(\{DV_k^j\}_{k=1,\ldots,d})}, \quad (1)$$

where M_j is the Magnet with the jth attribute, D_i the ith Dust particle, MM_j the magnitude of the jth Magnet $\in [0, 20]$, DV_i^j the jth attribute of the ith Dust particle, RT_j the repellent threshold of the jth variable, and d the total number of Dust particles.

In the case that the jth variable is ordinal,

$$attraction(M_j, D_i) =$$
$$\frac{MM_j(-1)^{R_j(DV_i^j)}[DV_i^j - \min(\{DV_k^j\}_{k=1,\ldots,d})]}{\max(\{DV_k^j\}_{k=1,\ldots,d}) - \min(\{DV_k^j\}_{k=1,\ldots,d})}, \quad (2)$$

where M_j is the Magnet with the jth attribute, D_i the ith Dust particle, MM_j the magnitude of the jth Magnet $\in [0, 20]$, DV_i^j the jth attribute of the ith Dust particle, $R_j(x) = 0$ if x is not a repellent value for the jth variable and $= 1$ if x is a repellent value for the jth variable, and d the total number of Dust particles.

Figure 11 Screenshots for how to apply the 'Repellent' feature: a slider for a quantitative attribute (left); checkboxes for ordinal attributes (right).

In summary, the magnitude of attraction is proportional to the magnitude of a Magnet and the matched value for a given piece of Dust. As previously discussed, the range of the matched value can vary according to the dimensions of the data, so min–max normalization is used to normalize any difference in scales. Also, if the variable is quantitative, a certain Magnet might have a repellent threshold, or, if the variable is ordinal, some values of variables are marked as repellent. In those cases, the value of attraction can be negative, which means that a Magnet can repel Dust instead of attracting it. This repulsion feature can be useful in order to diminish the amount of overlapping and to make clusters more distinctive. As mentioned previously, the magnitude of Magnets and the repellent threshold can be adjusted on the Magnet tab of the control view as shown in Figure 11.

A Dust particle is attracted toward or repelled from a Magnet while a user drags the Magnet. The speed of attraction/repulsion is determined by the magnitude of attraction. Because the Dust particle is attracted or repelled only while the Magnet is being dragged, a user can control how far the animation of Dust progresses. Additionally, it yields a comparable experience to encountering magnets for the first time in childhood. While a user drags a Magnet on the screen, Dust is pulled toward or pushed away from the Magnet. By observing this, the user can hopefully learn intuitively how DnM works. As shown in the scenario, initially all of the Dust is located in the center of the main view, so the user is forced to bring up a magnet and drag it around the Dust. The movements of Dust particles are different based on the values of each Dust particle for the given attribute. The speed and direction of each Dust particle give extra clues about the data set and the underlying logic of DnM.

If there are multiple Magnets on the main view, all of them attract/repel Dust simultaneously. Thus, the direction and magnitude of attraction is calculated by vector summation of each attraction between a Dust particle and every Magnet. This logic of vector summation is similar to that of Star Coordinates. However, users should be able to understand the underlying logic of DnM more easily because of the familiarity of the magnet metaphor. One might think that it is less intuitive to have all Magnets affect Dust simultaneously whenever one of the Magnets is dragged. However, this type of attraction was chosen because it was assumed to be more helpful to accomplish tasks involving multiple attributes than single attraction, where the Magnet being dragged would have solely affected Dust. If necessary, the single attraction can be simulated by adjusting the magnitudes of all Magnets to zero except for the desired Magnet, or by removing the other Magnets in the view.

Adjustment   As briefly mentioned before, the interaction technique of attraction alone is not sufficient to render DnM usable because two major problems occur: occlusion and a lack of reproducibility. DnM allows occlusion of Dust because it is natural to be located at the same or similar location if the attributes of the particles of Dust are the same or similar to each other, especially in the case of large data sets. However, it is difficult to see each piece of Dust, and trends of data sets can be misinterpreted when the Dust specks occlude each other. The other interaction problem is a lack of reproducibility. Because users can manipulate locations of Magnets freely, reproducing the same output from a data set is challenging. These two problems are unfavorable side effects of

the design principles of DnM, which permit high levels of interaction.

