The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information

Ramana Rao and Stuart K. Card

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
<rao,card>@parc.xerox.com

ABSTRACT

We present a new visualization, called the Table Lens, for visualizing and making sense of large tables. The visualization uses a focus+context (fisheye) technique that works effectively on tabular information because it allows display of crucial label information and multiple distal focal areas. In addition, a graphical mapping scheme for depicting table contents has been developed for the most widespread kind of tables, the cases-by-variables table. The Table Lens fuses symbolic and graphical representations into a single coherent view that can be fluidly adjusted by the user. This fusion and interactivity enables an extremely rich and natural style of direct manipulation exploratory data analysis.

KEYWORDS: Information Visualization, Exploratory Data Analysis, Graphical Representations, Focus+Context Technique, Fisheye Technique, Tables, Spreadsheets, Relational Tables.

INTRODUCTION

The size of information set which users can coherently bring together on the display of an interactive computer system limits the complexity of problems that can be addressed. In the last few years, we have been exploring the application of interactive graphics and animation technology to visualizing and making sense of larger information sets than would otherwise be practical by other means[3]. In this paper, we present a new visualization, the Table Lens, for manipulating large tables. Though elements of our design are broadly applicable, we have focused on the most widespread kind of table, variously called a cases-by-variable table, an object-attribute table, or loosely a relational table.

The Table Lens supports effective interaction with much larger tables than conventional spreadsheets do. A spreadsheet can display a maximum of 660 cells at once on a 19 inch display (at cell size of 100 by 15 pixels, 82dpi). The Table Lens can comfortably manage about 30 times as many cells and can display up to 100 times as many cells in support of many tasks. The scale advantage is obtained by using a so-called "focus+context" or "fisheye" technique. These techniques allow interaction with large information structures by dynamically distorting the spatial layout of the structure according to the varying interest levels of its parts. The design of the Table Lens technique has been guided by the particular properties and uses of tables.

A second contribution of our work is the merging of graphical representations directly into the process of table visualization and manipulation. Initially, graphical representations were incorporated because of their natural economy in showing cell values. However, a second, perhaps more important, advantage is the effectiveness with which humans are able to spot patterns and features in well-designed graphical renderings of collections of values. The combination of our focus+context technique and graphical mapping scheme, with a small set of interactive operators enables performing exploratory data analysis in a highly interactive and natural manner. After describing the focus+context technique, the graphical mapping scheme, and the Table Lens user interface, we illustrate this data exploration process with actual analysis scenarios.

TABLE LENS FOCUS+CONTEXT TECHNIQUE

Focus+Context techniques support visualizing an entire information structure at once as well as zooming in on specific items. This interplay between focus and context supports searching for patterns in the big picture and fluidly investigating interesting details without losing framing context. A number of such techniques have been developed in the last ten years including the Bifocal Display [7], Furnas's Fisheye techniques [1], the Perspective Wall [2], techniques for graphs and for direct manipulation of 2-D surfaces by Sarkar et al. [5, 6], and the Document Lens [4].

The Table Lens technique has been motivated by the particular nature of tables. The most salient feature of a table is the regularity of its content: information along rows or columns is interrelated, and can be interpreted on some reading as a coherent set, e.g. members of a group or attributes of an object. This is reflected in the fact that tables usually have labels at row and column edges that identify some portion of the meaning of the items in the row or column. These observations indicated a need to preserve the coherence of rows and columns and their labels despite distortions to the table. Thus, the Table Lens mutates the layout of a table

Reprinted From: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, April 1994, ACM

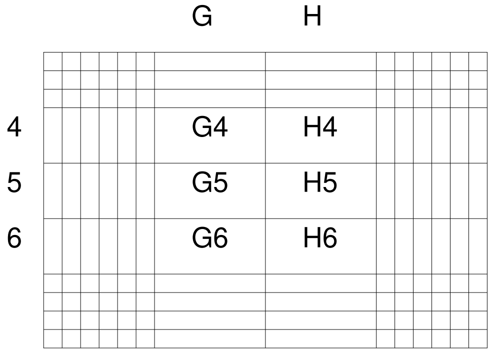

Figure 1: The Table Lens Focal Technique.

without bending any rows or columns. Cells in the focal area and the label row and column divide the total focus space of each dimension appropriately. Cells in the context divide the remaining space equally. Figure 1 shows a 10 by 14 table with a focus area of 3 by 2 cells.

