Using Metaphors in Dynamic Social Stratification Visualizations

Michael B. Gousie

John Grady

Ben Burrage

Robbie Grossman

David Machado

See next page for additional authors

Authors

Michael B. Gousie, John Grady, Ben Burrage, Robbie Grossman, David Machado, Sarah Milewski, and Christopher S. Stuetzle

Using Metaphors in Dynamic Social Stratification Visualizations

7 authors, including:

Michael B. Gousie
Wheaton College (MA)
20 PUBLICATIONS   **176** CITATIONS

SEE PROFILE

John Grady
Wheaton College (MA)
26 PUBLICATIONS   **227** CITATIONS

SEE PROFILE

Diego Machado
Universidade de Fortaleza
39 PUBLICATIONS   **182** CITATIONS

SEE PROFILE

Christopher Stuetzle
Merrimack College
21 PUBLICATIONS   **34** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Computer Science Pedagogy View project

Multi-User Interaction on Large-Scale Displays View project

Using Metaphors in Dynamic Social Stratification Visualizations

Michael B. Gousie[1], John Grady[2]
Ben Burrage, Robby Grossman, David Machado, Sarah Milewski, Christopher Stuetzle
[1]Department of Math & Computer Science, [2]Department of Sociology
Wheaton College, Norton, MA 02766 USA
mgousie@wheatonma.edu, jgrady@wheatonma.edu

Abstract

We present three information visualizations for studying social stratification. Each Web-based applet uses a different metaphor to display U.S. Census income data along with the categories of race, marital status, and profession. Each system is completely dynamic, affording the user the choice of categorical variable to compare, as well as the choice of categories within each visualization. Two different user interfaces have also been implemented. The systems are described, compared, and their respective merits and deficiencies discussed.

Keywords— **visualization, metaphor, social stratification, multivariate data, applet**

1 Introduction

Social stratification, as it pertains to income, race, and other factors, is a major study area in sociology, political science, and other disciplines. As such, researchers often use data available through the Current Population Survey (CPS) of the U.S. Census Bureau [3]. While the Bureau also has various on-line tools for customizing census data, most of the data are presented as simple tables. Some have opted to look at multivariate social statistics by other means, such as the poster by Rose that uses icons and colors to represent various social/ethnic categories that are then plotted on an income graph [11]. This poster shows 11 different occupations, four race categories, and all combinations of families and dependents. However, because visualizations such as this are static, there is no way to make head-to-head comparisons of individual or custom combinations of categorical variables.

We present three different prototype Web-based systems for the visualization of social stratification. Each system uses a different metaphor that makes new patterns in the data appear compared to what can be seen in static tables. The use of metaphors is useful in information visualization because it allows users to more easily move from familiar concepts to new ones [5]. A metaphor can also supply a consistent framework for the user even as data within a visualization changes. The systems are dynamic, in that the user can highlight a certain categorical variable to compare as well as choose specific categories within a variable. Two of the systems allow the movement of data in the visualization to further increase the number or quality of comparisons that can be done. The systems use CPS data based on the 2000 U.S. Census, and display populations at various income levels as a simple numerical quantity or as a percentage of the income being displayed in the visualization in part or as a whole. Each system is meant to be effectively used by a non-expert in the area of (electronic) information visualization.

2 Previous Work

There are many information visualization systems in existence, such as IBM's Data Explorer, OpenDX, an open source version of Data Explorer [14], to name but two. However, these systems are for "everything and anything," and as such, do not include visualizations that may be appropriate for a specific data set. Another approach to get a desired visualization is to custom build one using a toolkit such as **prefuse** [7]. However, in either case, the user must have some expertise and time to develop a desired outcome.

Young *et al.*[17] describe some of the various graphical displays for use with statistical data, many of which are straightforward visualizations of typical statistical tools, such as scatterplots. Mosaic plots for frequency data are described; social stratification data would fit into this paradigm. ViSta, the Visual Statistics System is also described, a tool for technically proficient statisticians to create visualizations.

There are many custom-applications and Web sites for use in information visualization. Some that include elements relevant to our systems are described here.

