ISSN 2281-4299

Dipartimento di Ingegneria informatica automatica e gestionale Antonio Ruberti

Sapienza
Università di Roma

Analyzing and Mitigating Issues in Radial Axes Plots for High-Dimensional Data Visualization

Mauro Ficorella
Giuseppe Santucci
Simone Lenti

Technical Report n. 02, 2023

Analyzing and Mitigating Issues in Radial Axes Plots for High-Dimensional Data Visualization

Mauro Ficorella[1] Giuseppe Santucci[1,2] Simone Lenti[1,2]

[1]University of Rome ''La Sapienza'', Via Ariosto 25, Rome 00185, Italy
[2]Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG)

Abstract

The analysis of high-dimensional data often requires visualizations that can effectively represent multidimensional information on a two-dimensional plane. Various techniques have been proposed, such as polyline mapping or radial axes visualizations. However, the latter can be counterintuitive and challenging to interpret, leading to misplaced points and false clutter.
One challenge is the difficulty users face in understanding the meaning of plotted points and visualizing them effectively. The presence of points may also necessitate the use of selection mechanisms, like squares or rectangles, which may not be suitable for all visualizations.
To address these challenges, this work proposes a novel approach that introduces representative point. The representative point helps users comprehend the true meaning of a point in its specific position, providing additional context and aiding interpretation. By using a trapezoidal selector coupled with a threshold selection mechanism, users can select and visualize only the points of interest based on their placement accuracy. This approach helps users focus on meaningful data points and reduces misleading clutter.
The proposed methodology not only aims to solve the issues associated with RadViz, but also seeks to offer a generalizable approach to other high-dimensional visualizations. By providing a systematic solution to these challenges, this work contributes to improving the interpretation and usability of visualizations for high-dimensional data.
As an illustrative case study, the paper conducts a detailed analysis of RadViz, demonstrating how the proposed approach can mitigate the aforementioned issues. The study showcases the effectiveness of the representative point concept and the trapezoidal selector with a threshold selection mechanism.
In conclusion, this work proposes a novel approach to address the challenges faced in visualizing high-dimensional data. By introducing representative points and suitable selection mechanisms, the proposed approach enhances the understanding and interpretation of plotted points, reducing clutter and improving visualization accuracy. This methodology can be applied not only to RadViz but also to other high-dimensional visualization techniques, providing a generalizable solution for improved data analysis.

Keywords: Radial axes visualizations, RadViz, High-dimensional data

1 Introduction

In the visual communication of data properties, the choice of representation format plays a crucial role in conveying accurate and comprehensive information. For instance, there are two primary categories of visualization methods: line-based and point-based approaches. The former entails representing data elements using lines or polylines, while the latter employs points to represent the data elements.
In this paper, our emphasis will be on point-based visualizations. This choice stems from their superior suitability for direct user interaction.

In the simplest case, data coming from datasets with 2 or 3 dimensions can be represented using a *direct representation* through Cartesian coordinates system, without any transformations or projections.
As the number of dimensions continues to increase (5, 6), additional visual coordinates such as size or color of the point can initially be employed to represent the additional dimensions of the data.
However, once this limit is surpassed, it becomes necessary to employ other techniques.

In such cases, one approach could be to take the multidimensional space and simply use Cartesian projection to project it onto a two-dimensional space.
On the other hand, another useful approach for representing data variability is to alter the original space in such a way that the two-dimensional representation conveys more information. This can be achieved through *dimensionality reduction* techniques such as PCA, MDS and t-SNE, transforming the n-dimensional points into a two-dimensional space.

However, two-dimensional projections and dimensionality reduction techniques, in any case, partially or completely lose the connection with the dimensions of the dataset or with its vast majority.
For example, two-dimensional projection and PCA somehow maintain the two coordinates, but only those and not others. Even worse for MDS and t-SNE, where the coordinates are completely lost.

On the other hand, when a connection with the original values (the original axes of the data) is important for the analysis, it is convenient to use radial axes visualizations in which there is an axis for each dimension and each axis is arranged radially around a common central point.