The Table Lens technique is similar to that of Sarkar et al.[6]. Besides the difference in metaphor, the Table Lens distorts based on discrete cell boundaries as opposed to smaller pixel or continuous units. It is unlikely that providing support for partial inclusion of a cell in the focal area would be useful, but even then, a cell "detente" or "gravity" mechanism that aids "clicking" or "snapping" into cell borders is necessary. Our technique is further complicated by variable cell widths and heights. In particular, as the focus is moved around the table, it may have to change size to remain on cell boundaries.

An important property of the Table Lens technique is that distortion in each of the two dimensions is independent from the other. This means that rows and columns aren't bent by the distortion, and can thus be scanned entirely by a single horizontal or vertical eye motion. Furthermore, this enables label display, multiple focal areas, and multiple focal levels. Multiple focus areas are important for a number of reasons including comparing distal areas of the table and maintaining focus on summary rows or columns while investigating other portions of the table. Multiple focal levels allows dealing with larger tables and opens up a promising new design space (our current implementation has started this exploration by adding a third level). Below, for concreteness, we illustrate the case of two levels (i.e. focus and context), but point out issues arising with multiple focal levels.

As can be seen in Figure 1, though cells are allocated spaces along each dimension independently, there is an interaction in cell geometry. In fact, four types of cell regions are created by the distortions on the two axis: focal, row focal, column focal, and nonfocal. *Focal* cells are in the focus area along both axes, *row focal* and *column focal* are both half focal in that they are in the focal area of only one of the two axes, and *nonfocal* are in the context area along both axes. For cases with multiple focal levels, there are n by m types of areas, where n and m are the number of focal levels in each dimension (e.g. nine area types in our current implementation). As later described, each of the cell region types may require separate graphical treatment.

Distortion Function Framework

The distortions produced by many focus+context techniques can be described using a general framework starting from the notion of a degree of interest (DOI) function as introduced by Furnas [1]. A DOI function maps from an item to a value that indicates the level of interest in the item. The DOI function can be used to control how available space is allocated amongst items. The DOI function changes over time because of user interaction (e.g. the focus is moved) or system activity (e.g. search processes).

In the Table Lens, a DOI function maps from a cell address to an interest level, and each of the two dimensions has an independent DOI function. In particular, with one focal area, each dimension has a block pulse DOI (as in Figure 2, in which the contiguous focus area cells are at a higher level of interest than the context area cells. Multiple focal areas are characterized by a DOI function which contains multiple pulses. Multiple levels of focus are characterized by a DOI function with block pulses at different levels.

An additional framework concept is that of a transfer function that maps from uniformly distributed cell addresses to "interest-warped" physical locations. In fact, such a transfer function is the integral of the DOI function scaled appropriatedly to fit in the available space. Thus it essentially maps from an item to the total area that has been allocated to items of lesser cell address.

Figure 2 shows a DOI and an associated transfer function for one dimension and the effect of this distortion. The DOI function shown at the top of the figure is a square pulse that comes in three pieces at two different levels. The transfer function has three pieces at two different slopes, which map cells to the three areas of warping, the central focus area and the two flanking context pieces.