A tool for analyzing historic hotel visits uses a table of square icons as part of its visualization system [15]; these dynamic icons can show patterns that may not be seen eas-
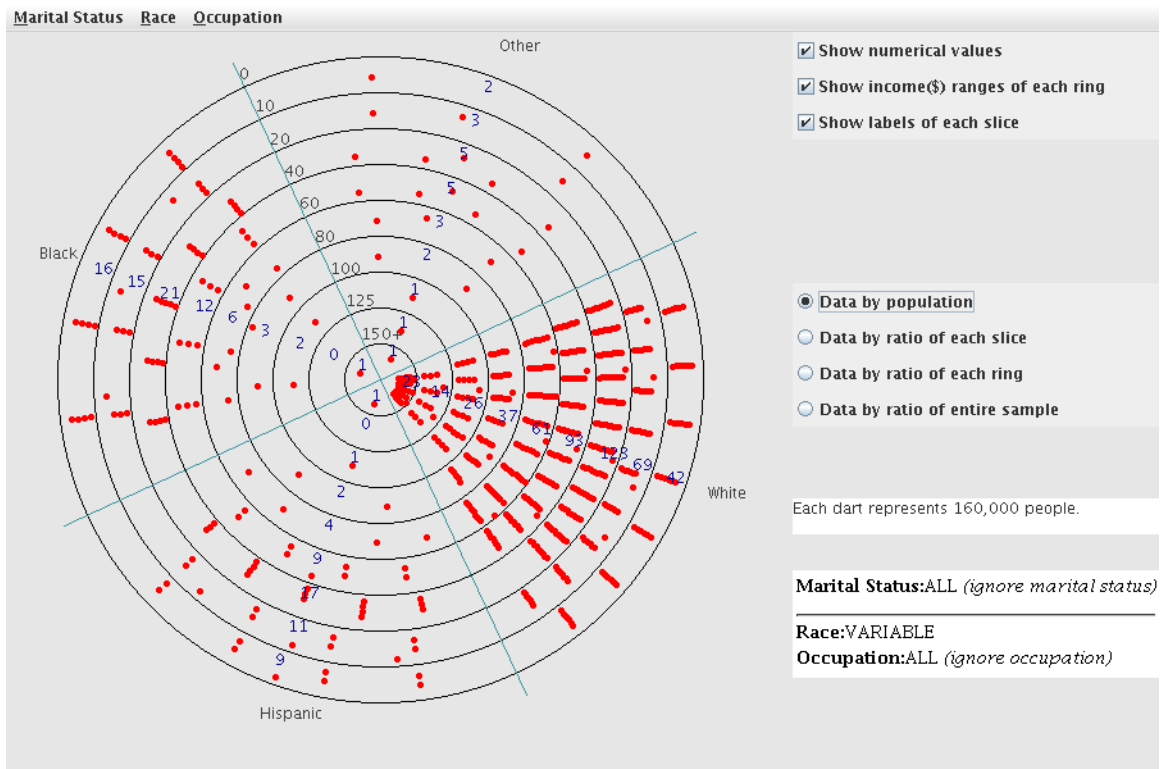
Figure 1: The Target showing income levels of the total population categorized by race/ethnicity.

ily through tables or other visualizations. In order to show more dimensions of data, PolarEyez [8] divides the visualization space in a radial fashion around a polar focus point; thus the same visualization works for different dimensionalities.

Social Explorer [1] is a Web site that includes several visualizations of census data, including income, ethnicity, etc. These values are layered, by color (darker=more/higher), on a map with zoom capabilities. However, only one kind of data can be accessed at a time and always in the context of geographical location, thus severely limiting the kinds of comparisons that can be performed.

Gapminder World [6] is a Web application written in Flash that uses animation in a two-dimensional table to plot multivariate data. Information about a country is displayed as a circle whose diameter is scaled relative to the magnitude of the data. One can view the data on the 2D graph or draped over a world map; the former makes most non-geographic comparisons easier. An additional dimension is displayed when the data circles are viewed in a time sequence, moving along the x and y axes as a Flash movie starting from the year 1995.

The New York Times has an interesting interactive Web

application that incorporates various visualizations depicting socio-economic conditions [13]. The portion that relates most to this work is the applet called "How Class Breaks Down," which displays a 3D bar graph visualization consisting of 25 rectangular parallelepipeds, where the height of each box represents the percentage of the population at the intersection of education (five different levels) and income (divided into five levels). Furthermore, the user can dynamically change the visualization to show just one row and column so as to compare high school equivalency and middle income, to give but one example. One can do the visualization based on all occupations or choose from 23 different categories. One can not change the x or y axes, however.

