# Interactive Radviz Visualization System Using Hierarchical Approach for Large-Multidimensional Data

Hyeonsik Gong, Hyoji Ha*
Lifemedia Interdisciplinary Program, Ajou University

Kyungwon Lee†
Department of Digital Media, Ajou University

## Abstract

This study aims to improve one of the methods used to express large amounts of multidimensional data in two dimensions, that is, Radviz. In particular, by enabling the grouping of specific nodes using a novel method that involves clustering large amounts of data hierarchically, this study alleviates node overlapping, which is one of the limitations of previous Radviz systems. The findings of this study allow the user to cluster repeatedly after setting a specific group of data, such that the user can efficiently group large amounts of data as well as explore the desired cluster of data. Furthermore, the findings of this study enable the user to grasp the similarity between the selected groups and determine by how much each dimension forces a specific group through given detailed views.

**Keywords:** Data visualization, similarity measurement, clustering, hierarchical structure

**Index Terms:** Information interfaces and presentation—User Interfaces—Graphical user interfaces

## 1 Introduction

In visualization field, renowned studies have been carried out steadily on multidimensional data. In particular, one of the most important issues in this field of research is placing multidimensional data in a human-visible space. One of the conventional methods for solving this issue, called dimensionality reduction (such as t-SNE), converts multidimensional data into two- or three-dimensional data for visibility. However, a limitation of this method is that it cannot define exactly how much each axis is affected by each dimension.

Another method of placing multidimensional data in a human-visible space that also solves the aforementioned disadvantage is Radviz [1]. Radviz is a visualization technique that places multidimensional data inside a circle with the attraction of dimensional anchors (DAs) located in the circumference of the circle. However, Radviz has a number of limitations too, such as possible overlapping of the nodes in the center and distortion of each node's position, which conceals the extent to which dimensions influence the nodes [2]. These disadvantages tend to become worse as the amount of data increases.

To address these problems, this study introduces a visualization system that can provide the conventional Radviz technique with a hierarchical data structure. In addition to the effects of previous research, which help the user grasp the relationship between clusters and the internal data of each cluster [3], this study allows the user to perform iterative clustering within specific data groups after the entire hierarchical data structures have been understood.

The four main objectives of this visualization are as follows: 1) Providing an exploratory visualization system via user-driven interactions 2) Enhancing clustering performance by dividing data between the user's actions 3) Grasping the hierarchical structure of large-scale data intuitively 4) Helping the user to understand the distribution of values for each dimension of selected Radviz nodes.

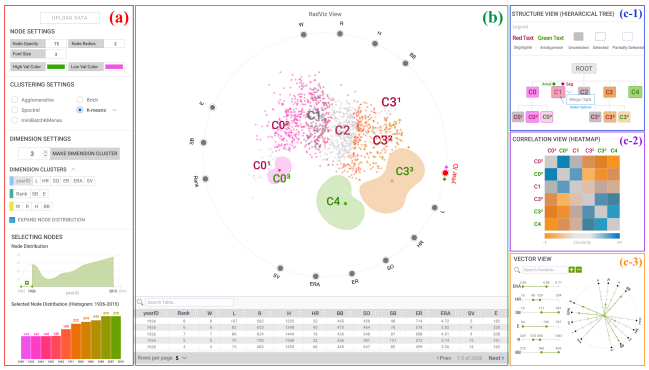


Figure 1: Overview of the visualization system in which user can set options that can be applied to visualization, such as setting a node's dimension range section and clustering method (a). Radviz arrangement based on options and table to actual data of nodes (b). Structural view (c-1) allows user to understand data hierarchically with some interactions included. Correlation view (c-2) expresses the relationship between selected groups with heatmap. Vector view (c-3) shows the direction and magnitude of force received by selected nodes on each dimension.

## 2 Methods

### 2.1 Navigation Layout

On the left side of the visualization system, the user can upload data and set options for the data nodes to be placed in Radviz. In the Node Settings section, the user can set graphical options such as transparency and radius of the nodes. In the Clustering Settings section, the user can select the clustering method and tune detailed parameters of the method. The Dimension Settings section allows the user to cluster each dimension to determine the position of the dimensional anchors that will be initially placed on the circumference of Radviz. This section would help the user understand a relatively similar dimension group.

Furthermore, the user can understand the distribution of data for a specific dimension via the Selecting Nodes section, and control the range of the dimension as desired, which will be reflected in the bottom node histogram. This allows the user to adjust data freely, preventing the automatic even division of the data distribution of numeric dimensions.

### 2.2 Main Layout

The Main Layout on the center consists of a Radviz and a table that shows the actual data of the nodes. The position of each node is determined by the sum vector of every dimension vector with magnitude value obtained via the min–max scaling of each dimension, and the direction was determined from the center of the circle to position of the corresponding dimensional anchor.

*e-mail: {overholic10, hjha0508}@ajou.ac.kr
†e-mail: kwlee@ajou.ac.kr

When the user clicks a specific anchor, the color of each node is determined subject to the color in the node distribution of the corresponding dimension. The user can select specific nodes via interaction by clicking on the node or dragging the area in the Radviz. At this time, the color of the unselected node is changed to gray, whereas the data in the table and the visualization of the Detailed Layout on the right side are updated based on the selected data.