This work systematically analyzes the weaknesses of various visualizations methods, such as the ones mentioned just above, and identifies mitigation strategies that improve their effectiveness.
As a case study, the work focuses on a detailed analysis of the RadViz visualization technique, implementing some of the aforementioned mitigations in a practical experimental environment.

The remaining of this paper is organized as follows. Section 2 provides an overview of data visualizations, covering high-dimensional visualizations, and in particular Radial Axes Visualizations, explaining their meaning, theoretical framework, problems, and key terminology. Section 3 analyzes issues impacting the above mentioned visualizations, categorizing them and suggesting solutions. Section 4 presents a case study on RadViz, focusing on representative points and trapezoidal selectors to enhance the visualization's effectiveness. This section also includes a validation comparing the selector with the traditional squared selector. Section 5 finally concludes the work and discusses future directions.

1.1 Paper Contribution

After conducting an in-depth analysis of the state of the art of high-dimensional visualizations, particularly the radial axes visualizations, the contributions of this paper can be summarized as follows:

1. Detailed analysis, in-depth study of all possible problems and their systematization
2. Ideation and systematization of possible solutions to these problems
3. Demonstrative case study on RadViz

2 Background and related work

High dimensional data visualizations refer to techniques used to represent and explore data that have more than three dimensions. In such visualizations, complex datasets with many dimensions are mapped onto a two-dimensional or three-dimensional space in order to reveal patterns and relationships that might not be obvious from the raw data.

Talking about point-based visualizations, it is a common misconception that adding attributes such as colors or sizes to 2-dimensional scatterplots can adequately represent data in higher dimensions beyond three. While this approach may suffice for up to 5-6 dimensions, managing more than that number of dimensions becomes increasingly inefficient. Therefore, alternative methods are required to analyze and visualize data in higher dimensions.

This type of high-dimensional data visualizations can take many forms, including classical cartesian projection, scatterplot or dimensionality reduction techniques such as principal component analysis (PCA) [25, 18, 29], multidimensional scaling (MDS) [4, 6] or t-distributed stochastic neighbor embedding (t-SNE) [27, 28], scatterplot matrices [5, 10].

When discussing techniques for dimensionality reduction, it is necessary to make a distinction between those that employ a classical projection and those that utilize an actual mapping. In general, it's important to note that, going from n dimensions to 2/3, there is *always* a loss of information.

Projection is a way to reduce the number of dimensions in a dataset by keeping some dimensions and dropping others. This is done by a mathematical transformation that maps a high-dimensional space onto a lower-dimensional one. The important thing is that the dimensions that are kept are the original ones, not new ones that are derived from them. The most commonly used technique for this is called Principal Component Analysis (PCA).

In this case, PCA is also considered because although it creates new n dimensions from the original n dimensions through a reversible transformation matrix in order to maximize points' variance, then it selects only 2/3 of them with the maximum variance.

On the other hand, *mapping* refers to a technique that transforms high-dimensional data into a lower-dimensional space while keeping important information intact. This technique takes a high-dimensional vector as input and outputs a lower-dimensional representation of the data, where the dimensionality is much smaller. The purpose of this mapping is to preserve the most essential features of the data, like relationships between points and patterns of variation, while reducing the dimensionality to a manageable level.

Mapping can also be divided in two types: *linear* and *nonlinear*.

Linear mapping techniques, such as Principal Component Analysis (PCA), use a linear transformation to project the data onto a lower-dimensional subspace, where the new basis vectors are the eigenvectors of the covariance matrix of the data. This linear mapping is invertible, meaning that the transformation can be reverted to the original values.

Nonlinear mapping techniques, such as t-Distributed Stochastic Neighbor Embedding (t-SNE), use a nonlinear function to transform the data into a lower-dimensional space, where the new coordinate system is learned from the data itself. Unlike linear mapping, nonlinear mapping is

non-invertible, meaning that it's not possible to reverse the transformation.
As previously discussed, it can be noted that PCA falls into both linear projection and mapping techniques. This is because, through mapping, it transforms the original n-dimensional data into a new n-dimensional space, and then, it employs projection to project the data elements onto 2/3 dimensions from these new dimensions, rather than the original ones.

