FanLens: A Visual Toolkit for Dynamically Exploring the Distribution of Hierarchical Attributes

Xinghua Lou*   Shixia Liu†   Tianshu Wang‡

IBM China Research Lab

ABSTRACT

Radial, space-filling visualization is very useful for representing the distribution of attributes in hierarchical data; however it also suffers from its drawbacks in terms of view transition, context preservation, thin slices, flexibility and large sized data support. To address these problems, we propose FanLens, an enhancement upon existing approaches with new features like incremental layout and fisheye distortion based selecting. This visual toolkit also features dynamic hierarchy specification, dynamic visual property mapping, smooth animation, etc. We illustrate the effectiveness of our technique with two examples of case study and results from informal user experiments.

Keywords: Radial space-filling visualization, dynamic hierarchy specification, fisheye distortion, visual property.

Index Terms: K.6.1 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces

1 INTRODUCTION

Information Visualization, especially space-filling methods, has been proved to be a useful approach to understanding the distribution of attributes e.g. terms in text [9], web search results [8], document content [4], stock market performance [17]. Among them the rectangular and radial layout methods are more popular. However, evaluation of these two methods [15, 14] indicates that benefiting from its explicit portrayal of structure, the radial method aids task performance more frequently in both correctness and time.

Several attempts have been made to visualize hierarchical data using radial, space-filling methods [5, 1, 16, 19]. However, they more or less suffers from drawbacks such as lack of flexibility, context loss or visual clutter (i.e. thin slices). In this paper, we present FanLens, a toolkit that enhances the conventional radial, space-filling visualization (e.g. Sunburst) mainly with incremental layout and fisheye distortion based selecting.

The remaining sections are organized as follows. In section 2 we give a brief summary of radial, space-filling visualization methods. Section 3 describes the supported data format of FanLens and corresponding data transformation process. Next, in section 4 we present its features, followed by section 5 where two case studies are introduced in detail. Then, section 6 discusses the advantages and pilot user experiments. Finally, section 7 presents the conclusions and potential future work.

2 RELATED WORKS

An early case of applying radial, space-filling method to visualize hierarchical data was based on Pie Chart. Dix and Ellis [5] enhanced the Pie Chart by allowing users to drill down into one slice

---

*e-mail: xinghua.lou@iwr.uni-heidelberg.de

†e-mail: liusx@cn.ibm.com
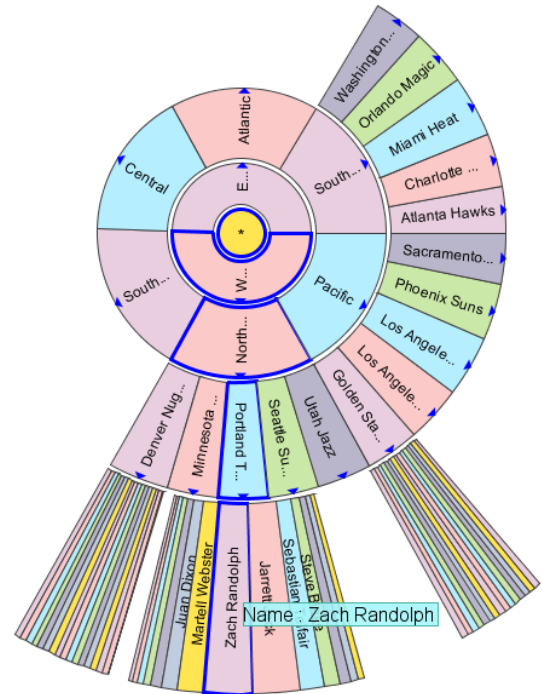
‡e-mail: wangtsh@cn.ibm.com

Figure 1: An example of FanLens visualization, an incremental, radial space-filling visualization.

and fill the area with its children. This method has the drawback of losing the important information of the hierarchy.

To preserve the hierarchy information, Andrew and Heidegger proposed Information Slices [1] which uses cascading semi-circular discs to compactly visualize large hierarchies. Selecting one slice in the overview disc, the subbranch (i.e. the selected slice and its descendants) is extracted and occupies the next semi-circular disc. This method forms an "overview + detail" scenario; however, Stasko and Zhang felt that the alternating between overview and focus is not smooth and flexible enough and they proposed three distinct methods to address this problem [16]: the angular detail method, the detail outside method, and the detail inside method. The basic idea of their methods is to shrink the overview and thus free more space for drawing the expanded focus. Their methods improve the alternating between overview and focus but this transition is still not smooth enough because each change of focus may cause the emerging and disappearing of new visual objects.