To avoid occlusion, Dust particles in the initial version of DnM were supposed to repel each other while they were moving. However, this repelling feature among the Dust introduced another problem. If Dust particles with low values surrounded a piece of Dust with a high value, the Dust particle with the high value could not escape from the surroundings because the adjacent particles of Dust blocked the internal ones. Consequently, the locations of the Dust failed to reflect the attraction among pieces of Dust and the Magnets correctly. There-fore, the 'Shake Dust' feature was implemented, instead. Whenever a user wants to see all of the Dust without occlusion, 'Shake Dust' can be used through the menu user interface or a shortcut key.

The primary concern in implementing this feature was that the automatic 'Shake Dust' feature might confuse users. It is possible to align Dust specks without occlusion instantly. If it happens, however, users can lose track of the Dust. Thus, smooth and gradual spreading-out is preferred.

The underlying idea of the 'Shake Dust' feature is simple. If a Dust particle occludes other Dust particles, it moves to avoid occlusion. Every Dust particle in turn dodges other Dust particles. In order to find the over-lapped Dust particles, DnM maintains an adjacency matrix internally that contains the distance between every two Dust particles. By looking up this adjacency matrix, DnM does not need to calculate the distance between every two Dust particles exhaustively. Whenever a Dust particle moves, DnM updates the adjacency matrix partially instead of thoroughly.

The pseudo-algorithm of 'Shake Dust' dictates that each piece of Dust is processed according to steps 1–3.

1. Update the adjacency matrix for a Dust particle (e.g., Dust particle 'A').
2. Find other Dust (e.g., any neighboring Dust particles) that occludes Dust 'A'.
3. If occlusion is found, do (a)–(c) for each neighboring Dust.

   (a) Calculate the relative location of a neighbor Dust to Dust 'A'.

   (b) Move Dust 'A' away from the neighbor Dust in one unit, which is corresponding to approxi-mately one pixel without zooming in/out.
   (c) Update the adjacency matrix for Dust 'A'.

As mentioned previously, the purpose of the 'Shake Dust' algorithm is not to align all the Dust without occlusion on the first try. Instead, the 'Shake Dust' feature gradually spreads out Dust as shown in Figure 12. Thus, each time the user clicks on the 'Shake Dust' menu item, DnM iterates through the above-mentioned steps once. The distance each particle moves is determined by how occluded it is by other particles, but the move-ment is normally small, allowing the user to track the movement. This gradual spreading-out might be too subtle, so using the shortcut key is easier because feeding multiple keystrokes is possible by holding down the shortcut key. This allows the user to choose the proper amount of spreading. Usually, when the grains of Dust are no longer moving, it is time to stop spreading.

Even with this help for removing occlusion, the lack of reproducibility still remains a problem. As it is an inherent result of the interactive, exploratory nature of DnM, the lack of reproduction seems difficult to resolve completely. Therefore, two additional features were implemented to help alleviate the problem. The first is the 'Center Dust' feature that returns all of the Dust to the original position, the center of the main view. This simple feature is useful when users want to clean up the view and run the attraction over again. If users leave the settings and orientation of the Magnets constant, then using the 'Center Dust' feature can allow users to repeatedly generate highly similar reproductions of the original (or previous) output. The second feature to alleviate the reproducibility problem is 'Attract Dust.' Instead of dragging a Magnet in the main view, the user can use the 'Attract Dust' command to attract the particles of Dust to all of the Magnets. Like 'Shake Dust,' each time 'Attract Dust' is used the Dust particles move a small distance: more precisely, two units at the most. This function is also accessible through a shortcut key, which can be more convenient for executing 'Attract Dust' multiple times. This method is less interactive, so users
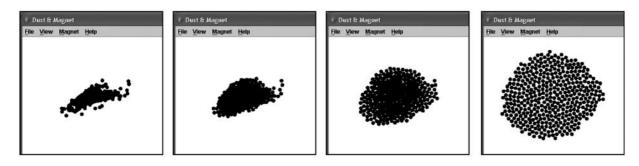

Figure 12 A series of screenshots demonstrating the results of the Shake Dust feature.

may miss some of the benefits of DnM. However, if the precise locations of the Magnets are important, the 'Attract Dust' feature helps to generate reproducible outputs from a data set.