Moreover, among other ways to represent high-dimensional data on a two dimensional plane, there are Radial Axes Visualizations.
Radial axes visualizations [22, 8, 9] are a type of projection method used to represent multidimensional data samples as points on a two-dimensional plane. This approach relies on a set of axis vectors, each associated with a data dimensions, arranged radially around a central point. Moreover each axis is spaced at an angle that varies depending on the number of dimensions being represented.
The advantage of radial axes plots is that they provide a way to visualize multidimensional data in a compact and intuitive format. By representing each dimension on a separate axis, these plots allow for easy comparison between them and identification of patterns or relationships within the data.
However, it is important to note that radial axes plots can also have some limitations. For example, they may not be suitable for data sets with a large number of variables, as the resulting plot can become overcrowded and difficult to read. Additionally, the choice of axis scales and ranges can have a significant impact on the resulting plot, so careful consideration should be given to these factors during the plotting process.

Additionally, a significant challenge in radial axis plots is the difficulty in accurately approximating high-dimensional data values within the two-dimensional visualization. This issue can hamper the ability to search for data with specific characteristics, analyze the most common data values within clusters, inspect outliers, and perform other important data analysis tasks.
To address these challenges, various solutions have been proposed. These include algorithms for automatic axis placement, methods for variable normalization and scaling and approaches for interactive exploration of high-dimensional data.
Thus, developing effective solutions to these challenges is crucial for improving the utility of radial axis visualizations for high-dimensional data.

In this section, a concise introduction will be provided for four types of radial axes visualizations, which have been derived from the existing literature: Star Coordinates, Principal Component Biplot, Adaptable Radial Axes Plot and RadViz.

Star Coordinates (SC) [19, 20, 23], shown in Figure 1, is a visualization technique that plots data points in a coordinate system where each dimension of the data is represented by a coordinate axis. By connecting the points to a central origin, a star-shaped pattern is formed, hence the name "star coordinates".

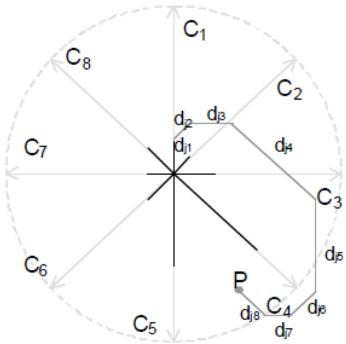

Figure 1: Star Coordinates example with 8 dimensions

In particular, SC aims to provide a multidimensional visualization that is easy for users to understand, in order to support them in the initial stages of their tasks to understand the data and gain insights. This approach assumes that once users have a good understanding of the data, they will know where to look for numerical details to perform further analysis. For this reason, great importance is given to simple and uniform visualizations across all dimensions. Additionally, it can be used for multifactorial analysis, which involves examining relationships between multiple variables. By plotting multiple sets of data using the same coordinate system, relationships between different variables can be visualized and analyzed.
Treating each dimension uniformly, this approach uses a coarse representation of data, limiting its accuracy and usefulness when seeking detailed information and making it difficult to estimate original attribute values visually. Furthermore, the normalization of data to fit into a two-dimensional (2D) plane results in the loss of numerical analysis capabilities. Additionally, the position of a data entry on SC does not provide any unique information due to multiple ways of representing the same point, making very difficult to estimate high-dimensional values associated with an embedded point by simply visualizing the plot [9, 24].
This ambiguity is known as the "depth perception problem". In SC, the problem arises because the 2D plot cannot fully represent the higher-dimensional space.
The authors [19] aimed to mitigate these issues by providing interactive methods for manipulating data such as scaling and rotation transformations, allowing users to select any configuration of axes, and therefore to generate plots associated with any linear mapping, in order to quickly gain insights during initial analysis [21, 23]. For this reason, SC is considered an highly interactive visualization.

Principal Component Biplot (PCB) [11, 14, 12, 13] are a type of graphical representation of data obtained from Principal Component Analysis (PCA).