This framework allows contrasting Table Lens to other distortion techniques. The Table Lens distortion is equivalent to the distortion function of the Bifocal Display, except that it uses two independent distortions for each of the dimensions. The DOI/transfer functions of the Perspective Wall and the Document Lens are somewhat more complicated. In both cases, the interest level for context cells falls off with distance from the focus area, so the DOI is a flat-topped mountain with sloped regions falling off from the top. Also in both cases, the dimensions interact so the DOI function can not be decomposed into independent functions for each of the dimensions. Thus, their DOI/transfer curves are actually z-surfaces over an x-y plane.

Interactive Manipulation of Focus

The Table Lens supports a number of operations for controlling the focal area. These operations are analogous to ones on the Document Lens and Perspective Wall, though these techniques didn't originally support all the operations. In particular, there are three canonical manipulation operations:

- *zoom* changes the amount of space allocated to the focal area without changing the number of cells contained in the focal area. It corresponds to pulling the lens closer to the viewpoint in the Document Lens and to increasing the width or height of the focus area of the Table Lens.

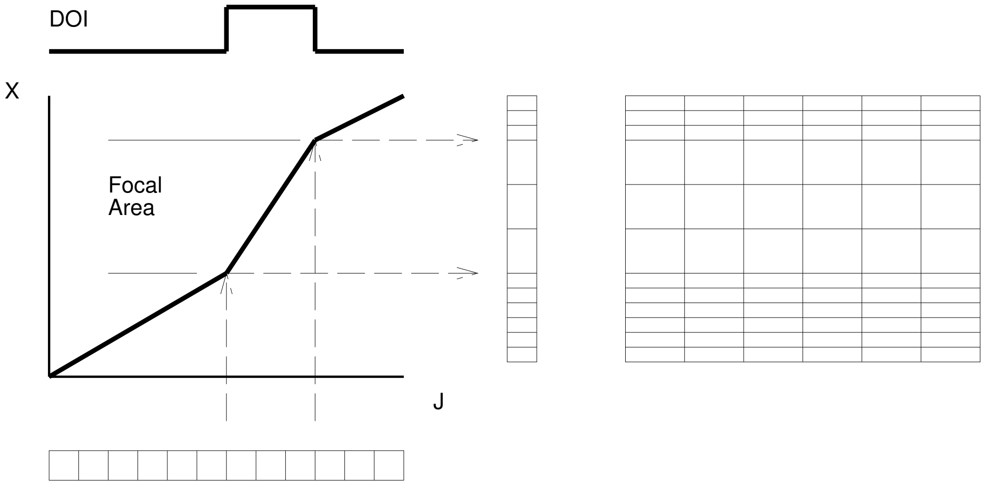

Figure 2: The DOI function maps from cell address to interest level, and the Transfer function maps from cell address to position along an axis. The Transfer function is the integral of the DOI function scaled to fit in the available space on the axis. Its effect can be visualized by the mapping of the cell array into a distorted cell array.

• *adjust* changes the amount of contents viewed within the focus area without changing the size of the focus area. It corresponds to stretching or shrinking the Perspective Wall, or pulling more or less of the table into the focus area.
• *slide* changes the location of the focus area within the context. It corresponds to sliding the Document Lens or Table Lens in its 2-d plane, or to panning to a different region on the Perspective Wall.

Each of these three operations can be understood visually as simple effects on the DOI and transfer functions as is illustrated in Figure 3. Zoom increases the slope of the focal area which also decreases the slope of the context pieces, since there is a fixed amount of available space. Adjust increases the number of cells in the focus area without changing the amount of space occupied by the focus area, thus the slope in the focal area must decrease. Finally, slide pushes the high slope area to a different spot on the axis.

Another important operation (motivated by actual use) is a coordinated adjust and zoom. A common need is to increase/decrease the number of cells in the focus without affecting their size. This requires doing an adjust plus enough of a zoom to add enough space to the focus to preserve the original cell sizes, an operation we call *adjust-zoom*.