Supporting features

Besides the two primary interactions, DnM has several supporting features, including well-known information visualization techniques such as zooming/panning, filtering, color/size encoding, detail view, and marking. These features are meant to enhance the usability of DnM and also to improve the overall effectiveness of the primary interaction techniques of DnM.

Zooming and panning DnM supports smooth zooming and panning features. The zooming-out feature is helpful when the number of Dust particles or Magnets overwhelms the allotted screen real estate. The zooming-in feature is helpful when a user wants to see a certain region more clearly, which can also serve as a simplified workaround to the occlusion of closely coupled data points. Consequently, panning is also necessary when a user wants to navigate the main display while the current view is smaller than the whole view (e.g., when zoomed in).

Filtering In some cases, certain Dust particles are not interesting and can even disrupt users in making sense of data sets, thus DnM provides a filtering capability. Depending on the nature of the dimensions, the filter has two appearances. If the variable is quantitative, a dynamic query function[23] is provided. Users can dynamically adjust the filter option while observing the changes on the main window instantly. If the variable is ordinal, the checkbox type filter is presented. Similarly, users can exclude or include certain values of the variable.

Color and size encoding Even though the most salient feature of DnM is tracking the locations of Dust particles, additional encoding on the Dust is helpful for understanding the trends and nature of the data. As shown earlier in the scenario, assigning different colors or sizes for the values of certain attributes might be useful at times.

Detail view and marking Although users are able to encode additional information besides the Dust particles' locations by using color and size, one might need to see the detailed information of several pieces of Dust. Especially, when the scope of candidates is sufficiently narrow, the attribute-by-attribute comparison can be very useful in making a solid decision or distinction.

In order to see the detailed information in the detail view, the users simply have to click on the Dust particle(s) of interest. For a comparison, multiple particles should be selected, which can be done by selecting multiple Dust particles while holding down the 'CTRL' key. Then,

selected Dust particles are labeled with the primary attribute (the name of the cereal in the case of the previous scenario) in red, and at the same time, the detail view shows the values of all of the attributes for the selected Dust particles. When another piece of Dust is selected without holding the 'CTRL' key, the previous selection is canceled out. Figure 7 is an example of these features. The marking of Dust can also be useful for tracking the movement of interesting Dust particles (i.e., data cases) while they are moving throughout the view.

Implementation

This application is written using Java2SE platform and based on Piccolo 1.0.[24] Piccolo is a Java toolkit that assists with creating zoomable user interfaces (ZUIs). Piccolo was helpful in the incorporation of the many features available in DnM: for example, the capability to zoom in/out as well as pan.

Evaluation

Participants

In order to assess the usability and effectiveness of DnM for information visualization tasks, a user evaluation was conducted. Six participants (mean age = 24.6) were recruited from the graduate student population at the Georgia Institute of Technology. Three students were recruited from the School of Industrial and Systems Engineering. These individuals represented experienced computer users who had no prior background in, or knowledge of, information visualization theories or techniques (the 'non-InfoVis group'). The remaining three participants were students recruited from a graduate-level information visualization class taught in the College of Computing. These three individuals represented the information visualization 'experts' with sufficient knowledge of the domain (the 'InfoVis group'). By involving the two different user groups, the authors sought to capture not only the perspective of novice users of information visualization systems, but also the perspective of experts who are familiar with the domain and other information visualization techniques. However, none of the six participants had been exposed to DnM previously. Each participant was compensated with 10 U.S. dollars for his or her voluntary participation.

Procedure

Following a background questionnaire, participants were given a comprehensive tutorial introducing the various features of DnM. Although participants were given detailed explanations of how each feature works and even encouraged to ask any questions during the session, the tutorial was designed to give participants minimal information about how to use DnM to accomplish specific tasks. As one of the key purposes of this user evaluation was to assess how easily individuals could learn to use DnM, the authors wanted to examine

whether users could accomplish various information visualization tasks without specific instruction on task performance using the tool.