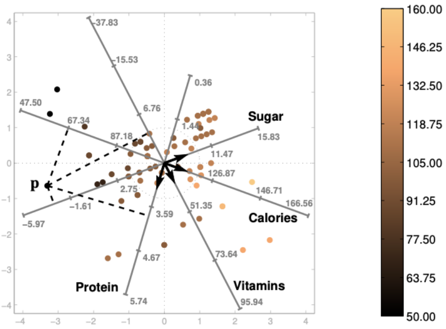

Figure 2: Principal Component Biplot

In a PCB (Figure 2, the data points are projected onto a two-dimensional plane, and the

axes on this plane represent the principal components of the data. Specifically, PCBs can be understood as a result of solving an optimization problem where the goal is to find a set of axis vectors and embedded points that allow for optimal estimation of attribute values by projecting points orthogonally onto calibrated axes. The resulting visualization can help to highlight the relationships between the different variables in the data.
PCBs can be used to optimize different properties of the data visualization, depending on the scaling factors used. For example, by preserving distances between samples, or by approximating correlations between variables. Also, calibration of PCB is an important step in creating a useful and interpretable visualization of the data. It allows for easy estimation of attribute values and comparison of different attributes, which can help users to gain insight into the underlying patterns and relationships in the data. Finally, centering the data facilitates labeling of the axes, and can improve the accuracy of estimates of the original data attributes.
However, unlike Star Coordinates, which allow users to update the axis vectors interactively, PCBs are a static visualization that does not allow for interactive updates such as selecting arbitrary directions for the axis vectors associated with the variables since they are constrained in order to maximize attribute estimation accuracy. However, this limits the user's ability to directly manipulate data points and impedes the comprehensive analysis of the visualization, thus hindering the ability to derive meaningful insights.

Adaptable Radial Axes Plots (ARAP) [26] is an innovative technique that integrates the advantages of Star Coordinates and Principal Component Biplots.

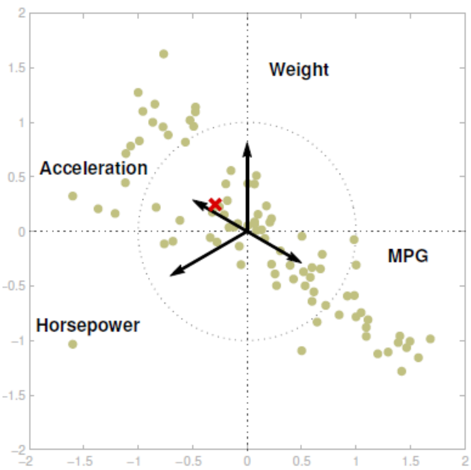

Figure 3: Principal Component Biplot

ARAP (Figure 3) enables users to update axis vectors interactively while optimally estimating attribute values through labeled axes. The projections are defined using convex optimization problems that extend the optimization problem of PCB. Moreover, ARAP minimizes the projection errors users may make when approximating high-dimensional attribute values by projecting mapped points orthogonally onto the axes.
ARAP offers users a diverse set of projections to explore, and it facilitates various data analysis tasks. The method enables users to modify the axis vectors freely and interactively, thus bridging the gap between Star Coordinates and Principal Component Biplot. The ARAP technique also allows users to explore nonlinear projections by using alternative norms, providing additional perspectives on the data. Additionally, ARAP includes the ability to incorporate weights and constraints in the optimization problem, enabling users to sort the plotted points according to one attribute. ARAP is a highly flexible technique that complements and enhances current radial methods for data analysis, extending their capabilities in various ways.
It is important to highlight that ARAP is primarily intended for exploratory purposes and for

obtaining an overview of the data. This is because the estimates obtained through projections onto the axes are only approximations of the true attribute values. Therefore, this technique should be regarded as an alternative in tasks where optimal estimation accuracy is not crucial, and where user interaction is essential to explore the data effectively.

RadViz [15, 1, 2, 16], shown in Figure 4, is a data visualization technique that allows for the representation of the n dimensions as points arranged equidistantly around the circumference of a circle.