Diverse systems have used the idea of metaphors, often in 2D and sometimes in 3D [5]. Users more readily find desired information when the visualization maps to something well-known. For example, an iceberg metaphor is used to show that the visible data represents only a small fraction of that needed to compute a particular result in a financial context [9]. The idea of a map as metaphor is used regularly in many visualization systems, even when the data is not geographically based [12]. An interesting use of a metaphor is that of dust and a magnet [16]. In this

multivariate system, data points are represented as "dust" and attributes are represented as "magnets." Thus, the user may choose a particular attribute, and the matching data is "attracted" to the magnet. The size of the dust particles represents the magnitude of that particular component. Users report that the metaphor made the system easy to learn. In [10], metaphor-based icons represent multivariate data that are used as mosaic tiles layered on a map. Avatars that have facial expressions are part of a 3D virtual environment that uses social metaphors to enhance communication among participants [4]. The metaphor of a 3D city is used to show the relevance of search results in [2].
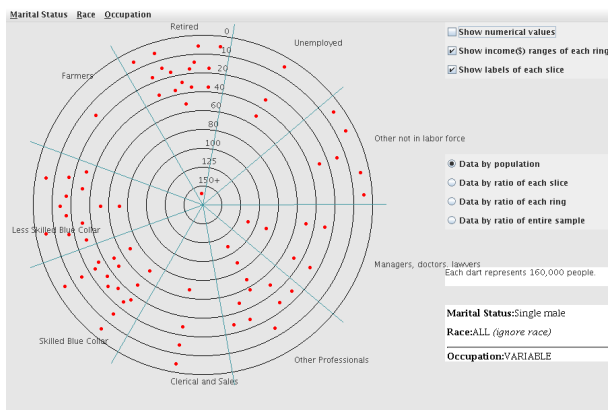

Figure 2: The Target showing percentages of income levels of single males categorized by occupation.

3 Three Metaphors

We have implemented three Java applets using three different metaphors to visualize social stratification, using data and categories taken from the U.S Census Bureau and the 2000 Census. Each applet displays the number or percentage of individuals at a particular income level. The individuals are chosen from three categorical variables: marital status (married husband, single male, single female, male head of household, and female head of household), race/ethnicity (white, black, Hispanic, other), and occupation (nine different possibilities, such as professional, skilled blue collar, etc.). Each of the visualizations let the user choose by which two categories the output will be indexed, as well as which values to use for the third category. The following sections discuss each of the three prototype systems: the Target, Mountain Climber, and Census[2], all available at http://cs.wheatonma.edu/mgousie/research.html.

3.1 The Target

The first applet is the Target, which uses the metaphor of a dart board. The goal of the individual is to hit the middle of the target, which represents the highest income level. Figure 1 shows the Target displaying the income levels by

population, where each red dot or "hit" represents 160,000 individuals. The highest income level is represented by the "bull's eye" at the center, with the progressively larger diameter concentric rings representing lower income levels. The target is further divided into four slices, one for each of the race/ethnicity categories. The other categories are turned on to "all," meaning all individuals with any marital status and any job type are included in the visualization.

The power of this applet (as well as the others) is that the categories can be changed dynamically. Figure 2 now shows the slices representing occupation, with the income figures taken from single males of any race/ethnicity. Furthermore, the data is displayed as a percentage of the whole, where each hit represents 1%; labels are turned off so as to keep the visualization uncluttered. New patterns now emerge where one can view clear differences among categories and income levels.

While the dynamic nature of the Target may improve social stratification investigations, there are a few shortcomings of the system. For one, the position of the slices is fixed, therefore eliminating the possibility of side-by-side comparisons between arbitrary groups. Also, the user interface is not optimal. For example, in informal tests, users often had difficulty in determining which variable would be assigned to the slices. In the pull-down menus, the option "Occupation: VARIABLE" indicates that occupations will be assigned to slices, whereas "Occupation: ALL" means the slices will be assigned to a different category, e.g. marital status. The current options are displayed in the lower right, but still presented difficulties to users. Finally, perhaps the worst deficiency is the clustering problem: Because the diameters of the rings necessarily get smaller the closer they are to the center, the "hits" are clustered closer together. The density of hits may give a false impression of the importance of a particular ring/slice. This can be seen in Figure 1, where it is not obvious as to how the different rings within the White category compare. The numerical values help in this regard, but somewhat obviate the idea of the visualization. A second version of this applet filled in areas in the rings using both the thickness of the filled areas and color gradations to represent the magnitude of the data rather than individual hits. Although it alleviated the clustering problem, this solution proved unsatisfactory because the magnitudes were more difficult to compare across categories compared to the user observing individual hits.

