

A Scalable Correlation-Driven Framework for Compact Visualization of Multidimensional Data using Adaptive Projection Technique

Shridhar B. Dandin, Santosh Kumar Das, and Megha Sinha

Cite as: Dandin, S. B., Das, S. K., & Sinha, M. (2026). A Scalable Correlation-Driven Framework for Compact Visualization of Multidimensional Data using Adaptive Projection Technique. International Journal of Microsystems and IoT, 4(4), 1876–1883.
<https://doi.org/10.5281/zenodo.21980786>



© 2026 The Author(s). Published by Indian Society for VLSI Education, Ranchi, India



Published online: 30 April 2026



Submit your article to this journal:



Article views:



View related articles:



View Crossmark data:



<https://doi.org/10.5281/zenodo.21980786>



A Scalable Correlation-Driven Framework for Compact Visualization of Multidimensional Data using Adaptive Projection Technique

Shridhar B. Dandin¹, Santosh Kumar Das², Megha Sinha³

^{1,2}Department of Computer Science and Engineering, Sarala Birla University, Ranchi, Birla Knowledge City, Mahilong, Purulia Road, Jharkhand, India-835103

³Department of Computer Science and IT, Arka Jain University, Jamshedpur, Jharkhand, India-832108

ABSTRACT

In the era of rapidly developing digital technology, the application of data visualization has increased exponentially. But in data visualization, the graphical charts or tools are only limited to represent the relationship among two or three parameters of the dataset. However, most real-world datasets contain many more attributes. Projecting two parameters on to two-dimensional display prevents the discovery of multidimensional associations like correlations. In this paper, a scalable correlation-driven framework is proposed for Compact Visualization of Multidimensional Data (ComVisMD). The proposed method, ComVisMD, is based on an adaptive projection technique that maintains essential relationships among visualization elements. It is simulated using the 'graphics' package in the Python programming language. It is used to display five dimensions for mapping five attributes of dataset into 2D compacted formation. It easily helps to illustrate the hidden information of the dataset. Moreover, users can interactively map the display to search for data insights. The performance of ComVisMD is compared with other common graphical methods. Additionally, comparative analysis is also done with some existing methods. Finally, it is validated with a cricket players' dataset based on visual analytics parameters.

KEYWORDS

Multidimensional Data Visualization; Data Projection; Data Aggregation; Correlation Analysis; Visual Analytics for Complex Data; Intelligent Decision Support Systems.

1. INTRODUCTION

The rapid development of digital technologies consists of several cutting-edges applications in the area of artificial intelligence, data science, cloud computing, internet of things, etc. Each of the applications provides numerous benefits in real-life society [1, 2]. But the capacity of the dataset increases exponentially which causes issues with storage, speed, data quality and visualization. Moreover, converting the dataset into meaningful information becomes a crucial point in different domains. In this case, data visualization plays a vital role in transforming the crucial information into a meaningful graphical representation [3, 4]. There are several representations of data visualization such as charts, dashboards, graphs, maps, networks, etc. The most important method is the chart which helps in different visualization application such as comparison, trend analysis, composition, distribution, relationship analysis, project management and performance monitoring, etc. [5, 6]. It is easy to understand and visualize data based on an effective decision-making system. It is expressively used in different applications such as smart cities, education, cybersecurity, finance, healthcare, business intelligence, etc. [7,8].

1.1 Motivation

In recent decades, numerous approaches have been proposed in the field of data visualization which are discussed in the next section.

However, most of these methods do not focus on compact visualization techniques for multi-dimensional datasets. With the rapid advancement of artificial intelligence, new visualization techniques are emerging continuously, while the size and complexity of datasets are growing exponentially. This creates a significant challenge in effectively representing high-dimensional data in an interpretable and space-efficient manner. Therefore, there is a strong need for a tool that can visualize multi-dimensional datasets in a compact form, tailored to user requirements.

1.2 Contributions

The contributions of the proposed method, ComVisMD, are:

- i. A compact map-based visualization for multi-dimensional data using cell representation.
- ii. A two-dimensional partitioning approach with limited categories to reduce visual clutter.
- iii. Integrated multiple visual encodings (color, shape, internal patterns) for enhanced information representation.
- iv. Interactive dimension switching to explore patterns and correlations.
- v. Effectiveness demonstration through scatterplots' comparison and improvement in visual analysis capability.

1.3 Organization of the paper

Section 2 discusses several existing works as a literature review. Section 3 describes the main methodology of the proposed method and in the Section 4 discusses simulation analysis of the designed methodology. Finally, in the last section i.e. Section 5 discusses conclusions and future work.