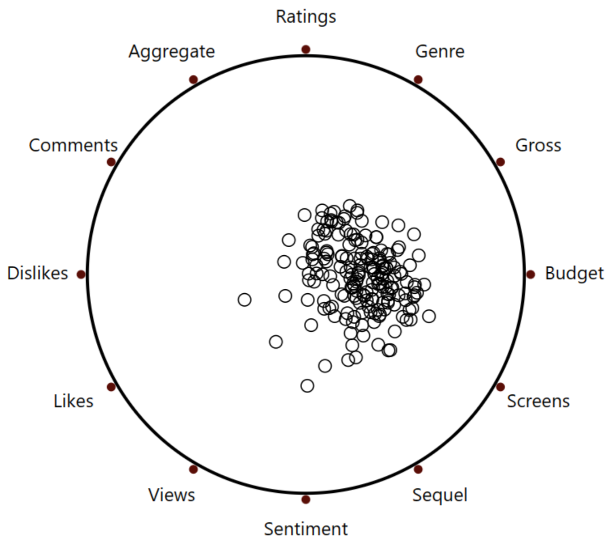

Figure 4: RadViz

The radial positions (x and y coordinates) of the data points are determined by a set of n springs, with one end of each spring attached to a perimeter point and the other end attached to the corresponding data point.
Spring constants are mathematical parameters that can be used to represent the strength of the relationship between two points in a system. The spring constant of each spring is determined by the value of the i-th dimension of the fixed point. All data point values are normalized to ensure that they fall within a standardized range, typically between 0 and 1. The data point is then displayed at a location where the sum of the spring forces acting on it is equal to zero, indicating a state of equilibrium in the system.
The authors have found some inherent problems that can affect the quality of the representation in RadViz. *Clumping towards the center* problem arises when points accumulate near the center of the RadViz plot [7]. This is due to the fact that, as the number of dimensions increases, it is more likely that points have approximately equal (normalized) values for the dimensions that lie on opposite sides of the circumference. This causes them to be pushed towards the center of the plot. *Many-to-one plotting* occurs when points that are far apart in the multidimensional space are represented very close to each other in the RadViz plot. This can happen when multiple dimensions have similar values for different points, causing them to be plotted in close proximity to each other. *Effectiveness of point position* means that the quality of the RadViz plot can be highly influenced by the order in which the dimensions are arranged. Some dispositions of dimensions may be more effective than others in representing the normalized values of the points. Therefore, it is important to choose the most effective disposition to avoid misrepresenting the data. Moreover, RadViz employs a dimension-balancing technique to represent high-dimensional data in a 2D space. This technique results in a projection of the data that reflects the *proportional* magnitudes of each dimension. However, the 2D projection does not preserve the Euclidean distances in the original $\mathbb{R}^n$ space, and the actual values of the

data in each dimension cannot be directly inferred from the plot.
The equilibrium position of a point in RadViz is determined by the balance of the values across all dimensions. Therefore, the relative magnitudes of the dimensions and their relationships to each other play a crucial role in the resulting plot. Consequently, RadViz is a useful tool for exploring the relative proportions and relationships among high-dimensional data points, but not for precise quantitative analysis of the original data values.

3 Problems systematization

After conducting an analysis of various existing high-dimensional visualizations, in particular radial axes visualizations, an examination of the individual problems affecting them was carried out.

Regarding Dimensionality reduction techniques, the distances between pairs of points do not accurately reflect the actual dimensions of the data; moreover, in PCA, new dimensions are first created and then only two of them are projected into the plane, and in t-SNE and MDS, there are no direct links to the original dimensions since they are not represented on the plane.

In radial axes visualizations, although all dimensions are present, information loss occurs when n-dimensional points are represented on a 2D plane; furthermore, it is difficult to read a point when there are a large number of dimensions.
In Star Coordinates, normalization results in a loss of numerical analysis, making it challenging to understand the meaning of a point and the position of a point does not provide unique information since there may be multiple ways to represent a point in space.

In Principal Component Biplot, since the points and dimensions are constrained to achieve maximum estimation accuracy, there is a tradeoff, and there is no interaction between the user and points.

In Adaptable Radial Axes Plot, the points are obtained through estimates with projection and are therefore approximations of the actual attribute values of the point. Consequently, it is not well-suited for a precise analysis of the meaning of a point in terms of its attributes.

Finally, in RadViz, clumping towards the center and many-to-one plotting problems occur. Moreover, the effectiveness of point position can be misleading and also, RadViz is useful for analyzing attribute proportions, but not for a precise analysis of their numerical values.