Regarding the limitations of the previous methods, the InterRing proposed by Yang et al. [19] achieves better visualization quality by providing powerful interactive distortion functions including circular distortion and radial distortion. Circular distortion deals with the sweep angle of the focus. It increases or decreases the sweep angle of the focus and meanwhile decreases or increases the angles of the rest slices. Radial distortion works the same way as circu-

lar distortion but changes the radii of the rings (i.e. the hierarchy layer) instead. The InterRing preserves the hierarchy context well and also maintains smooth transition; however, the distortion methods also introduce another problem: losing the quantitative attribute context. Usually the slice angle in radial, space-filing visualization represents quantitative attributes, e.g. file/directory size in disk space visualization [16], term occurrences in document content visualization [4]; even if the raw data contains hierarchy information only, the angle is proportional to the branch size. As a result the circular distortion method breaks this important context between slices inside and outside the focus. Another problem is that though the distortion methods enlarge the thin slices in the focus, they may also cause the emergence of new thin slices outside the focus which forms new visual clutter.

3 DATA TRANSFORMATION

Basically, FanLens supports input data with hierarchy, e.g. family tree, network hierarchy and organization structure, and the data should be saved in GraphML [7] format. However, to free the users from the tiresome labor of converting their data into GraphML format, we support dynamic data transformation from tabular data to hierarchical data.
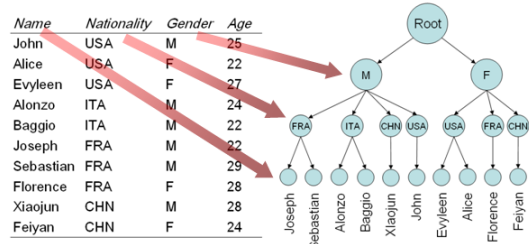

Figure 2: An example of structuring the hierarchy from the tabular data in the order of *Gender*, *Nationality* and then *Name*.

Hierarchy can be structured by breaking down the table in order of its attributes (columns), as shown in Figure 2. Dynamic data transformation means that users can create various hierarchies on demand [18]. We followed the dynamic hierarchy specification idea introduced in Treemap 4.0 [3]. For nominal attributes, users can structure the data according to the value; for quantitative attributes, users can bin them into different ranges and impose the results on structuring the hierarchy. We also implemented easy-to-use interfaces for specifying the hierarchy and binning the data (Figure 3). For advanced users, we provide the Hierarchy Specification Script based on basic XML format where the relevant parameters are defined. In addition, we also provide semi-automatic structuring of the hierarchy, which first breaks down the data according to nominal attributes ordered by their possible values ascendingly and then allows users to modify the specification.

4 VISUALIZATION DESIGNS

In this section, we consider the visual design for the FanLens approach. We first present the basic principles of our incremental, space-filling visualization method, followed by the solution to the typical thin slice problem. The rest talks about other aspects including dynamic visual property mapping, animation and navigation cues.

4.1 Incremental Layout

The incremental layout method used in FanLens follows the general format of the traditional Sunburst visualization. Inspired by the idea of SpaceTree [12], we apply incremental layout which follows the following two principles:
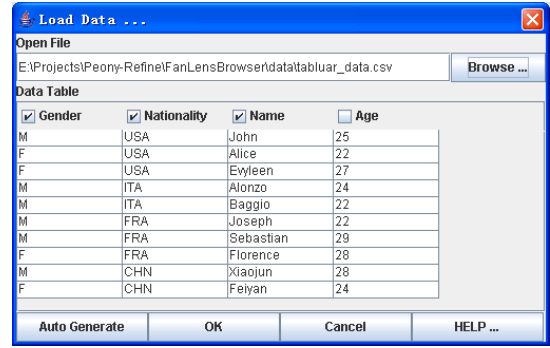

Figure 3: The interface for dynamic hierarchy specification.

- *Base level visualization*, which means when visualizing complex data FanLens does not lay out the entire hierarchy initially but displays only several most important levels which are regarded as the base levels (Figure 4(a)). Base levels represent the high-level information and are always displayed. By default, these base levels contain the top three levels of the hierarchy but they can also be configured by the users (Figure 4(b)).
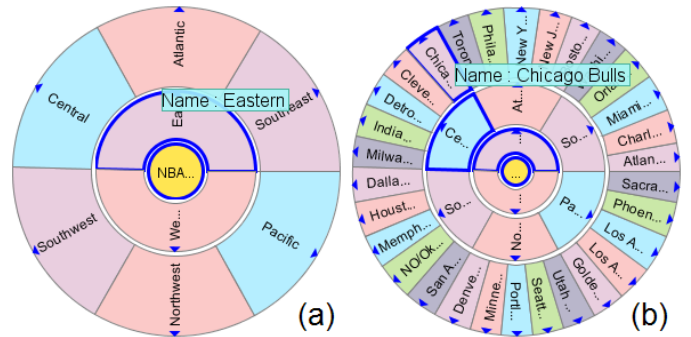
Figure 4: Examples of base level visualization and redefinition. (a) Three levels are defined as base levels by default; (b) Increase the base levels into four levels.

- *Expanding/collapsing mechanism*, which allows users to drill down into lower levels by expanding one branch from the higher level after the base levels are displayed. The newly expanded branch will be incrementally laid out around the periphery of its parent slice, radially, and is regarded as the focus. And this focus will be emphasized by increasing its radius and meanwhile decreasing the radii of its ancestors in every higher level (Figure 5(a)). In this way, users can browse the branch level by level (Figure 5(b)). Multiple foci in one display is also supported. Users may browse into a new branch while keeping the previously visited ones, as shown in Figure 6. This feature is very useful when comparing different data points that locate far from each other in the raw data.