2. LITERATURE REVIEW

In the last few years, several works have been proposed in the area of data visualization. Some of the works are discussed in this section. T. Houichime and Y. Amrani [9] proposed a method for call graphs that used pattern detection and visualization. It helps in dynamic warping and data profiling for data visualization. The framework of the proposed method is divided into two parts. The first part is used to visualize runtime data. The second part is a method of visualization for symbolic sequencing in the pattern. It also helps to analyse the classification process, optimal length, and symbolic and compact translation. M. Fan and L. Zhou [10] proposed a method for medical visualization aimed at design purposes. The proposed method consists of cognitive habits and analytics logic for distinguishing among the design, target user analysis, and classifying the task based on descriptive analysis. Finally, it analyses three models of medical visualization. M. Cuesta et al. [11] proposed a method for high-dimensional datasets to support class separability. The name of the method is CSViz, which indicates class-separable visualization. It is a novel visualization method that is used to overcome the issue of visualization labelled with the help of visualization analytics. Finally, the proposed method is simulated and analysed based on 50 datasets with different features. F. Hong et al. [12] proposed a method for ensemble data aimed at visual analytics. The main purpose of this work is ‘realization set analysis’ with large-scale of data. It is also used to find relations among data. It helps to design a two-dimension matrix for ensemble data. The visualization contains similar feature in a subspace for each intersection. T. Xiao et al. [13] proposed a method for a social network service to support a document embedding system and topic transition visualization. It analyses the abstract of the document based on two-dimensional space. The visualization of the work uses time series analysis for word analysis. Finally, it analyses different data such as user rating, COVID-19 data, and nuclear disaster data. G. Wu et al. [14] proposed a method for visual analysis in the context of damage data from simulations. It uses an integrated visual method for defect analysis, including clustering and identification and graph tracking. It overcomes the issue of blur modes and defect dynamics in the visualization. C. Kelleher and A. Braswell [15] designed and illustrated a recommendation approach for an introductory overview based on data visualization. The visualization method is designed for scientific data visualization with respect to large set of data. The main reason behind this is the rapid growth of a large set of data along with big data. The work preserves complex data

visualization in multiple dimensions. Finally, the work outperforms static visualization with interactive visualization using animation. In [16], the authors analysed the color assignment for optimization purposes. It is based on visual separability for strong influence. It is also based on a multi-class structure that uses the visualization method of scatterplots. The method used a genetic algorithm to optimize a good color assignment. V. Molchanov and L. Linsen [17] proposed a method for preserving shape in star coordinates. In this method, all dimensions are explored within the whole set of attributes. Analysis tasks: detecting clusters, trends and outliers. The method focuses on minimizing loss of visualization related information. In this paper, the authors used a 2D visual space technique along with an encoding method for mapping more than one-dimensional data. It improves the tool for exploration by linear mapping. A different color scheme is used for visualizing membership information. A. Sarikaya and M. Gleicher [18] proposed a survey for scatterplot analysis that helps the designer make choices in the visualization. It includes several design decisions that highlight different features. Three basics are connected in this survey such as design choice, data features and analysis tasks. Based on this survey, it overcomes the limitation of traditional scatterplots. The methodology designed by S. Ferré et al. [19] is used for a multi-dimensional cube. It helps to organize different subsets of the cube with the help of query. It provides a set of navigation based on structured dimensions. Finally, it describes several representations such as multi-value context, sets of objects, aggregate value, etc. The early origins of the methodology designed by M. Friendly and D. Denis [20] for analyzing scatterplots in visualization. This paper traced several developments of the scatterplot in history. It also discussed modern inheritors and recommended its unique benefits. It includes several graphical components to enhance the quality of the visualization including annotation. M. Y. Huh and K. Kim [21] proposed a method for data visualization in multi-dimensional data. The method is named as Grand Tour technique. It helps to change 2D projection of the data and illustrates several structures and patterns. Finally, it helps to avoid unused motion and enhance the clarity and interpretability of the visualization.

3. METHODOLOGY OF THE COMVISMD

The description of the proposed method ComVisMD is divided into three major subsections. The first subsection gives an outline of the proposed method including the framework of the model along with flowchart. The second subsection describes a rigorous and comprehensive analysis of the proposed method with the algorithm and mathematical model. The third subsection illustrates the GUI design with different visualizations.

3.1 System Architecture and Workflow

In this subsection, the workflow of ComVisMD is discussed based on the system architecture. The analytical framework of the proposed method is shown in Fig. 1. It describes the process of designing the model based on a multidimensional dataset. Data is pre-processed for normalization and cleaning. Then attributes are mapped for object portioning to categorized and group. It helps to visualize a multidimensional dataset into a compact 2D form. So, it helps to find hidden correlations and patterns in the dataset. The overall flowchart of the proposed method is shown in Fig. 2. This flowchart illustrates the overall execution process of the proposed method ComVisMD. The execution flow is based on several tools and tips of the GUI. Its execution is designed in an iterative ways fusion of some operations such as selection of dimension, map view, hole, color, cell shape, and change scenario, etc. This operation is user-driven and runs whenever users choose to execute the model.

3.2 Algorithm and Mathematical Modelling

In this subsection, the proposed algorithm is described as pseudocode along with different mathematical modelling that are used in the background of the proposed method. The pseudocode of the ComVisMD is shown in Algorithm 1. The algorithm contains the process of proper execution of the model including the dataset in the background as Comma Separated Values (CSV) format. This dataset consists of different attributes for handling visualization outcomes. The event-driven approach of the proposed model consists of two rows with two buttons as horizontal and vertical. The execution process is based on operations such as color, hole, cell shape, map view and dimension selection. The proposed method supports two primary types of data: numerical and categorical. The input data is expected in CSV format in two pre-decided top rows. The first row contains the attribute names (column headers). The second row specifies the data type for each attribute, using one of the types mentioned in Table 1.