After the tutorial, the experimental session started. The screen activities and participants' facial expressions/comments were video- and audio-taped for later review. The cereal data set, used in the scenario section of this paper, was presented to each participant. Five tasks were then given to each participant one by one. The five tasks were as follows:

- Task 1: Identify the cereal that contains the most sodium among Kellogg's cereals.
- Task 2: Is there a correlation between carbohydrates and sugar content?
- Task 3: Does Post make more cereals than other companies?
- Task 4: Please categorize cereals into two types: one consists of healthy cereals (more vitamins and fiber) and the other consists of unhealthy cereals (more sugar and calories). Which category has more cereals?
- Task 5: Write down the conditions for your ideal cereal and identify it using Dust & Magnet.

The first four tasks were chosen to exemplify different task types ('identify,' 'correlate,' 'compare,' and 'cluster') according to the common taxonomies of visualization techniques by Wehrend[25] and Roth and Mattis.[26] The last question (Task 5) was designed to give the participant a chance to use the features of DnM more freely.

After performing the tasks, participants were asked to complete a post-task questionnaire consisting of multiple-choice, Likert scale, and open-ended questions. The experiment concluded with an interview designed to elicit more personal, subjective views based upon the observations during the actual experiment. As a result, some of the questions asked in the interview varied from participant to participant.

Results

Task-by-task observation   The first task was to 'Identify the cereal that contains the most sodium among Kellogg's cereals.' All participants responded correctly without any major problems, and they rated this task as the easiest among the five tasks. Common usage patterns and strategies were as follows:

1. applying a filter to screen out other manufacturers except for 'Kellogg's,'
2. attracting with the 'Sodium' Magnet,
3. identifying the fastest Dust particle using the detail view.

The second task required a response to the question 'Is there a correlation between carbohydrates and sugar content?' According to a question in the post-task questionnaire, it was perceived as the most difficult out of the five tasks. All participants reported difficulty

interpreting the meaning of the distribution of Dust after performing attraction with the 'Sugar' and 'Carbohydrates' Magnets. Thus, only three participants responded in a manner we would consider correct, that is, that there is a (weak) negative correlation. This is probably partly because of the relatively weak negative correlation between sugar and carbohydrates (Pearson coefficient $= -0.332$). Another cause might be the complexity and dynamic nature of multiple Magnets. As this task shows, finding a correlation between two variables is not trivial using DnM. Actually, one of participants who succeeded in this task did not rely on the attraction feature, but relied on the color encoding feature to correctly derive the answer.

The third task required a response to the question, 'Does Post make more cereals than other companies?' Because DnM does not have an explicit feature that counts the number of Dust particles, the participants reported having a difficult time accomplishing a task of this nature. Eventually, all participants ended up using attraction with single Magnet with color encoding to sort the Dust particles in the order of manufacturer. They also used 'Shake Dust' to see the Dust particles without occlusion for counting Dust particles correctly. Five out of six participants answered the question correctly. However, the participant who answered incorrectly also employed the same usage pattern.

The fourth task was to answer the question, 'Please categorize cereals into two types: one consists of healthy cereals (more vitamins and fiber) and the other consists of unhealthy cereals (more sugar and calories). Which category has more cereals?' Interestingly, all participants accomplished this task using similar strategies (see Figure 13), even though the tutorial did not contain any specific instruction about how to accomplish this type of task. As shown in the figure, all participants put two Magnets corresponding to healthy variables (vitamin and fiber) on one end and the other two Magnets corresponding to unhealthy variables (sugar and calories) on the other end. Except for the subtle difference in the arrangement of Magnets, all of the participants accomplished the task in the same way, providing evidence that the basic interaction of DnM seems to support this kind of clustering task in an intuitive way.

All participants responded correctly and rated this task easier than Task 2, which appeared to be a simpler task because it involved only two variables. Thus, this shows that the number of magnets used in a task does not necessarily determine the task's complexity.