4.2 Thin Slice Problem

Thin slice problem is one particular weakness of radial, space-filling visualization, which arises when the angle of one slice is too small to distinguish it from its sibling slices (Figure 7(a)). In this section, we will introduce our solution to this problem in terms of zooming and selecting.

Figure 5: Examples of expanding/collapsing mechanism. (a) Expanding one branch from the base levels; (b) Drill down deeper into the branch.

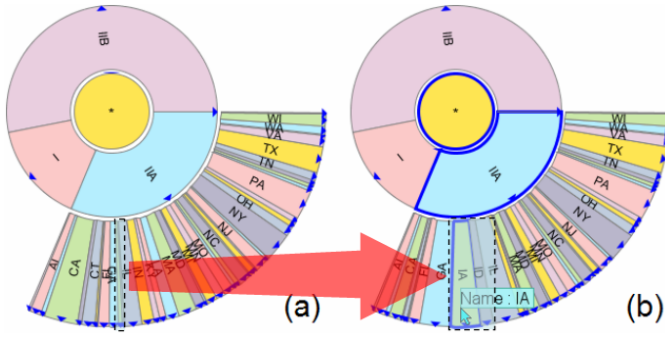Figure 7: An example of zooming in FanLens. (a) Thin slices in radial, space-filling visualization; (b) Zooming enlarges the sweep angle of the focus.

and $\theta_m + \theta_0$ to avoid unnecessary change of the slices further from the distortion center.

Figure 6: An example of multiple foci in FanLens.

Figure 8: Illustration of the basic idea of fisheye distortion.

4.2.1 Zooming

Zooming is the classic method to solve this problem which is implemented by enlarging the sweep angle of the focus so all the thin slices in it are enlarged as well [16, 19]. Our solution follows this effective and intuitive idea (Figure 7 (b)) and brings two advantages from the expanding/collapsing mechanism. For one thing, it maintains smooth transition of views because the focus was changed where it was and no reposition is needed. For another, it preserves the context of quantitative attributes better because the angles of slices outside the zooming area are unchanged.

4.2.2 Selecting

Selecting the thin slice is another potential problem that has not been addressed specifically. A previous solution to this was zooming before selecting, which surely works but also lowers the efficiency. In many occasions users need to locate one slice quickly and precisely to find some detailed information or drill down into deeper levels.

We proposed an effective solution to this problem by applying fisheye distortion [6] to the slice angles in one certain hierarchy layer. Figure 8 shows the basic idea. The angle corresponding to mouse position θ_m is the center of distortion where the slice borders are repelled from neighboring borders thus the slices are enlarged. The range in which fisheye distortion affects is fixed within $\theta_m - \theta_0$

The fisheye distortion formula we used is basically the classic one which works in three steps. Firstly, angles θ within the range $\theta_m - \theta_0$ and $\theta_m + \theta_0$ are normalized into range $[-1, 1]$ (Equation 1). Then fisheye transformation is applied to all normalized angles according to Equation 2. Finally, the transformed angles θ'_f are mapped to the range $\theta_m - \theta_0$ and $\theta_m + \theta_0$ (Equation 3).

$$\theta' = \frac{(\theta - \theta_m)}{2\theta_0} \quad (1)$$

$$\theta'_f = \begin{cases} \frac{(1+d)\theta'}{(1-d\theta')} & \theta' \in (-1, 0] \\ \frac{(1+d)\theta'}{(1+d\theta')} & \theta' \in (0, 1) \end{cases} \quad (2)$$

$$\theta_f = \theta'_f \theta_0 + \theta_m \quad (3)$$

Figure 9 shows the effect of fisheye distortion. Figure 9 (a) is the view before applying the distortion where several clusters of thin slices exist and slices in them cannot be distinguished from neighboring ones. But with fisheye distortion turned on, the thin cluster is enlarged when the cursor moves nearby and users can clearly inspect its contents, as shown in Figure 9 (b).

Figure 9: An example of fisheye distortion based selecting. (a) View with no fisheye distortion; (b) View with fisheye distortion turned on.

4.3 Dynamic Visual Property Mapping

We make use of two visual properties in FanLens: angle and color. Both of them can represent quantitative attributes and color also works for enumerative attributes. FanLens supports dynamic visual property mapping, namely users can change the mapping of angle/color to attributes on demand and the view is updated accordingly.