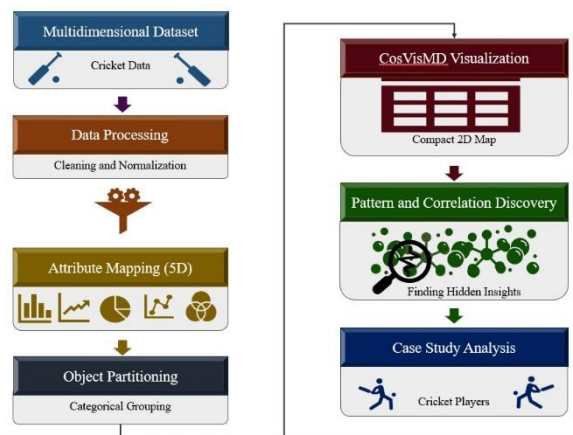


Fig. 1 Analytical framework of the ComVisMD.

The data type informs the system how to process and encode each attribute properly. For example, arithmetic operations like sums or averages are meaningful only for numerical data

as REAL or INT. In contrast, categorical data like CATS or CATL represent qualitative distinctions such as gender or country and require discrete encoding methods. The proposed method ComVisMD is demonstrated on a cricket players dataset with 2170 records from 16 countries affiliated to the International Cricket Council (ICC). Attributes include player ID, years active, matches played, runs, batting averages, bowling stats, catches/stumpings, and country affiliation. The overall data preprocessing of the proposed method involved converting fields to consistent types, cleaning missing values, and specifying attribute types.

The mathematical modelling of the proposed method depends on several data types. The main aim of this modelling is to classify numerical data into distinct performance regions based on their statistical distribution. The goal is to provide a compact visual representation of performance levels using color-coded regions derived from finite data statistics. The detailed visualization statistics are shown in the next subsection as GUI Design and Visualization. The underlying logic is formulated using mathematical notations as follows:

Let the numerical data, as shown in Eq. 1, be:

$$X = \{x_1, x_2, x_3, \dots, x_n\} \subset \mathbb{R}, \text{ where } n \in \mathbb{N} \quad (1)$$

where, X represents the set of values of any numerical attribute (e.g. Runs), the x_i represents individual data (e.g. runs scored by a batsman, column value of respective record) from X and n represents the total number of records (rows).

Then, the central break point 'BP' is evaluated as shown in Eq. 2:

$$BP = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

where, x_i represents individual data from X and n represents the total number of records (rows).

Then, in the lower half subset, let BP_{min} be the average value in that subset and we can evaluate BP_{min} as shown in Eq. 3.

$$BP_{min} = \frac{1}{n_{low}} \sum_{x_i \in x_{low}} x_i \quad (3)$$

where, x_i represents the individual data from the subset of X and n_{low} represents total number of records or rows of that subset as shown in Eq. 4.

$$X_{low} = \{x_i \in X \mid x_i \leq BP\}, n_{low} = |X_{low}| \quad (4)$$

Then, in the upper half subset, let BP_{max} be the average value in that subset and we can evaluate BP_{max} as shown in Eq. 5.

Algorithm 1 The pseudocode of the ComVisMD.	
Step 1: Start	
Step 2: Input dataset in CSV format, with the first-row column headers indicating dimensions (attribute names) of the dataset and the second row specifying the datatype of each dimension	
Step 3: Execution of the ComVisMD tool	
Step 4: Display of two rows, in the form of a menu	
(a) Top row indicates three options: SelectDimensions, MapView and Quit	
(b) Second row indicates five dimensions	
(i) Horizontal-red colored	
(ii) Vertical-grey colored	
(iii) Color-blue colored	
(iv) Hole-green colored	
(v) CellShape-yellow colored	
Step 5: Click on the 'Horizontal' button, now all the dimensions (attributes) of the dataset are displayed in the third row after the second row of five dimensions	
(a) Click on one of the dimensions	
(b) The selected dimension is displayed in red color indicating the selected dimension for 'Horizontal' dimension	
Step 6: Click on 'Vertical' button	
(a) Click on one of the non-selected dimensions	
(b) Selected dimension is displayed in grey color indicating 'Vertical' dimension	
Step 7: Click on 'Color' button	
(a) Click on one of the non-selected dimensions	
(b) Selected dimension is displayed in blue color indicating 'Color' dimension	
Step 8: Click on 'Hole' button	
(a) Click on one of the non-selected dimensions	
(b) Selected dimension is displayed in green color indicating 'Hole' dimension	
Step 9: Click on 'CellShape' button	
(a) Click on one of the non-selected dimensions	
(b) The selected dimension is displayed in yellow color indicating 'CellShape' dimension	
(c) The categories (values) of the selected dimension are displayed as options in the next row	
Step 10: Select the options up to a maximum of SEVEN from the options of 'CellShape' and once finished the selection, click on 'DONE'	
Step 11: Click on the 'MapView' of the top row to visualize the map of the objects based on dimensional attributes	
Step 12: Complete the Visual Analysis task, to know all the details of any object, click on any of the cells to display a 'detail's window'	
Step 13: To change the scenario, or to restart the 'dimension selection process', click on the 'SelectDimensions' option from the top row and repeat the process from Step 5 .	
Step 14: To exit from the tool, select the 'Quit' button.	
Step 15: Stop	

$$BP_{max} = \frac{1}{n_{high}} \sum_{x_i \in x_{high}} x_i \quad (5)$$

where, x_i represents the individual data from the subset of X and n_{high} represents total number of records or rows of that subset as shown in Eq. 6.