The fifth and last task was 'Write down the conditions for your ideal cereal and identify it using Dust & Magnet.' Figure 14 shows views from the displays of all participants while performing this task. Each participant set up his or her own criteria; some chose relatively complex criteria (e.g., involving eight variables as shown in the bottom middle screen shot of Figure 14). All participants successfully identified their own ideal cereal without
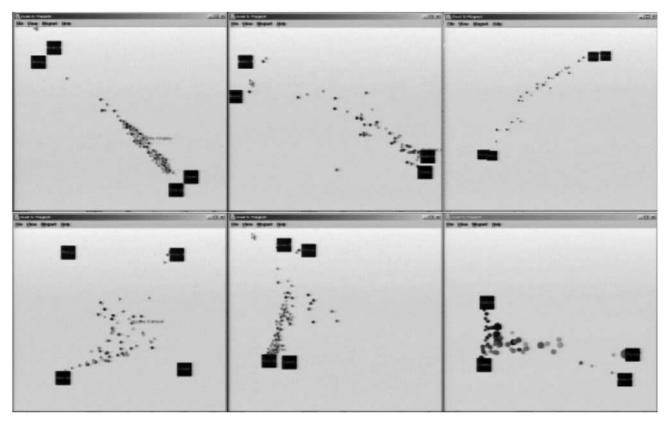
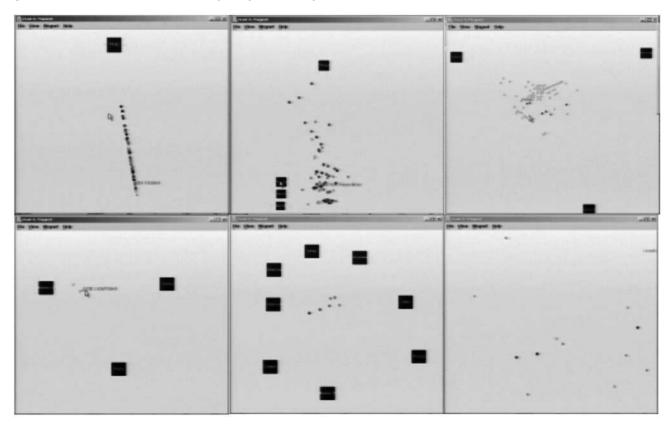**Figure 13** Screenshots of DnM when all participants accomplished Task 4.

Figure 14 Screenshots of DnM when all participants accomplished Task 5.

reporting any major problems. The general usage pattern was as follows:

1. attracting Dust particles with multiple Magnets,
2. narrowing down to a couple of candidates using the filter or color encoding features,
3. making the final decision after comparing the candidates using the detail view.

While accomplishing Task 5, participants demonstrated some advanced and creative usages of DnM. One participant stacked several Magnets (some were repelling, and some were attracting) on top of each other (see the top left screenshot in Figure 14). Another participant kept adjusting the magnitudes of Magnets to see the subtle differences between the final candidates (see the top middle screenshot in Figure 14). Another participant solely relied on the filter feature to trim out unwanted Dust particles (see the bottom right screenshot in Figure 14). All of these variations represent creative applications of the tool's features.

These varying usages during task performance demonstrated that DnM does not have limited usage patterns, which one might infer from the unanimous usage patterns demonstrated in accomplishing Task 4. Moreover, the results show that the magnet metaphor, which DnM relies on, does not suffer from the same limitations as the metaphors discussed in the Background section of this paper. Instead, the six participants who participated in this user evaluation seemed to understand the magnet metaphor (as demonstrated in the results of Tasks 1, 3, and 4) and took advantage of the interaction beyond the instruction (as demonstrated in the results of Task 5).

Generally, as described for each task, the participants used DnM effectively for accomplishing most of the tasks (finding a correlation was the most challenging). For other tasks involving multiple variables, DnM could be used easily and also creatively.

Overall impression   The overall impression that users had of DnM as a tool for information visualization was also of interest in this evaluation. Thus, the post-task questionnaire asked users to rate Dust & Magnet using a 5-level Likert scale.

Overall, the participants rated DnM between 'Good' and 'Great' (4.00–4.83, on a scale from 1 to 5 with 1 being 'Terrible' and 5 being 'Great') in terms of being 'easy-to-learn,' 'easy-to-use,' 'interesting-to-use,' and 'helpful.' Specifically, five out of six participants rated DnM as 'Great' for 'interesting-to-use.' For the 'easy-to-use' and 'helpful' criteria, DnM was rated slightly lower, although there were no negative assessments, such as 'Bad' or 'Terrible.'