### 2.3 Detailed Layout

First, the user is able to determine the hierarchical structure of data in Radviz through structural view. By clicking a button, the user can merge specific clusters or split a cluster with additional clustering. Moreover, the corresponding data on Radviz can be viewed as a group or individually via an amalgamate/segregate button. The corresponding nodes are highlighted on Radviz as they hover around each cluster group in a tree structure. Via this process, users can identify a large amount of data structures instantly and seek desired data by narrowing down the cluster.

The correlation view shows the similarity between groups containing the selected data. This similarity is represented by a value that is changed to a scale of -1 to 1 after adding up and averaging the similarity calculated between each data dimension in the corresponding group. This view helps the user understand similar groups and notifies the extension of clustering in the hierarchical structure of the data.

In vector view, the user can observe the vectors of each dimension, which are the elements that determine the location of the data selected in Radviz. The user can also filter the selected data by specifying a range of values of specific dimensions. Using this view, the user is able to grasp exactly how much of the data in Radviz is affected for each dimension.

## 3 Study & Insights

To clearly explain each layout, we obtained an annual MLB team data from SeanLahman.com [4] and applied it to the system. For an easy explanation of data, we filtered 14 relatively similar variables from the data before uploading them to the system.

First, with the Dimension Clustering section, the user can obtain three dimension clusters $\{yearID, L, HR, SO, ER, ERA, SV\}$, $\{Rank, SB, E\}$, and $\{W, R, H, BB\}$. If the user wants to see the data from 1926 to 2015 alone, the yearID button is pressed, which then controls the data in the corresponding range of Node Distribution and divides them into 9-year units that can be checked later in the Selected Node Distribution.

Next, the user sets the color of nodes with yearID and confirms that data of similar colors are gathered perfectly in relatively close proximities. It is plausible to infer that the reason for the location of the nodes on the Radviz is because the number of baseball games increased as time changed, together with the increasing number of strikeouts and home runs in recent years, which led the user to the conclusion that the strategy of each team changed with time.

The user then re-clusters the C0 and C3 groups for an in-depth understanding after determining five primary cluster groups in the structure view. To view the data distribution of a specific cluster group in detail, the corresponding nodes are spread on Radviz by clicking the Segregate button for each group. Subsequently, via the Correlation View section, the user can check the correlation between cluster groups to which the data currently selected on Radviz belongs. Consequently, it can be observed that the $\{C0^3, C0^2, C1\}$ and $\{C3^2, C3^3, C4\}$ groups are similar (Figure 2).

Finally, among currently selected data in the vector view section, the user can check the vectors that determine the location of data with an ERA of 4 or less and a number of 135 home runs or more (Figure 3). Consequently, the user can verify that the corresponding data was mostly influenced in the $\{W, H, HR\}$ order of dimension and was less influential in the $\{SB, RANK, E\}$ order. Therefore, it can be observed that the teams belonging to this data are the teams with many wins and fewer errors.

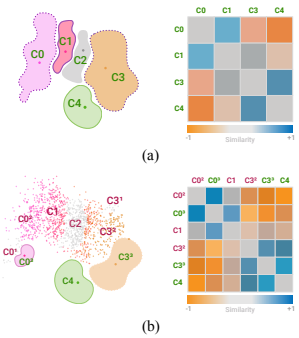


Figure 2: (a) Initial Radviz and heatmap. (b) Radviz and heatmap after performing cluster segregation and additional cluster split.

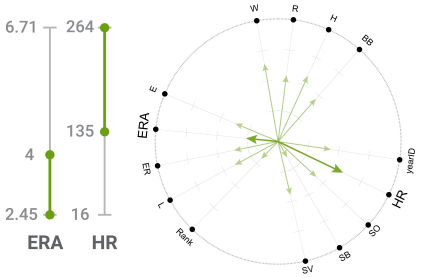


Figure 3: Example of vector view that can filter selected data according to specific conditions (ERA 4 or lower, HR 135 or higher), and view direction vectors that determine location of data.

## 4 Conclusion

The proposed visualization system enables the user to interactively explore a large amount of multidimensional data by iteratively performing various types of clustering methods. Via this system, the user is able to check the hierarchical structure of the data, understand the similarity between each group, and observe the direction and intensity of dimension vectors that determine the location of specific data in the Radviz space. The proposed system can accommodate additional processes, such as an elaborated version of clustering algorithms and the option of determining the order of clustering and dimensionality reduction in a pipeline to allow a user to freely participate in the processing of data. Future studies can contribute to the formation of more sophisticated Radviz, including the method of dealing with categorical dimensions when forming Radviz and that of determining the position of dimensions, such as adjusting the weight of each dimension.

## References


[1] Enrico Bertini, Luigi Dell'Aquila, and Giuseppe Santucci. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In *Coordinated and Multiple Views in Exploratory Visualization (CMV'05)*, pages 22–29. IEEE, 2005.

[2] Jorge Henrique Piazentin Ono, Fabio Sikansi, Débora Cristina Corrêa, Fernando Vieira Paulovich, Afonso Paiva, and Luis Gustavo Nonato. Concentric radviz: Visual exploration of multi-task classification. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 165–172. IEEE, 2015.

[3] John Sharko, Georges Grinstein, and Kenneth A Marx. Vectorized radviz and its application to multiple cluster datasets. *IEEE transactions on Visualization and Computer Graphics*, 14(6):1444–1427, 2008.

[4] Sean Lahman. Lahman's baseball database.