$$X_{high} = \{x_i \in X \mid x_i \leq BP\}, n_{high} = |X_{high}| \quad (6)$$

Then, the classification function is derived to map each value 'x' that belongs to 'X', as Eq. 7:

$$f(x) = \begin{cases} \text{Region 1: Very Low,} & \text{if } x < BP_{min} \\ \text{Region 2: Low,} & \text{if } x = BP_{min} \\ \text{Region 3: Mid-Low,} & \text{if } BP_{min} < x < BP \\ \text{Region 4: Mid,} & \text{if } x = BP \\ \text{Region 5: Mid-High,} & \text{if } BP < x < BP_{max} \\ \text{Region 6: High,} & \text{if } x = BP_{max} \\ \text{Region 7: Very High,} & \text{if } x > BP_{max} \end{cases} \quad (7)$$

Table. 1. The list of data types.

Types	Description
REAL	Floating point numbers (continuous values)
INT	Integer values
STR	String or text (non-categorical)
DAT	Date/time values
CATS	Categorical data with up to 7 unique categories
CATL	Categorical data with 8 or more unique categories

3.3 GUI Design and Visualization

In this subsection, the GUI designed is described along with complete event-driven visualization. In the proposed method ComVisMD, the map displays all or a selected subset of dataset objects as cells arranged by row and column clusters. Users can dynamically reorder the mapped attributes to explore new patterns. The cells are colored by their mapped attribute using a 7-color scale, hole sizes represent one attribute proportionally, and cell shapes indicate categorical classes. Clicking a cell reveals object details, facilitating detailed inspection alongside high-level pattern recognition. During execution of the proposed method, a menu is presented with two rows of buttons initially. The top row has three option buttons, 'Select Dimensions', 'Map View' and 'Quit', as shown in Fig. 3. By clicking the 'Select Dimensions' option the user can select or change the attributes to map the respective dimension. The second option, 'MapView' is made available for users to choose once all five attributes are mapped to the respective five dimensions. The last option 'Quit' can be selected to close the tool.

The second row of the menu displays 5 options representing 5 dimensions of the map. At present, the user has to choose the options strictly in the order, 'Horizontal' to 'CellShape'. Each option is identified with a distinguishable color, which will be an indicator during tool execution. The option 'Horizontal' is chosen to select an attribute to map in the horizontal direction. As this is the first option, once clicked all the attributes of the chosen CSV format file are displayed in the next line. User has to complete the selection by clicking on one of the attributes. The selected attribute is marked in red color (same as 'Horizontal'). The next option, 'Vertical' can be clicked to select the attribute to map in the vertical direction. The selected attribute is colored in gray which is similar to the 'Vertical' option. In the same way, the next three options, 'Color' (blue colored), 'Hole' (green colored) and 'CellShape' (yellow colored) are clicked in order to map the selected attribute to the respective dimension. For the last option, it is necessary to select a categorical attribute for representing the cell shape. Then the user has to select the available options to assign one shape each to categorical values. Maximum of seven options

can be selected to assign a cell shape. Then, the option, 'DONE' is selected to complete the mapping of attributes with dimensions. Now, 'Map View' is available for selection. If the user selects the 'Map View' option then the compact map is displayed. For the respective dimensions, the legends are displayed with same background color to show the similarity

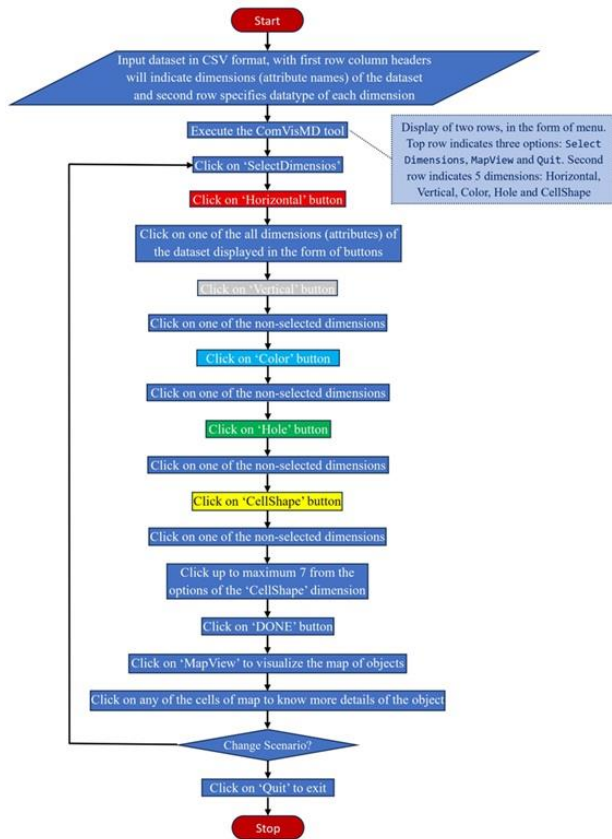


Fig. 2 Flowchart of the ComVisMD.