3.2 Mountain Climber

The second applet, Mountain Climber, uses the metaphor of a mountain. The highest income level is at the summit of the mountain, with equi-spaced horizontal lines denoting lower income levels as one moves toward the base. The visualization produces multiple mountains based on the number of categories to display (Figure 3).
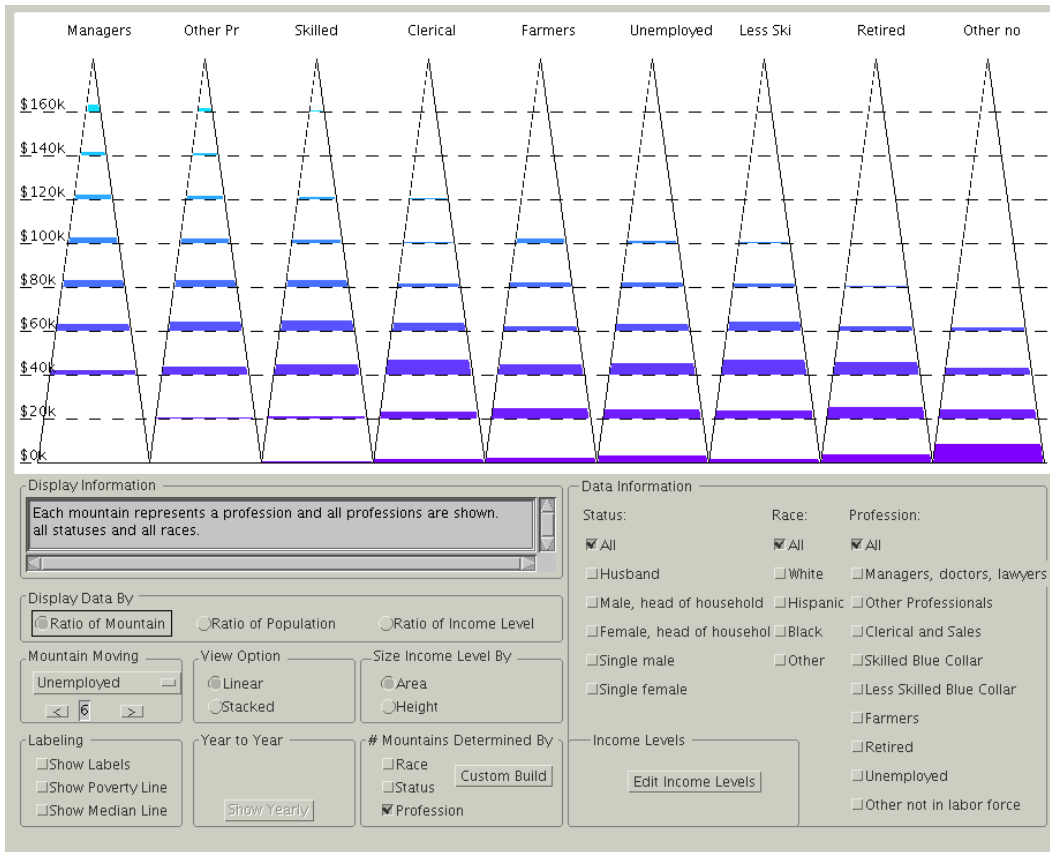
Figure 3: Mountain Climber showing income levels of all individuals, categorized by occupation.