If multiple focal areas are supported, then the various operations can be done on individual focal areas or on all of them as a set. If a single level of interest is desired then zoom must be applied across all the focal areas simultaneously, adjust behaves poorly and should be disallowed, and adjust-zoom can be performed on each of the focal areas independently. Multiple focal levels opens up a complex design space that needs further exploration.

GRAPHICAL MAPPING SCHEME

The Table Lens currently uses a graphical mapping scheme that is tailored for the most common type of table: the cases-by-variable array. In particular, this means that the underlying table represents a number of cases (the rows) for each of which values of various variables (the columns) are provided. For example, we use a table of baseball players performance/classification statistics for 1986 below.[1] In particular, this table contains 323 players by 23 variables, 17 quantitative (e.g. At bats, Hits, Home Runs, Salary '87) and 6 category (e.g. Team, Offensive Position, Team '87).

The Table Lens uses a number of different types of graphical representations, called presentation types, to display the contents of the table cells. In particular, presentation types utilize graphical vocabulary including text, color, shading, length, and position to represent underlying cell values. Six factors affect which presentation type and how in particular the type is used (examples given can be readily seen in the five color plates):

• *Value*. The cell value is depicted in some way by a presentation type. In a text representation, the cell's value is printed as a number or a string. In a bar representation, a quantity can be represented by the length of the bar.
• *Value Type*. The cell's value type determines which presentation type is used. In particular, a presentation type is chosen for each column of a cases-by-variables table. Quantitative variables are presented using a bar representation and category variables are presented using shaded, colored and/or positioned swatch representations.
• *Region Type*. Cells in focal, column focal, row focal, or non-focal region are treated differently. In particular, a focal cell uses a textual presentation as well as a graphical presentation so that the focal area is integrated into the graphical presentation of the context. A column that is focal uses a more colorful or detailed presentation type than nonfocal columns. The more focal the region is, the brighter the background shade (e.g. we currently use 3 gray levels).
• *Cell Size*. The cell size depends on the region type, but also depends on the amount of available space, and the size of the

[1]The data, obtained from the CMU StatLib server, was collected by the American Statistical Association from Sports Illustrated and the 1987 Baseball Encyclopedia Update, Collier Books.

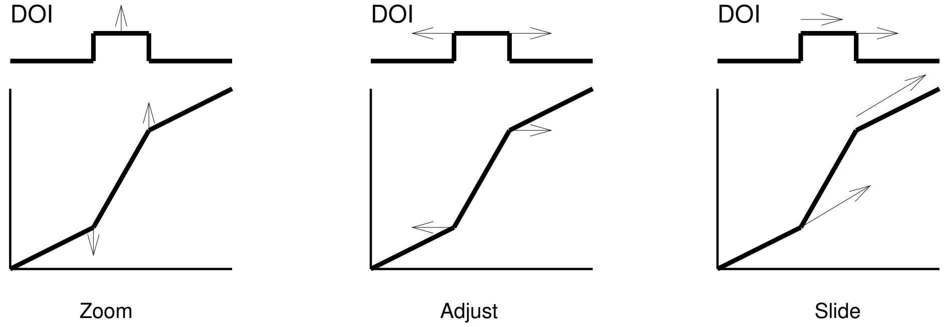

Figure 3: The three canonical focal manipulation operations are simple transformations on the DOI/transfer function. In each of the cases, the direction of the arrows can be reversed all at once.

table. Presentation types may have a range of sizes in which they are usable and paint different amounts of ink depending on the space available.

- *User Choices*. A presentation type may provide a variety of options which can be directly set by users. For example, the bar presentation allows the user to choose how the bars are scaled to fit the available space including whether the left edge is zero or the minimum value. The category variable presentation allows the user to control the number of colors used and the mapping from category values.
- *Spotlighting*. Particular cells can be accented based on some criteria chosen by the user. For example, the user can spotlight quantity values that match some numerical predicate or special values like medians or quartiles or particular category values.