Fig. 3 ComVisMD menu to explore selected dataset.

With reference to Fig.4, the display area presently shows the records (objects, called cells in a map) of 7 countries that are represented with distinct shapes. As the user has selected the columns, Match, HighScore and Runs for dimensions mapping, the focus of the view is on batsmen. It would have been the other way around for bowlers if the attributes Balls and Wicket were selected for mapping. A total of 1023 records (players) from 7 out of 16 countries are displayed for analysis. Firstly, we are grouping the players on the basis of HighScore, the highest score of the players in their played ODIs. This information is represented in the horizontal direction of the map, the first dimension of the dataset Team16Countries.csv. Secondly, on the basis of the decreasing order of number of played matches the cells (representing players) grow in the horizontal direction on the map, second dimension, Match. Thirdly, the color of the cell represents runs scored, an indicated analysis point, and is the third dimension, represented by Runs. Fourthly, the hole on

cell represents batting average, an additional indicator for analysis, is the fourth dimension, BatAvg. Finally, the shapes of the cells representing the categorical attribute Country of the dataset, the fifth dimension assist the observer in the analysis in terms to find players from one country and concurrently compare with players from other countries.

When the user clicks the left mouse button keeping the mouse pointer on a cell, the object details are displayed in a rectangle below the map for the final view of the selected player, with the same background color as the player square. For example, the details of player, 'Ishant Sharma' from 'India' are displayed as shown in the bottom right rectangle of Fig. 4. On click of 'Select Dimensions' from the top line of the menu, the user goes back to change the attributes to map the respective dimensions, which is basically the initial point of the system ComVisMD for compact visualization. In this way, the user of the compact visualization system ComVisMD can analyse the map to produce different meanings intuitively. We can pick a few interested players among large number of players and display further details of only those interested players to know more about them one-by-one.

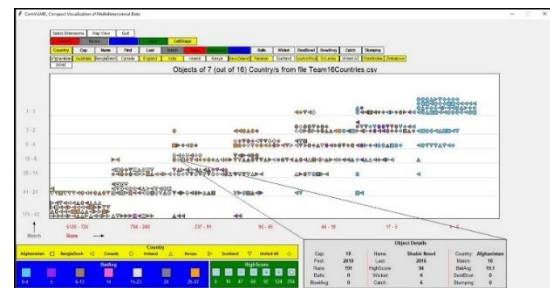


Fig. 4 Display of 1023 records in a compact 2D space.

4. PERFORMANCE EVALUATION

In this section, performance evaluation is analyzed theoretically and empirically. The data available in a table (two-dimensional structure), normally has a limited scope to explore content as a whole. Users can get a full picture of the whole data through visualizations. A visualization technique, scatterplot, displays two quantitative variables [22]. More data in terms of other dimensions can be visualized to assist user for data analysis. ComVisMD tries this by providing 3 additional dimensions in 2D display. Encoding data points by the position of a visual mark like dot, allows user to reveal several different relationships in the data [23]. ComVisMD explores such relationships like, correlation, compact clusters, distinct classes within clusters, shapes, separation of data, etc. As scatterplots represent each object in a dataset with a point, identified by Sarikaya and Gleicher in [18], in ComVisMD, we are representing the objects through map in the form of different shaped cells. During the map view when the user clicks on any of the cells, the details of the object are displayed in a separate viewing area with the same background color as the cell.

In Grand Tour [21], a series of scatterplots obtained through a smoothly changing sequence of projection methods, are examined. Here, Huh and Kim project multi-dimensional data on to 2D window. This approach doesn't help to find patterns. We are proposing our tool, ComVisMD, to explore such patterns from the dataset through mapping all attributes one-by-one in five dimensions. The matching can be changed as per requirement after making series of selections to find the patterns, clusters, correlations, etc. In our system, all five dimensions are scaled to a magical number 7. It means, 7 clusters, 7 classes, 7 colors, 7 shapes, etc. Our compact visualization tool extends the scatterplots to represent the objects with more power or additional information to make it more effective data visualization tool for pattern matching or finding correlations intuitively. The additional information can be provided in the form color of data points, data points in the form of cells, cells in different shapes, etc.

Scatter diagram as identified by Friendly and Denis in [20], is a plot having two variables plotted along its two axes and in which points are placed to show the values of these variables for each of several objects, so that the form of the association between variables can be seen. In ComVisMD, the same very old concept of horizontal versus vertical axes is used to project cells on the map. To perceive correlation, the cluster, the class, the color, the hole and even the shape dimensions are added to ease visual analysis for the user. Users will have an opportunity to make any combinations of these dimensions for producing visual maps for analysis.