4.3.1 Mapping of Angle

Mackinlay's hypothesis [10] about perceptual accuracy of visual properties on quantitative data indicates that the human's perception of angle is more accurate than that of color. Thereby angle should be mapped to the key attributes in the data. For example, Figure 10 (a) viusalizes the AAUP (American Association of University Professors) 1994 Salary Survey data with angle mapped to professor's salary; if the mapping is switched to the number of professors, the view looks quite different, as shown in Figure 10 (b).
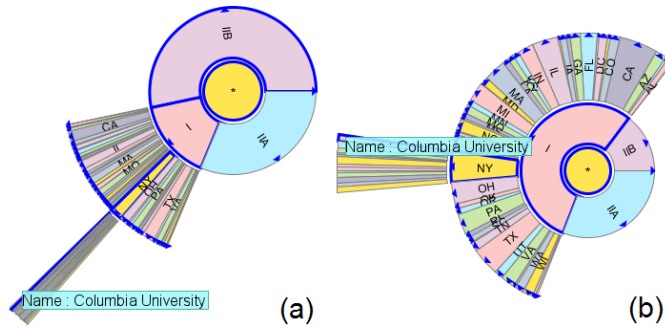

Figure 10: An example of dynamic mapping of angle. (b) Mapping angle to professor's salary; (b) Mapping angle to number of professors.

4.3.2 Mapping of Color

Color mapping is optional in FanLens. If no mapping is specified, FanLens will automatically assign a color to each slice and the purpose is to make neighboring slices easier to be distinguished. On the other hand, the coloring strategy can be described from the following two aspects:

- We use HBS colors with variations in brightness of a given hue and saturation to represent the quantitative attribute following the Color Use Guidelines by Cynthia Brewer [2] as well as the points stated in Bernice Rogowitz's paper [13].

- We follow the structure-based coloring strategy proposed in the InterRing [19]. The color of a parent slice is derived by averaging the colors of its children.

Figure 11 (a) shows the coloring with no color mapping specified; Figure 11 (b) shows the result if color mapping is targeted at professor's salary (with angle mapped to the number of professors).
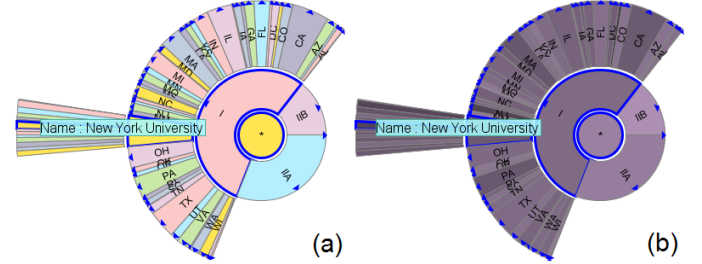

Figure 11: An example of dynamic mapping of color. (a) Without color mapping specified, colors are automatically assigned to distinguish nearby slices; (b) Color is mapped to professor's salary and the darker color represents higher payment.

4.4 Visual Cues

In this section, we introduce the design of visual cues that help users understand the transition of views and guide their exploration.

4.4.1 Animation

In FanLens animation was implemented regarding two rules as follows. For one thing, we use the slow-in, slow-out timing [20] instead of straight linear timing, which inspires the users to anticipate the change. For another, the newly expanding branch gradually grows out of its parent slice (Figure 12) while the collapsing one shrinks into it, which is really an intuitive design that accords with the meaning of 'hierarchy'.
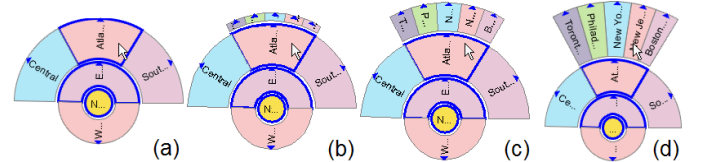

Figure 12: An example of animation: expanding branch grows out of its parent slice.

4.4.2 Exploration Navigation

Two visual cues are designed for users' navigation. The first one is to show an expandable mark (e.g. an arrow) at the outer periphery (Figure 13) of a slice which indicates that it has child slices and can be expanded. The second design is to use landmarks (highlight) to help users remain oriented of their exploration path, which is quite necessary when the focus is zoomed. Figure 14(a) shows an enlarged focus but its parent is ambiguous. This issue can be addressed by highlighting (e.g. thicker slice border) the entire path from root to the focus, as shown in Figure 14(b).