TABLE LENS USER INTERFACE

The screen interface for the current Table Lens system is shown in Color Plates 1 and 2. All plates are based on the baseball statistics data described above. Interaction is based on a small number of keyboard commands and pointer gestures. In particular, we use two mouse buttons: one for "touching" and another for "grasping." Pointer gestures are performed by pressing the touch button and drawing a stroke (e.g. "flicking" in various directions). Objects are dragged using the grasp button.

Focal Manipulation is supported using control points and pointer gestures. Grasping the control point at the upper-left corner cell is used to zoom all cells, and control points on each focus are used to adjust-zoom that focus. Touching any region in the context will slide the current focus to that location. Grasping any focus allows dragging (sliding) that focus to a new location. New foci are added by grasping a cell in the context which initiates a drag of a control point for adjust-zooming the new focus. Keyboard commands allow removing all focal spans on each dimension.

A number of operators are provided on columns. They can be moved between three levels of focus (hidden, non-focal, or focal) with "flick left" and "flick right" gestures (Color Plate 2 shows columns at each level of focus). Columns can be rearranged by grasping the column label and moving it to a new position in the table. Columns can be sorted in descending or ascending order with "flick down" or "flick up" gestures. All color plates show some column that has been sorted. Finally, new columns can be added and derived by a calculation over other columns, constants, and standard series. Color Plate 1 shows "Avg" and "Career Avg" columns that have been derived by dividing "Hits" by "At Bats."

Graphical mapping parameters and spotlights can be selected using column-specific or global dialog boxes. Controls include selection of applicable presentation types and presentation type parameters for the column (e.g. what value the left edge and right edges of the column represent in the bar presentation or how colors are assigned in the category swatch presentation). The user can also spotlight particular values in a column and focus on spotlighted rows. In Color Plate 2, the extremes, quartiles, and median in the "Hits" column and the right fielders in "Position" column are spotlighted. In addition, the focus has been set to the rows containing spotlighted values in the "Hits" columns with a keyboard command.

DATA ANALYSIS SCENARIOS

Color Plates 3, 4 and 5 illustrate various ways the Table Lens can be used to explore patterns in the data and investigate various explanatory models. Interestingly, every observation made below would be readily acknowledged by any baseball fan (of course, statistics-loving) as baseball-sensical.

In Color Plate 3, the quantitative variable "Career At Bats" is sorted. This reveals fairly clear correlation in the "Years in Major" and "Career Hits" Column. Inspecting the latter, the two values that stick out the most off the generally decaying "Career Hits" curve, are Wade Boggs and Don Mattingly. To confirm, what the eye sees, the "Career Average" column is derived by dividing "Career Hits" by "Career At Bats." This column confirms that the two have the highest two career averages, and furthermore reveals a reasonably flat pattern (the observed correlation) showing, as a statistician might quip, a regression to the average. (Also, note the increased noiseness with decreasing number of "Career At Bats.") This pattern remains prominent even when the column is non-focal and other sorts have been performed (as in other plates). Notice that both Boggs and Mattingly are, justifiably, paid relatively well for their topflight batting yields.

In Color Plate 4, a series of sorts on the category variables "Position," "Team," "Division," and "League" reveals the nesting of the last three variables, and the fact that every team carries a complete stable of player positions. The sporadic positions

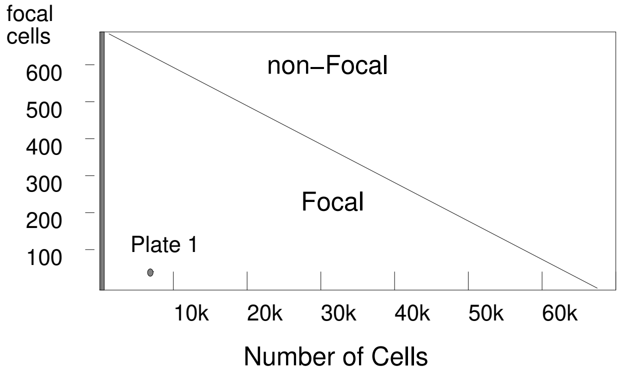

Figure 4: The gray strip indicates the region reachable using a spreadsheet. The Table Lens by trading full-sized cells for non-focal can show over 2 orders of magnitude more cells.

at the right of the column represent combination codes for players that divided their play significantly among more than one position. Also apparent by the patterns in the category variables "League '87" and "Team '87" (at the extreme right) is the relatively small amount of migration between years.