The map displays all or a selected subset of dataset objects as cells arranged by row and column clusters. Users can reorder mapped attributes to dynamically explore new patterns. Cells are colored by their mapped attribute using the 7-color scale, hole sizes represent one attribute proportionally, and cell shapes indicate categorical classes. Clicking a cell reveals object details, facilitating detailed inspection alongside high-level pattern recognition. Scatterplots allow regression analysis by fitting lines and calculating R2 coefficients, providing explicit quantitative measures of bivariate relationships. ComVisMD, focusing on rich visual encodings over five dimensions, does not directly compute regression metrics but supports pattern and cluster detection visually.

1.1 First Scenario

Fig. 5 and Fig. 6 give a comparison between the visualization tools ComVisMD and Scatterplot. Fig. 5 is the view generated using the tool ComVisMD, which is a 4-dimensional compact display. It shows the objects from the cricket dataset (belonging to the Country, 'India') after selecting the attribute Match in the Horizontal dimension, Runs in the Vertical dimension, HighScore as the Color dimension, and BatAvg as Hole dimension. Fig. 6 is the scatterplot, plotted using ggplot() in R for the generation of scatterplot. Match is used as the x-axis parameter; Runs is used as y-axis parameter; to plot the objects as points. The parameter HighScore is used as third parameter for the

identification of object through color. When we compare both the figures, there is a clear identification of the problem in the scatterplot where the objects are concentrated at one place without clear distinction at the bottom-left corner. In the case of ComVisMD, there is a clear distinction of objects where the observer can easily check the details of a particular object and in fact there is no overlapping of the objects. The objects are clearly classified and clustered based on the algorithm.

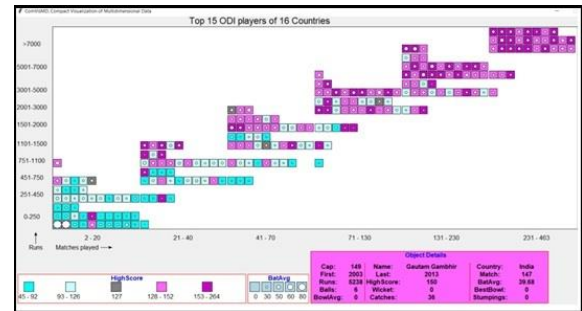


Fig. 5 ComVisMD map view selecting Match in the Horizontal dimension and Runs in Vertical dimension. Third dimension, 'HighScore' (Color dimension) and Fourth dimension 'BatAvg' (Hole dimension).

Additionally, a fourth dimension for the distinction of objects in terms of a property batting average (attribute BatAvg) is provided for finding exceptions. The size of the hole indicates a player's batting averages. Two players (objects) at the bottom left corner have very good batting average but played a smaller number of matches and scored fewer runs too. In comparison with the scatterplot, ComVisMD has a compacting factor of about 4, visualizing four dimensions on a compact 2D display space. The views helped the users to find correlations among selected attributes as well as to spot the exceptions. The system was tested through fixed or earlier decided attributes for classification, for scaling, for coloring, etc. In the latest improved version of ComVisMD, we have achieved a compacting factor of 5, as shown in Fig. 7.

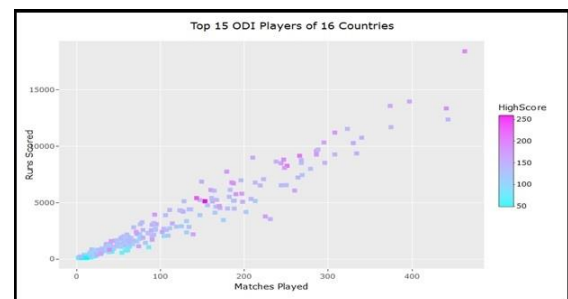


Fig. 6 Scatterplot using Match as x-axis. Third dimension, 'HighScore' (color scaling).

4.2 Second Scenario

Fig. 7 and Fig. 8 give another comparison between the visualization tools Scatterplot and ComVisMD. Fig. 7 shows

a scatterplot of objects, the players, of the cricket dataset with 3 of the dimensions of Fig. 4: HighScore, Match and Runs. For the generation of a scatterplot using Python, the attribute Runs is converted to categorical type from the numerical type. The correlation trend is as clearly visible as with ComVisMD which is shown in Fig. 8. However, the best players (in orange) are scattered in a large area of the scatterplot whereas in ComVisMD, as previously discussed, they are concentrated in a compact region

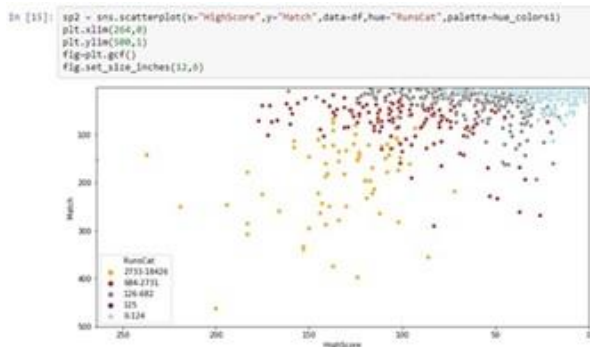


Fig. 7 Scatterplot using HighScore as x-axis, Match as y-axis and third dimension, 'Runs' as color dimension.