5 CASE STUDY

To evaluate the effectiveness of FanLens, we performed two case studies using our research prototype.
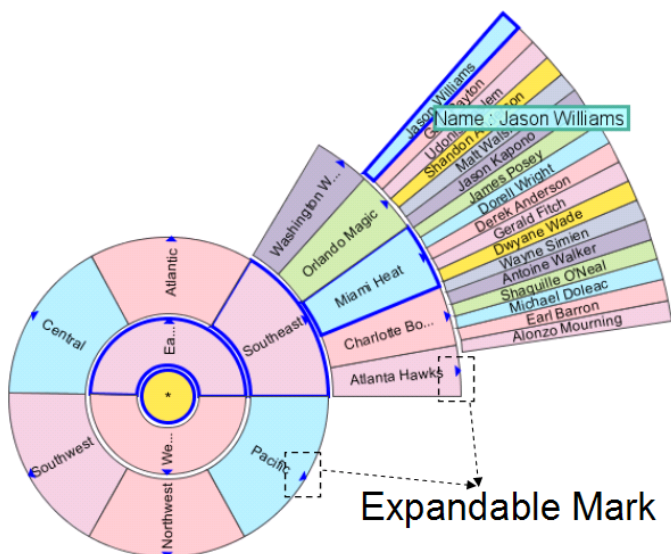
Figure 13: An example of expandable mark that indicates whether users can drill down from one slice.
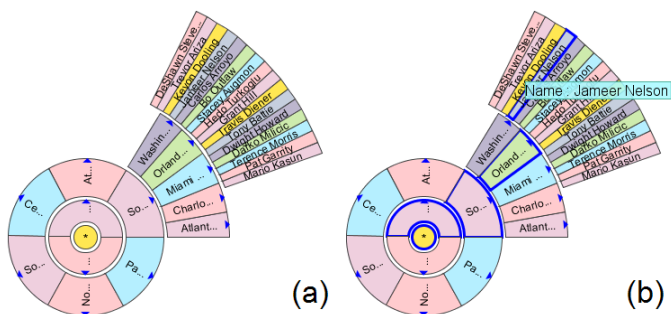

Figure 14: An example of exploration landmarks that deal with the ambiguity of the hierarchy. (a) The parent of the focus is ambiguous; (b) The entire exploration path is explicit

5.1 AAUP Data

The AAUP dataset comes for the ASA Statistical Graphics Section's 1995, containing information on faculty salaries for 1161 American colleges and universities. It has four categorical attributes, including FICE (Federal ID number), college name, state (postal code) and type (I, IIA, or IIB). All the rest attributes are quantitative, such as average salary of faculties, average compensation of faculties, number of full professors, etc.

5.1.1 Unusual Data Detection

Suppose more information about the full professors needs be discovered, the data can be hierarchically structured in order of *type*, *state* and *college name*. After mapping the slice angle to the number of full professors and the slice color to the average salary of full professors, a start-up display with default base level definition is shown in Figure 15(a). Considering the semantic meaning of visual presentations, it is immediately clear that type I has more professors than the other two in all and the average salary there is the highest. Increasing the base levels by one we arrive at Figure 15(b) showing the breakdown of each type by state, from which we can conclude that California has the most full professors and notice several states with higher pay of full professors including California, New York, Massachusetts, Pennsylvania, etc. Re-defining the base levels to cover the entire hierarchy will result in the traditional Sun-

burst visualization where it is easy to detect universities or colleges with extreme value in the same way.
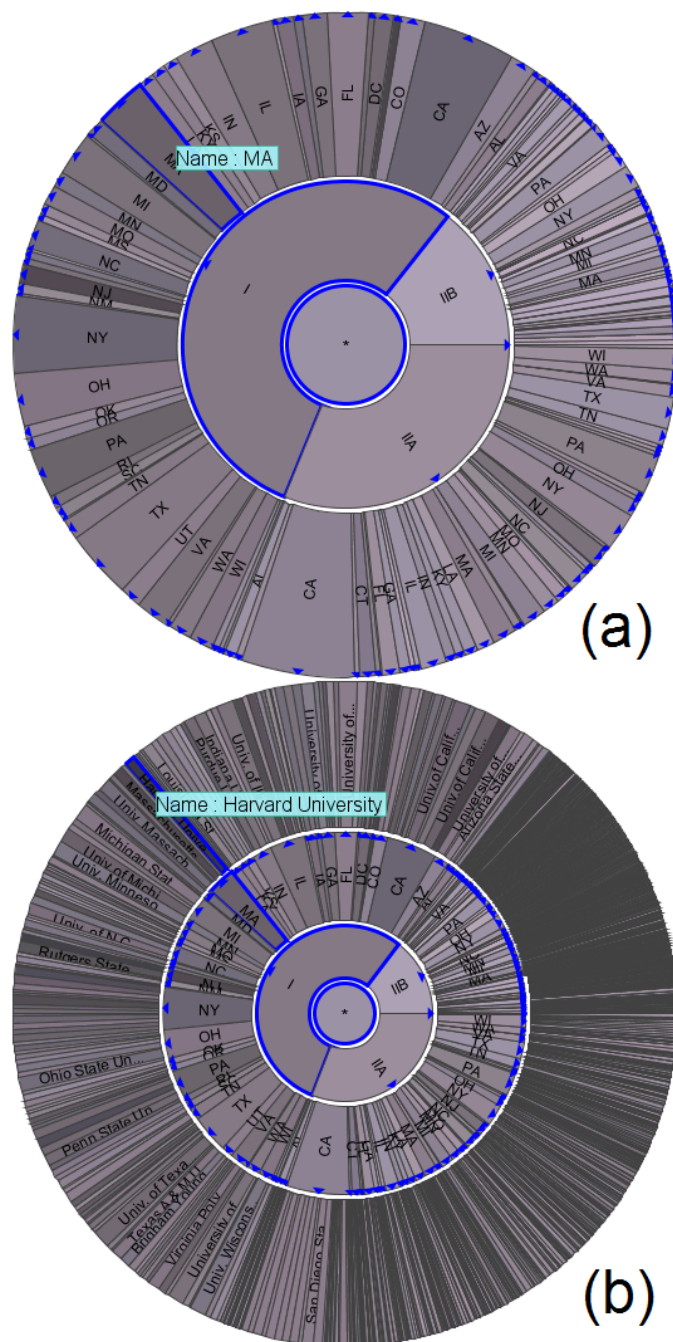

Figure 15: An example of re-defining the base levels to discover extreme value. (a) Default base levels cover the top two levels; (b) Increase the base levels to discover unusual value in lower levels.