Additionally, subjects were asked, 'Can you describe your overall impression of the software using one or two adjectives?' during the interview session. The following verbal responses were given, with V1, V2, and V3 referring to the three participants from the InfoVis group

and NV1, NV2, and NV3 referring to the three participants from the non-InfoVis group.

- NV1: 'Really cool. I want to have a copy.'
- NV2: 'It's potential. More potential to be great. I am really into information, so I might be a little bit biased subject.'
- NV3: 'It has a lot of potential. However, it's personal tool rather than professional tool. For example, SPSS.'
- V1: 'I like it. I like the metaphor.'
- V2: 'It's easy-to-use, so intuitive. Easy-to-pick up.'
- V3: 'I like the metaphor. Clever. Actually insightful.'

All responses were more positive and encouraging than expected. Because DnM is a new user interface for all of the participants, and because the given tasks forced them to deal with multivariate data sets, the tasks could have generated some frustration among users. However, that was not the case. As two participants mentioned, they liked the magnet metaphor used in DnM and thought that DnM was interesting, intuitive, and easy-to-use. These responses show that the major objective in this paper was accomplished at least for the six participants involved in the evaluation.

Discussion

As the results of the study suggest, one of the strengths of DnM might be that it is easy-to-learn and easy-to-use, especially for users without much knowledge of multivariate information visualization. As the participants' responses suggest, this strength most likely relies on the magnet metaphor and the interactive implementation of the magnetism. Due to the magnet metaphor, the logic of vector summation can be smoothly and implicitly delivered to the users. Compared to other multivariate information visualization techniques that do not use a metaphor (e.g., Star Coordinates), DnM can be more easily understood because the comprehension of complex mathematical algorithms is not necessary. The effectiveness of the magnet metaphor is also bolstered by the animated interaction of DnM. A magnet metaphor itself is already used in explaining the concept of POI of WebVIBE, but the authors believe that DnM differs from WebVIBE from the perspective of taking advantage of interaction to support the magnet metaphor. Animated interaction used in DnM piques the curiosity of users and provides clues about the underlying logic of the tool. Additionally, the interaction in DnM is more similar to the real physical phenomenon between a magnet and ferrous particles. Magnets in DnM attract Dust particles until the particles end up sticking to the Magnets. In contrast, data points in the WebVIBE system reside on the screen based on the proximity between POIs, or magnets, which is less realistic and might confuse users.

DnM can allow users to undertake sophisticated tasks, but in a simplified manner. As shown in the scenario and evaluation sections, adjustment of criteria can simply be performed by placing multiple Magnets and adjusting the magnitudes of the Magnets. Magnets can be positioned

on opposite sides of the Dust, in order to represent positive criteria at one end and negative criteria at the other. Assigning repellent thresholds is another way to adjust criteria. The sizes of Magnets also reflect the magnitudes or strengths of the Magnets, so users can get visual feedback as well. When comparing the DnM approach to accomplishing these kinds of tasks to the use of a spreadsheet application, it is easy to recognize how much DnM can improve the procedure of analyzing complex, multivariate data.

The supporting features, such as zooming/panning, filtering, color/size encoding, providing the detail view, and data marking, can also help the process of making sense of data. Even though these features are very common in modern information visualization tools, their use in DnM can greatly enhance its more unique interaction techniques as shown in some screen shots from the user evaluation.

Despite its benefits, DnM has several limitations. Above all, DnM has a lack of reproducibility. Because DnM allows users a high degree of freedom in adjusting and manipulating dimensions, it is challenging to explicitly document certain clustering schemes. Yet, if the user keeps applying the 'Center Dust' feature, it is possible to regenerate similar clustering. However, the dimensions of DnM can move all over the main view, so DnM does not have the reproducibility that other visualization techniques have. This interactive characteristic also introduces a lack of precision. Currently, the authors are working on a 'Saving' feature to save the arrangement of the main view, so this problem will be alleviated. Logging exhaustive movements of all the Magnets might generate copious amounts of data, but it can be a comprehensive solution for this problem. However, the problem of reproducibility is really a trade-off with the heightened ability for users to freely explore and manipulate the data with a high degree of freedom.