Furthermore, exceptions are less visible, for example in Fig. 7, there is a blue left-triangle clearly visible in region 11-19Match/44-66HighScore: it corresponds to an Indian player with a number of runs smaller than the other players in that region. That player is not visible on the scatterplot as the corresponding dot is amongst the crowded part. Thus, the partitioning process of ComVisMD into regions gives equal importance to each cell. In so doing, more space can be given to each cell, it enables to add information in them such as our 'Hole' and 'Shape' dimensions. They could be added in a scatterplot but would probably not be as legible as in ComVisMD. One can also note that the attributes of scatterplots must be numerical, whereas the partition approach of ComVisMD allows both numerical and categorical data to be displayed.



Fig. 8 ComVisMD map view selecting HighScore in Horizontal dimension and Match in Vertical dimension.

Third dimension, 'Runs' (Color dimension), Fourth dimension 'BatAvg' (Hole dimension) and Fifth dimension 'Country' (Shape dimension).

5. CONCLUSION AND FUTURE WORK

A visualization framework, ComVisMD, focuses on

compact and multi-dimensional data visualization. It uses dynamic scaling for coloring the cells of different shapes, like square, hexagon, left-sided triangle, right-sided triangle, up-sided triangle and down-sided triangle. The cells represent the objects of a dataset on a two-dimensional display called map. At present, through our system, we have contributed 5-dimensional visualization, an improvement from a 4-dimension display and object comparison without query. The system enhances the visual analysis capability of the user to derive correlations from different scenarios of visualization and automatically find the exceptions. ComVisMD is constrained with prepared or processed dataset in the form of data file in CSV format. In future, the system can be optimized according to the user's choice. ComVisMD supports categorical attribute too in comparison to only numerical data attribute supported by scatterplot. Our compact visualization tool extends the scatterplots to represent the objects with additional information to make it more effective data visualization tool for finding the correlations intuitively. At present, extension support covers the color of data points, data points in the form of cells, cells in different shapes, compaction through scaling, etc. Further, the system may be augmented with saving of visualization scenarios, provision of multiple views side-by-side, filtering, aggregation and navigation methods. The future work is extending the work to large scale datasets based on high dimensional dataset and to enhance the map display without losing performance.

REFERENCES

1. Shahzad, M., Tah, J. H., Younas, M., & Almkhtar, A. (2025). Technologies and Techniques in Digital Twins for Real-time Data Visualisation in Building Maintenance: A State-of-the-Art Review. *Journal of Infrastructure Intelligence and Resilience*, 100185.
2. Han, L., Song, J., Chen, S., & Zhong, Z. (2025). Visualization and interaction techniques for single-text digital reading: A survey. *Computers & Graphics*, 104481.
3. Wang, J., Deng, Z., Deng, D., Wang, X., Sheng, R., Cai, Y., & Qu, H. (2025). Empowering multimodal analysis with visualization: A survey. *Computer Science Review*, 57, 100748.
4. Das, S. K., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2023). A non-linear multi-objective technique for hybrid peer-to-peer communication. *Information Sciences*, 629, 413-439.
5. Brosolo, M., Asmitha, K. A., Conti, M., KA, R. R., KP, M. S., Nicolazzo, S., ... & Vinod, P. (2026). Security through the eyes of AI: How visualization is shaping malware detection. *Computer Science Review*, 61, 100914.
6. Das, S. K., & Tripathi, S. (2018). Adaptive and intelligent energy efficient routing for transparent heterogeneous ad-hoc network by fusion of game theory and linear programming. *Applied Intelligence*, 48(7), 1825-1845.
7. Chen, Y., & Huang, L. (2025). Graphics and data visualisation. In *MATLAB Roadmap to Applications: Volume I Fundamental* (pp. 241-311). Singapore: Springer Nature Singapore.
8. Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2019). Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE transactions on visualization and computer graphics*, 26(1), 12-22.
9. Houichime, T., & el Amrani, Y. (2026). Introducing a Novel Technique for Call Graph Visualization and Design Pattern Detection through Runtime Data Profiling and Dynamic Warping. *Science of Computer Programming*, 103445.