5.1.2 Partial Data Exploration

California drew our attention due to its large amount of full professors. So we would like to explore the schools in this state for more detailed information. However, because of the thin slice problem, this task is difficult to complete in the Sunburst visualization which only portrays the overview. Therefore, we returned to the default base level and drilled down into the target branch level by level.

The coloring strategy made a thin (small number of full professors) but dark (higher pay) slice noticeable. By selecting the slice with the support of fisheye distortion or zooming we found out that it represents California Institute of Technology, as shown in Figure 16.

Figure 16: An example of partial data exploration for colleges in California.

Figure 17: An example of multiple foci data analysis for inspecting colleges in California with different types.

5.1.3   Multiple Foci Data Analysis

After the partial data exploration of California's type I colleges, we got interested in learning about all colleges in California. This could become a tiresome job if the data is represented in other formats, e.g. Excel, Information Slice [1], because the interested data may be scattered and users have to use several windows to show it within one view. Even if the users can use the SORT feature well, there may still be too many data points so they have to scroll up and down to find a special one. However, in FanLens, users can easily inspect them in one view using multiple foci function (Figure 17).

5.2   Basketball Player Statistics

This data contains the statistics of all NBA players for season 2005-2006 [11], such as PPG (Points Per Game), RPG (Rebounds Per Game), TO (Turnovers Per Game), etc. Categorical attributes include Conference (Eastern, Western), Division (Pacific, Central, etc), Team, Position and Player.

5.2.1   Overall Evaluation

Figure 18 shows the result of structuring the data in order of *Conference*, *Division*, *Team* and *Player* and defining the base levels to cover the entire hierarchy. The slice angle and slice color were mapped to player's offensive ability and defensive ability, respectively. The offensive ability corresponds to the sum of PPG(Points Per Game) and APG(Assists Per Game) and the defensive ability is the sum of the following attributes: RPG (Rebounds Per Game), SPG (Steals Per Game) and BPG (Blocks Per Game). This overall evaluation indicates that, even though player's offensive and defensive abilities vary a lot, the league is quite balanced in the levels of Conference, Division and Team because the angle and color of one team, division or conference are very close to its siblings. That is

Figure 18: An example of overall evaluation for studying the balance of the NBA league.

one important reason why NBA games are usually exciting because they are always close match-ups.

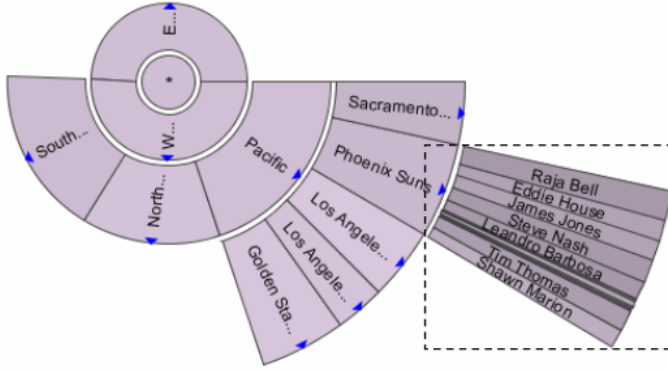## 5.2.2 Special Pattern Discovery

Figure 19: An example of special pattern discovery for studying the 3-Point shooting ability of the NBA teams.

Figure 19 visualizes the same hierarchy but was dedicated to analyze the 3-point shooting ability. The angle and color were mapped to 3PM(3-Points Made per game) and 3P%(3-Point shooting Percentage), respectively. The larger angle of the Pacific Division guided us to explore it and find a larger slice corresponding to the Phoenix Suns, which actually was a team that attempted 3-points shooting a lot. One reason could be found by observing a special pattern that most of its players had close 3-Point shooting percentage (corresponding to the color). That is to say most players in that team had close 3-Point shooting ability so everyone was willing to do it.

5.2.3 Hypothesis Testing

To study the correlation between player's scoring ability and mistakes, we first attempted to use Figure 20(a) where angle and color were mapped to scoring ability (PPG, Points-Per-Game) and mistakes (TO, turnovers), respectively. This visualization is not really intrinsic but it guided us to hypothesize that players with high scoring ability also make more turnovers. To test this hypothesis, we reconfigured the hierarchy by ranging the players into five categories according to their PPG (see the following table) and visualizing the new hierarchy with the same visual property mapping, as shown in Figure 20(b). This new visualization proves our hypothesis that players with stronger scoring ability also give more turnovers as the color of slices in the "Super" and "Strong" categories are mostly very dark.