In Color Plate 5, quantitative performance based on category is explored. "Hits," and then "Position" are sorted. This shows the hits distributions within each position. The 2nd and 3rd batches from the bottom stand out somewhat; these are the right fielders and first basemen, reasonable positions in which to place strong hitters with less than stellar fielding skills. Confirming this, the right fielders are spotlighted (red bands) in Color Plate 2, and there is a clear bunching of the bands toward the top against the complete hits distribution. Also in Color Plate 5, on the offensive statistics, catchers and first basemen show the largest volume of "Put Outs" and the three infielders dominate in "Assists."

DISCUSSION

Interestingly, all of the observations made in the last section arose during demos to baseball fans who forgot about the demo and wanted to explore some particular relationship in the data. In addition, we have tested the Table Lens informally on a half dozen different datasets (Cars, Place Rated Almanac, Stock Market Data, Breakfast Cereal). In each of them, we were able to quickly find interesting correlations or patterns that made sense based on a basic understanding of the domain. For example, many relationships between car origin (American, European, or Japanese) and other properties, and changes to cars over time were observed. In several cases, we needed deeper domain knowledge to explain or confirm the validity of easily-observed patterns.

The baseball statistics table contains 323 rows by 23 columns for a total of 7429 cells. This is 11 times (an order of magnitude) more cells than our estimated maximum of 660 cells in a standard spreadsheet display. We calculate that the maximum size table the Table Lens can display on a 19 inch screen is about 68,400 cells more than two orders of magnitudes greater than a spreadsheet. Figure 4 depicts the advance in size of information sets achieved by our technique. The gray strip shows the displayable region of a typical spreadsheet program, where all cells are focal. The rest of the figure shows how a larger information set can be handled by progressively converting focal area into non-focal area.

Moreover, most of the patterns easily found using Table Lens would have been much harder or impossible to detect using a traditional spreadsheet. Most exploratory data analysis packages (e.g. S) require much greater overhead to learn and don't offer Table Lens's ease of interaction. Further work is necessary to systematically measure, compare, and explain the costs of extracting various information or performing various tasks using the Table Lens, spreadsheets, and exploratory data analysis packages.

CONCLUSION

Focusing on tables, we have gone beyond the usual design of a general focus+context mechanism to the complete design of end-user functionality. Perhaps the most interesting aspect of this work is the powerful way in which the Table Lens fuses graphical and symbolic representations into a coherent display. This fusion in combination with a small set of interactive operations (sorting and search) enables the user to navigate around and visualize a large data space easily isolating and investigating interesting features and patterns.

REFERENCES

1. George W. Furnas. Generalized fisheye views. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 16–23. ACM, April 1986.

2. J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 173–179. ACM, April 1991.

3. G. G. Robertson, S. K. Card, and J. D. Mackinlay. Information visualization using 3d interactive animation. *Communications of the ACM*, 36(4), 1993.

4. George G. Robertson and J. D. Mackinlay. The document lens. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov 1993.

5. Manojit Sarkar and Marc H. Brown. Graphical fisheye views of graphs. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 83–91. ACM, April 1992.

6. Manojit Sarkar, Scott Snibbe, and Steven Reiss. Stretching the rubber sheet: A metaphor for visualizing large structure on small screen. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov 1993.

7. Robert Spence and Mark Apperley. Database navigation: An office environment for the professional. *Behavior and Information Technology*, 1(1):43–54, 1982.