10. Fan, M., & Zhou, L. (2025). A design study process model for medical visualization: M. Fan, L. Zhou. *Journal of Visualization*, 1-21.
11. Cuesta, M., Lancho, C., Fernández-Isabel, A., Cano, E. L., & Martín De Diego, I. (2024). CSViz: Class Separability Visualization for high-dimensional datasets: M. Cuesta et al. *Applied Intelligence*, 54(1), 924-946.
12. Hong, F., Shao, H., Mei, X., Pang, A., & Yuan, X. (2024). Visual analytics of ensemble data using coupled subspaces. *Journal of Visualization*, 27(5), 867-884.
13. Xiao, T., Oda, N., & Onoue, Y. (2023). Visualization of topic transitions in SNSs through document embedding and dimensionality reduction: T. Xiao et al. *Journal of Visualization*, 26(6), 1405-1419.
14. Wu, G., Lin, D., Wang, H., & Liu, L. (2022). Visual analysis of defect clustering in 3D irradiation damage simulation data. *Journal of Visualization*, 25(1), 31-45.
15. Kelleher, C., & Braswell, A. (2021). Introductory overview: Recommendations for approaching scientific visualization with large environmental datasets. *Environmental Modelling & Software*, 143, 105113.
16. Wang, Y., Chen, X., Ge, T., Bao, C., Sedlmair, M., Fu, C. W., ... & Chen, B. (2018). Optimizing color assignment for perception of class separability in multiclass scatterplots. *IEEE transactions on visualization and computer graphics*, 25(1), 820-829.
17. Molchanov, V., & Linsen, L. (2018). Shape-preserving star coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 449-458.
18. Sarikaya, A., & Gleicher, M. (2017). Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics*, 24(1), 402-412.
19. Ferré, S., Allard, P., & Ridoux, O. (2012, May). Cubes of concepts: multi-dimensional exploration of multi-valued contexts. In *International Conference on Formal Concept Analysis* (pp. 112-127). Berlin, Heidelberg: Springer Berlin Heidelberg.
20. Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41(2), 103-130.
21. Huh, M. Y., & Kim, K. (2002). Visualization of multidimensional data using modifications of the grand tour. *Journal of Applied Statistics*, 29(5), 721-728.
22. Cleveland, W. S. (1993). *Visualizing data* Hobart Press, Summit, New Jersey. Referenced on, 84.
23. Munzner, T. (2014). *Visualization analysis and design*. CRC Press

AUTHORS



Shridhar B. Dandin completed his Bachelor of Engineering in Computer Science from Karnataka University, Master's degree in Software Systems from BITS-Pilani and doing Ph.D. at Sarala Birla University, Ranchi, Jharkhand in collaboration with INRIA Labs (LIS team) of INSA-Rennes, France. His area of research includes Compact Visualization of Multidimensional Data. Data Structures, Compiler Design, Data Analytics, Python and Object-Oriented Programming are his core areas. Over 31 years of academic and administrative experience he has developed several international collaborations and relationships. Under his guidance more than 110 students successfully completed higher education even leading to PhD and got fruitful placements in France, Bangkok and other parts of the world. He is fellow of IET, IETE and member of ACM, ISTE. After completing 25 years of service at Birla Institutions, Pilani (Rajasthan), he joined Sarala Birla University, Ranchi, Jharkhand in 2020 as Dean of Engineering and Applied

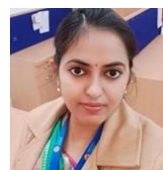
Sciences and presently Registrar at Sarala Birla University, strongly believes in religious ethnicities of India and views that, "global exposure, awareness and adoption with the root of Indian Vedic culture, will make successful managers, engineers and technocrats needed by today's world".

E-mail: Corresponding Author E-mail: sb.dandin@sbu.ac.in



Santosh Kumar Das received his Ph.D. degree in Computer Science and Engineering from Indian Institute of Technology (ISM), Dhanbad, India, in 2018 and completed his M. Tech. degree in Computer Science and Engineering from Maulana Abul Kalam Azad University of Technology (erstwhile WBUT), West Bengal, India, in 2013. Previously, he was an Assistant Professor at School of Computer Science and Engineering, National Institute of Science and Technology (Autonomous), Institute Park, Pallur Hills, Berhampur, Odisha, India. He is currently working as Assistant Professor at Department of Computer Science and Engineering, Sarala Birla University, Birla Knowledge City, P.O.-Mahilong, Purulia Road, Ranchi, India. He has more than 8 years of teaching experience. He has authored/edited 9 books with Springer in series as Lecture Notes in Networks and Systems, Tracts in Nature-Inspired Computing and Studies in Computational Intelligence, Taylor & Francis, Apple Academic Press, and IGI Global, including Scopus indexed. He has contributed more than 60 research papers, along with SCI and Scopus indexed. His research interests mainly focus on Applied Artificial Intelligence, Data Visualization, Blockchain, Deepfake, Data Visualization, and Wireless Ad-hoc Networks. His h-index is 20 with more than 1300 citations

E-mail: sunsantosh2014@gmail.com



Megha Sinha completed her B.E. from AVIT Vinayaka Mission's Chennai Campus and MTech. in Computer Science and Engineering from Birla Institute of Technology Mesra Ranchi (B.I.T Mesra) Jharkhand. She received her Ph.D. in Computer Science and Engineering, from Rabindranath Tagore University (VNTU) Bhopal. Previously, she was an Assistant Professor, Department of Computer Science and Engineering, Sarala Birla University, Ranchi, India. She is currently working as an Associate Professor, Department of Computer Science and IT, Arka Jain University, Jamshedpur, Jharkhand, India. She has 17 years of academic experience, with 8 years of research focus. She has successfully coordinated conferences, Faculty Development Programs (FDPs), workshops, technical seminars, and Project Exhibitions. She has authored a book titled "Cyber Security on web" and published 15+ research papers in Scopus-indexed Journals, national & international Conferences. She is a member of ABC, IETE, ISAC and IAENG. Her areas of interest include Data Mining, Data Science, IoT, Artificial Intelligence and Machine learning.

E-mail: csemeghs@gmail.com