PPG	Category
$25.0 < $ PPG	super
$18.0 < $ PPG ≤ 25	strong
$10.0 < $ PPG ≤ 18.0	regular
$5.0 < $ PPG ≤ 10.0	low
PPG ≤ 5.0	poor

6 EVALUATION AND DESIGN EXPERIMENTS

Incremental layout is the most remarkable feature of FanLens, which brings several benefits as follows:

- The major benefit is the flexibility. The traditional Sunburst lays out the entire tree at the start-up, providing the overview but lacking partial data exploration and effective interactions on the display. Overview is certainly important but on some occasions users need to focus on one branch and wish to understand its contents exclusively. On the other hand, overview
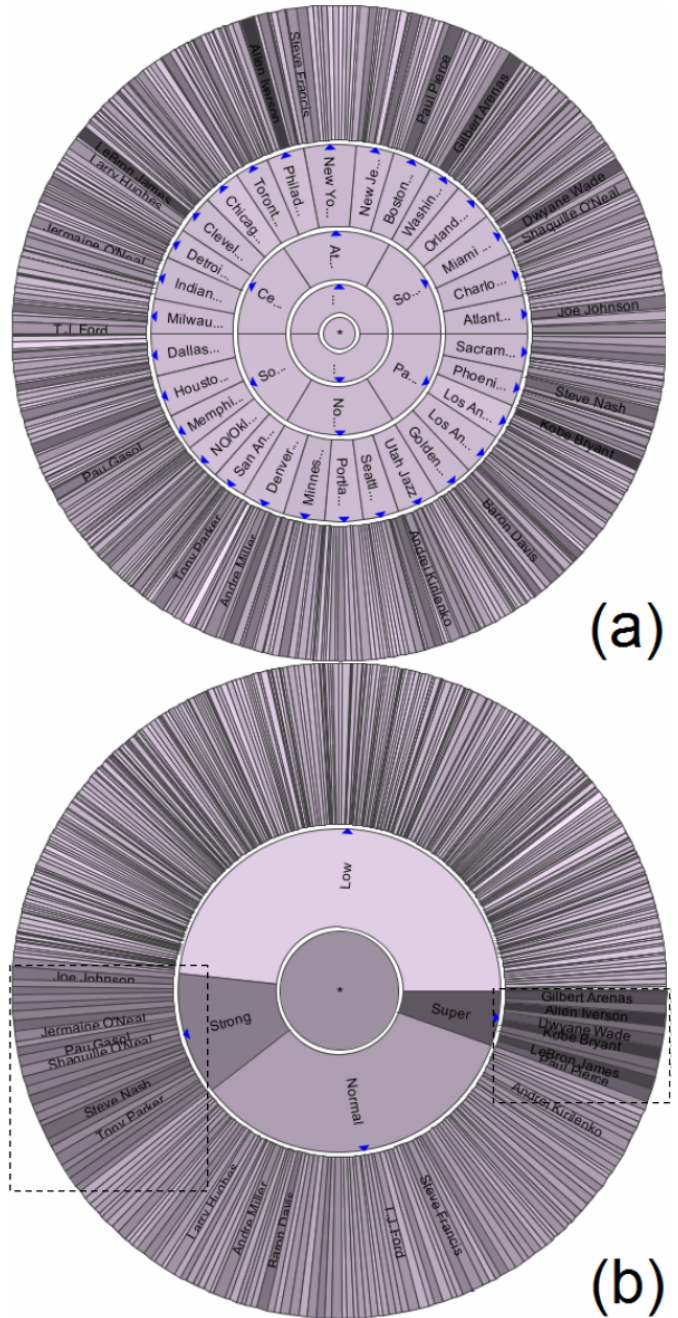

Figure 20: An example of hypothesis testing for studying the connection between players' scoring ability and turnovers. (a)Use overview to evaluate players' scoring and turnover. (b) Reconfigure the hierarchy and get a direct understanding of the connection between players' scoring ability and turnovers.

is also available in FanLens by redefining the base levels to cover the entire hierarchy, which will identically create the Sunburst visualization.

- Another advantage is the readability which is first achieved by the expanding/collapsing mechanism, offering the users with clear view of the exploration path and the structure of the focus. It is even improved by the emphasis on the focus as described above. It should be noted that increasing the radii will

not break context of quantitative attribute because the angle is preserved. Figure 21 (a)(b) shows the comparison of readability in the Sunburst and the FanLens.
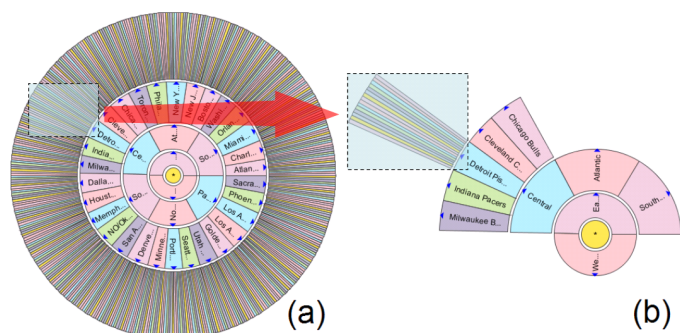

Figure 21: Comparison of readability between Sunburst and FanLens. (a) Focus in Sunburst; (b) Same focus in FanLens.