Occlusion was one of the biggest problems in the early stages of development. Now, several features are implemented to deal with this problem, including repellent options for Magnets and the 'Shake Dust' feature. Neither operation is perfect, although they help to remedy confusion resulting from occlusion.

Scalability, a common problem with other visualization tools and techniques, is another issue with DnM. Despite the enhancements to Java technology and Piccolo, there are still limitations in processing large data sets. Based on our test trials, conditions in which data sets with more than 700 data points are analyzed begin to slow down the interaction noticeably. However, this is primarily true for certain hardware profiles, including our test conditions: Dell Inspiron 8200 notebook with 2.2 GHz Mobile Pentium 4 and 256 MB of RAM running Microsoft Windows XP Professional, Service Pack 1. Additionally, the number of Magnets and Dust can be limited by the screen size of the main view or the interpretability of users. If too many Magnets are placed on the main view,

it may be difficult to explain the movement of the Dust clearly because too many factors are involved. However, the number of attributes inherent in the data set does not severely affect performance. As long as a given attribute Magnet is not placed on the screen, the algorithm is not bothered by the existence of this attribute or additional unused attributes.

The repulsion feature sometimes repels Dust particles outside of the current window of the main view, potentially causing users to lose track of the Dust particles. Putting a boundary around the main view was considered, but not implemented. Because a user might want to have a large space in order to make sense of his or her data, it is appropriate to give users full freedom to use the possible space. This problem only happens with Magnets with a repellent threshold and has not been observed during the user evaluation. If users apply this feature carefully, the problem is easily avoidable and when it happens, the 'Center Dust' feature allows the user to retrieve the missing particles.

Additionally, the participants of the user evaluation suggested several features that make DnM more usable. The most frequently requested feature was a reset feature, which would reset all settings, such as color and size encoding, filter, the magnitude and repellent threshold of Magnets, and zooming rate, to the original values. Another suggested feature was a summary statistics feature that would show descriptive statistics of selected Dust particles. The current version of DnM can show the detailed information of each Dust particle, but it does not have a feature to show summary information, such as count, mean, or standard deviation, of selected Dust particles. In addition, more sophisticated color encoding or filter features were requested. Fortunately, the suggested features do not conflict with the essential part of DnM, so the authors believe that these features can be incorporated into DnM relatively easily.

Finally, one might question whether DnM can be effective for people who do not know how to articulate a question. This is a fair question because DnM does not supply any systematic approaches to producing this query initially. Therefore, the user should pose the question in advance in order to find answers using DnM effectively. Conversely, as shown in the user evaluation, the fact that DnM is easy and interesting to use might encourage users to explore data sets more vigorously, which smoothly lead them to pose proper questions.

Conclusions

From the scenario and detailed introduction of interaction techniques, the effectiveness and ease-of-use of DnM is demonstrated. A user evaluation with six participants also partially verified its effectiveness. By introducing the metaphor of attraction, the complex multivariate information visualization seems to be explained intuitively and easily. Even though some dimensionality distortion

can happen in DnM, the interactive movement of Dust might be helpful for general users, who probably do not have a background in multivariate information visualization techniques, to understand the underlying logic.

Additional optimization of DnM is still necessary, along with consideration of any possible usability problems. The authors also hope that DnM will encourage the development and use of other simple, but elegant metaphors for information visualization tools.