Instead of "hits," the number of individuals within an income level is shown as a filled-in area, where the color and height/area represent the number of individuals in that category.
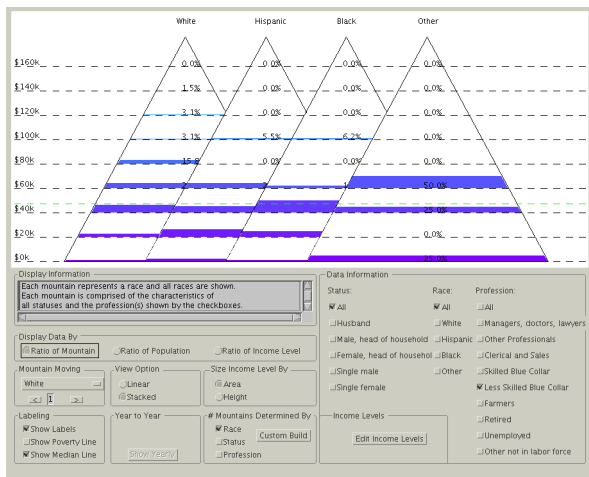

Figure 4: Stacked mountains comparing less skilled blue collar workers categorized by race/ethnicity.

There are several advantages of this visualization over the Target. One is that the user interface is improved by not using pull-down menus at all and instead showing all options at all times. In informal tests, this has been shown to be superior than the menus in the Target. Another feature is that the individual mountain positions can be moved; that is, the user may move the Skilled Blue Collar mountain next to Less Skilled Blue Collar to better compare the two categories or to see patterns among all the categories. The mountains may also be stacked to further aid comparisons. Figure 4 shows stacked mountains for all Less Skilled Blue Collar individuals categorized by Race. Note that percentages relative to each mountain are turned on in this figure; the percentages represent the ratio of the income relative to each mountain (shown), the entire population, or the income level. The visualization also adds a green line to indicate the median income.

Another feature of the Mountain Climber is the ability to create a custom mountain. The user can choose whatever

	$ 0-10	10-20	20-40	40-60	60-80	80-100	100-125	125-150	150+
Managers, Doctors, Lawyers	0	0	2	2	2	2	1	1	3
Other Professionals	0	0	5	5	4	3	2	1	2
Clerical and Sales	0	2	3	2	1	0	0	0	0
Skilled Blue Collar	0	2	8	8	6	4	2	1	1
Less Skilled Blue Collar	1	3	7	4	3	0	0	0	0
Farmers	0	0	0	0	0	0	0	0	0
Retired	7	11	11	5	2	0	1	0	0
Unemployed	0	0	0	1	0	0	0	0	0
Other Not in Labor Force	2	2	1	1	0	0	0	0	0

Selected Options

Marital Status: Show All

Race: White

Occupation: Sort By

Data By: Population

Each line represents 480,000 people.

Show Percentiles:

☐ Lower Percentile 0
☐ Upper Percentile 0

Data Over Time:

Start Year
End Year

Play   Pause

Select A Demographic To Save:

Feature is currently non–operational

Display All Saved Demographics

Data Highlighting Selections: ◉ None   ○ Below Poverty   ○ Below Median Income

Figure 5: Census[2] showing income for all white individuals, categorized by occupation.

mix of parameters (from the available categories) to create a completely custom visualization. For example, one could now compare married, white male professionals with single male professionals of any race/ethnicity. This feature affords comparisons not available in the Target.

Among the deficiencies of Mountain Climber is that the number of individuals can not be easily determined within a mountain without the numerical quantity labels turned on. The visualization also has somewhat the same problem as the Target in that the mountain gets smaller as the income level increases, akin to the clustering problem. Although the population with a certain income may be displayed via the height or the area of the colored area within each level, there may be instances where the user may get a false impression when comparing data at different income levels.

3.3   Census[2]

The final visualization is Census[2]. Although it has, perhaps, the weakest metaphor, that of a simple square box being filled, the simplicity gives this version much power. In the same way that hotel visits may be easily compared and patterns found by using rows and columns of icons of identical size [15], so too the boxes may be stacked in Census[2]

to afford easy comparisons between categories. Figure 5 shows all occupations for all White persons, categorized by occupation (same data as Figure 3). Each red line within a box represents a group of individuals (480,000 in this case) as shown in the user interface at the right. The number of bars is clearly displayed so the user need not resort to counting, and is less intrusive than the other systems. The user may also highlight those boxes below the median income or poverty levels. In this visualization, the income levels are on the x axis so that the number of categories can be modified easily in the vertical direction without crowding (which may happen in the other systems). The regularity of the boxes enables them to be stacked; indeed, a categorical row may be moved, by simple click-and-drag, to any position for comparison purposes. In informal tests, this system allowed users to see patterns more quickly than the other versions, and also did not have the problem of clustering which could lead one to incorrect conclusions about the data.