- Incremental layout accords with incremental data loading. This makes FanLens advanced in visualizing large-sized data or real-time data. Only the high-level summarization data is fetched at the beginning and the reset detail data is loaded when required.

We carried out experiments throughout the entire design progress to evaluate alternative approaches and collect useful suggestions. For example we have performed preliminary study on the effectiveness of fisheye distortion based selecting method. The data sets used in the evaluation were the AAUP Data and Basketball Player Statistics introduced above. The participants were asked to perform tasks like listing the colleges in a certain state whose faculty number is less than 50 or finding the player with least turnovers for each team. Three approaches were implemented for comparison, including selecting after zooming, selecting without zooming and selecting with fisheye distortion. The result indicates that the performance of first two approaches is very close and the fisheye distortion based approach can save up to 30% of the time needed to complete the same task.

7 Conclusions and Future Work

We have introduced the FanLens, an approach for dynamically exploring the distribution of hierarchical attributes. Our primary contribution is an incremental, radial space-filling visualization method and a fisheye distortion based selection method. It also supports dynamic hierarchy specification, visual property mapping and exploration navigation. FanLens better preserves the context, provides extra flexibility and maintains smooth view transition. We believe that the ideas in this paper, e.g. the incremental layout and the fisheye distortion based selecting, can also be applied to other hierarchical data exploration or radial, space-filling visualization methods.

Pilot user experiments have initially showed the effectiveness of our design and now we are working on the formal evaluation of this approach. In addition, we plan to improve the design with new features, e.g. hierarchy specification using direct DnD (Drag and Drop) operation on the display and automatically hierarchy generation functions using techniques like hierarchial clustering or SOM.

References

[1] K. Andrews and H. Heidegger. Information slices: visualising and exploring large hierarchies using cascading, semicircular discs. In *IEEE Symposium on Information Visualization, Late Breaking Hot Topic Paper*, pages 9–12, 1998.

[2] C. Brewer. *Visualization in Modern Cartography*. Elsevier Science, 1994.

[3] G. Chintalapani, C. Plaisant, and B. Shneiderman. Extending the utility of treemaps with flexible hierarchy. In *Proceedings of the Eighth International Conference on Information Visualisation*, pages 335–344, 2004.

[4] C. Collins. Docuburst: Document content visualization using language structure.

[5] A. Dix and G. Ellis. Starting simple - adding value to static visualisation through simple interaction. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 124–134, 1998.

[6] G. W. Furnas. Generalized fisheye views. In *Proceedings of the ACM-SIGCHI Conference on Human Factors in Computing Systems*, pages 16–23, 2000.

[7] GraphML. graphml.graphdrawing.org.

[8] Grokker. www.grokker.com/grokker.html.

[9] M. Hearst. Tilebars: visualization of term distribution information in full text information access. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 59–66, 1995.

[10] J. Mackinlay. Acm transactions on graphics. In *Automating the Design of Graphical Presentations of Relational Information*, pages 110–141, 1986.

[11] NBA. www.nba.com.

[12] C. Plaisant, J. Grosjean, and B. B. Bederson. Spacetree: supporting exploration in large node link tree, design evolution and empirical evaluation. In *IEEE Symposium on Information Visualization*, pages 57–64, 2002.

[13] B. Rogowitz and L. Trelnlsh. Data visualization: the end of the rainbow. *IEEE Spectrum*, 35:52–59, 1998.

[14] J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald. An evaluation of space–filling information visualizations for depicting hierarchical structures. In *International journal of human–computer studies*, pages 663–694, 2000.

[15] J. Stasko, M. Guzdial, and K. McDonald. Evaluating spacefilling visualizations for presenting hierarchies. In *IInformation Visualization Symposium 1999, Late Breaking Hot Topics*, pages 35–38, 1999.

[16] J. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualisations. In *IEEE Symposium on information visualisation*, pages 9–12, 2000.

[17] M. Wattenberg. Visualizing the stock market. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, pages 188–189, 1999.

[18] R. Wilson and R. Bergeron. Dynamic hierarchy specification and visualization. In *Proceedings IEEE Symposium on Information Visualization*, pages 65–72, 1999.

[19] J. Yang, M. O. Ward, and E. A. Rundensteiner. Interring: an interactive tool for visually navigating and manipulating hierarchical structures. In *IEEE Symposium on Information Visualization*, pages 77–84, 2002.

[20] K. P. Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of graphs with radial layout. In *Proceedings of Information Visualization 2001*, pages 43–50, 2001.