References

1 Fayyad U, Piatetsky-Shapiro G, Smyth P. KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 1996; **39**: 27–34.
2 Saraiya P, North C, Duca K. Evaluation of microarray visualization tools for biological insight. *The IEEE Symposium on Information Visualization*. 2004 (Austin, TX, USA); 2004; 1–8.
3 Bolshakova N. Microarray software catalogue [WWW document] http://www.cs.tcd.ie/Nadia.Bolshakova/softwaretotal.html (accessed 27 October 2004).
4 Leung YF. Functional genomics [WWW document] http://genomics home.com (accessed 27 October 2004).
5 Cleveland WS. *Visualizing Data*. AT&T Bell Laboratories; Hobart Press: Murray Hill, NJ, Summit, NJ, 1993; 360pp.
6 Feiner SK, Beshers C. Worlds within worlds: metaphors for exploring n-dimensional virtual worlds. *The Third Annual ACM SIGGRAPH Symposium on User Interface Software and Technology* 1990 (Snowbird, UT, USA), ACM Press: New York, NY, 1990.
7 Pirolli P, Rao R. Table lens as a tool for making sense of data. *The Workshop on Advanced Visual Interfaces* 1996 (Gubbio, Italy), ACM Press: New York, NY, 1996; 67–80.
8 Hand D, Mannila H, Smyth P. *Principles of Data Mining*. MIT Press: Cambridge, MA, 2001.
9 Chambers JM. *Graphical Methods for Data Analysis*, Vol. xiv. The Wadsworth Statistics/Probability Series, Wadsworth International Group. Duxbury Press: Belmont, CA, Boston, 1983; 395pp.
10 Chernoff H. The use of faces to represent points in k-dimensional space graphically. *Journal of American Statistical Association* 1973; **v68**: 361–368.
11 Kandogan E. Visualizing multi-dimensional clusters, trends, and outliers using Star Coordinates. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001)* 2001 (San Francisco, CA, USA), Association for Computing Machinery, 2001; 107–116.
12 Bellman R. *Adaptive Control Processes*. Princeton University Press: Princeton, NJ, 1961.
13 Cox TF, Cox MAA. *Multidimensional Scaling*, 2nd edn. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC: Boca Raton, 2001; 308pp.
14 Inselberg A, Dimsdale B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. *Proceedings of the First IEEE Conference on Visualization. Visualization '90 (Cat. No. 90CH2914-0)*. (San Francisco, CA, USA), IEEE Computer Soc. Press: Silver Spring, MD, 1990; 361–378.

15 XmdvTool home page [WWW document] http://davis.wpi.edu/ ~xmdv (accessed 27 October 2004).
16 Donoho D, Ramos E. PRIMDATA: data sets for use with PRIM-H. *American Statistical Association (ASA) Second Exposition of Statistical Graphics Technology* 1983 (Toronto, Canada), 1983.
17 Artero AO, Oliveira MCFd, Levkowitz H. Uncovering clusters in crowded parallel coordinates visualizations. *The IEEE Symposium on Information Visualization* 2004 (Austin, TX, USA), 2004; 81–88.
18 Peng W, Ward MO, Rundensteiner EA. Clutter reduction in multi-dimensional data visualization using dimension reordering. *The IEEE Symposium on Information Visualization* 2004 (Austin, TX, USA), 2004; 89–96.
19 Dix A, Finlay J, Abowd G, Beale R. *Human–Computer Interaction*, 2nd edn. Prentice-Hall: Englewood Cliffs, NJ, 1997; 638pp.
20 Bederson BB, Hollan JD. Pad++: a zooming graphical interface for exploring alternate interface physics. *Proceedings of the ACM Symposium on User Interface Software and Technology* 1994 (Marina del Rey, CA, USA), ACM: New York, NY, USA, 1994; 17.
21 Olsen KA, Korfhage RR, Sochats KM, Spring MB, Williams JG. Visualization of a document collection: The VIBE system. *Information Processing and Management* 1993; **29**: 69.
22 Morse EL, Lewis M. Why information retrieval visualizations sometimes fail. *Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics. Part 2 (of 5)* 1997 (Orlando, FL, USA), IEEE: Piscataway, NJ, USA, 1997; 1680–1683.
23 Shneiderman B. Dynamic queries for visual information seeking. *IEEE Software* 1994; **11**: 70–77.
24 Bederson BB. A structured 2D graphics framework [WWW document] http://www.cs.umd.edu/hcil/piccolo/index.shtml (accessed 28 August 2004).
25 Wehrend SC. A categorization of scientific visualization techniques. M.S. thesis, Department of Computer Science, University of Colorado, Boulder, CO 1990.
26 Roth SF, Mattis J. Data characterization for intelligent graphics presentation. *The SIGCHI Conference on Human Factors in Computing Systems: Empowering People* 1990 (Seattle, WA, USA), ACM Press: New York, NY, USA, 1990; 193–200.
27 Wehrend S, Lewis C. A problem-oriented classification of visualization techniques. *The First Conference on Visualization* 1990 (San Francisco, CA, USA), IEEE: Piscataway, NJ, USA, 1990; 139–143.