Note that in this visualization, filling a box only represents the highest value within a category, and not the highest income. However, the simplicity of the visualization is such that users have had no trouble deciphering the output.

There are some problems, however. The pull-down menus of the Target are reprised here, along with their deficiencies as discussed above. The information on the right side of the interface could also be improved. The "Data Over Time" animation and "Save" features have not yet been implemented, but would increase the system's usefulness.

Conclusions and Future Work

We have presented three different visualizations for studying social stratification. Each uses a metaphor to display census income data along with the categories of race, marital status, and profession. Each system is dynamic, affording the user the choice of multiple variables to compare. The data can be shown as simple numerical quantities or as a percentage of the whole, of the category, or of a particular income level. Mountain Climber and Census^2 allow the user to manipulate the visualization so as to achieve truly side-by-side comparisons. Mountain Climber has a user interface that most clearly shows the current status of the system, while Census^2 may have the most flexible data visualization.

Future work includes updating the systems to use more up-to-date CPS data, regularizing the user interface, and performing usability testing. The best system may be a combination of the visualization of Census^2 with an interface similar to that in the Mountain Climber. Furthermore, we hope to add more functionality, such as graphing capability using values chosen by the user and an animated visualization to see changes in social stratification over time.

References

[1] Andrew A. Beveridge, Ahmed Lacevic, Susan Weber, and Jordan Segall. Social explorer. http://www.socialexplorer.com/pub/maps/home.aspx; last accessed 5 February 2008.

[2] Nicolas Bonnel, Alexadre Cotarmanac'h, and Annie Morin. Meaning metaphor for visualizing search results. In *Proceedings of the Ninth International Conference on Information Visualization (IV'05)*, volume 4, pages 467–474. IEEE, 2005.

[3] U.S. Census Bureau. Current population survey (cps). http://www.census.gov/cps/; last accessed 5 February 2008.

[4] Steve DiPaola and David Collins. A social metphor-based 3d virtual environment. In *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, San Diego, CA, 2003.

[5] Cristina Russo dos Santos. Using 3d visual metaphors in information visualization. In *Proceedings of the 3rd IASTED International Conference on Visualization, Imaging, and Image Processing*, pages 890–895, Benalmadena, Spain, 2003.

[6] Gapminder Foundation. Gapminder world 2006. http://www.gapminder.org; last accessed 5 February 2008.

[7] Jeffrey Heer, Stuart K. Card, and James A. Landay. prefuse: A toolkit for interactive information visualization. In *CHI 2005*, pages 421–430, Portland, Oregon, USA, 2005.

[8] Sanjini Jayaraman and Chris North. A radial focus+context visualization for multi-dimensional functions. In *IEEE Visualization 2002*, pages 443–450, 2002.

[9] Miriam Lux. Visualization of financial information. In *Proceedings of the 1997 Workshop on New Paradigms in Information Visualization and Manipulation*, pages 58–61, Las Vegas, 1997.

[10] Thomas Nocke, Stefan Schlechtweg, and Heidrun Schumann. Icon-based visualization using mosaic metaphors. In *Ninth International Conference on Information Visualization*, pages 103–109, 2005.

[11] Stephen J. Rose. *Social Stratification in the United States*. The New Press, New York, 2000.

[12] A. Skupin. From metaphor to method: Cartographic perspectives on information visualization. In *IEEE Symposium on Information Visualization (InfoVis 2000)*, pages 91–97, Salt Lake City, UT, 2000.

[13] The New York Times. How class works. http://www.nytimes.com/packages/html/national/ 20050515_CLASS_GRAPHIC/index_03.html; last accessed 5 February 2008.

[14] Visualization and Imagery Solutions, Inc. Opendx. http://www.opendx.org; last accessed 12 February 2008.

[15] Chris Weaver, David Fyfe, Anthony Robinson, Deryck Holdsworth, Donna Peuquet, and Alan M. MacEachren. Visual exploration and analysis of historic hotel visits. *Information Visualization*, 6:89–103, 2007.

[16] Ji Soo Yi, Rachel Melton, John Stasko, and Julie A. Jacko. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4:239–256, 2005.

[17] Forrest W. Young, Pedro M. Valero-Mora, and Michael Friendly. *Visual Statistics: Seeing Data with Dynamic Interactive Graphics*. Wiley–Interscience, 2006.