Subsequently, analogies and potential differences were identified between these individual problems, and it was determined whether certain problems affected multiple visualizations. This led to the development of a systematization of the problems and allowed for a general analysis.

3.1 Curse of dimensionality

The curse of dimensionality [3] is a phenomenon that occurs when working with high-dimensional data, where the number of dimensions is very large. It refers to the fact that as the number of dimensions increases, the amount of data required to adequately represent the data increases exponentially, making it difficult to process and analyze the data effectively.
In other words, data with increasing number of dimensions can introduce the main problem of the *difficult reading of a point* users can face in interpreting it.
An example could be scatterplot matrices, in which the number of possible pairwise scatterplots increases rapidly with the number of dimensions, making it difficult to interpret them for

high-dimensional data since they need to be synchronized between each other in order to offer a meaningful interpretation.
In general, this problem can be found also in radial axes visualizations: a high number of dimensions means a high number of axes, closer to each other, leading to a more a more difficult analysis by users. An example of this can be found in RadViz visualization: as the number of dimensions increases (i.e. 30 dimensions), it becomes more difficult to interpret and understand the proportion between all attributes of a data point.

3.1.1 Mitigation

To address this challenge, an auxiliary visualization such as a bar chart or a line-based visualization can be utilized. In such a visualization, each line represents the actual value or proportion for an individual dimension, each of them ordered in the same way as the dimensions axes in the visualization. By providing a clear and rapid visual understanding of the dimensions' values or proportions for a given point, this auxiliary visualization enables the user to avoid the manual and confusing process of comparing dimensions one by one, particularly in cases where the number of dimensions is high.

3.2 Distortion

All dimensions are present through all the axes, but they are distorted, such as in the case of radial axes visualizations (i.e., RadViz, Star Coordinates).
In this case, despite retaining all original dimensions, a distortion occurs, which means that the distance between data points in the visualization is not proportional to their actual distance in the original high-dimensional space. Consequently, this may lead to incorrect interpretations of the relationships between data points and ultimately result in erroneous conclusions.

3.2.1 Mitigation

In cases of distortion in radial axes visualizations, the concept of error metric can be applied to indicate to the user which points should not be considered in their analysis. This error metric can be based on the color coding of each point in the visualization, making it easily understandable visually. This way, despite the distortion, the user can still try to perform some data analysis considering only the "best" data points.

3.3 Missing original dimensions

In the context of point-based high-dimensional data visualizations, it would be good to retain all the original dimensions without loss of information, but in practice, this is not possible. There are several reasons that may contribute to the situation at hand.
One possible reason is the creation of a new set of axes, which entirely replaces the original ones. This occurs, for example, in Principal Component Analysis (PCA), where n new dimensions are created from the original ones, and then two new axes are formed as a linear combination of these new dimensions. Consequently, this can pose difficulties in identifying the original attributes of a data point.
Another instance is when normalization is applied, leading to dimensions that are identical to the previous ones but with different scales, which are then considered distinct.
Another reason could be attributed to the preservation of some original dimensions, while others are eliminated, which may lead to challenges in comprehending and analyzing the data when users intend to compare values of different attributes present in the dataset but are no longer visible in the current visualization.
Another possible case involves dimensionality reduction techniques such as MDS and t-SNE, where the axes and, consequently, the references to dimensions disappear entirely following the

transformation from n to 2 dimensions.
In general, this problem results in a limited understanding of a data point with respect to only the attributes related to the two or three axes plotted on the plane, while the values of the other attributes are completely lost, making it challenging to interpret the meaning of the dimensions to understand the relationships between the variables in the data.

3.3.1 Mitigation

In cases where some or all dimensions are missing, a possible solution could be to use an auxiliary visualization or even a simple pop-up tooltip that allows the user to compensate for the missing dimensions by offering the corresponding values for all attributes, including those related to the missing axes, either visually or numerically. This way, even without a complete view of all the axes and dimensions for a certain data point, the user can still perform a more accurate analysis based on the values or proportions of the various attributes.

3.4 False positives and false negatives

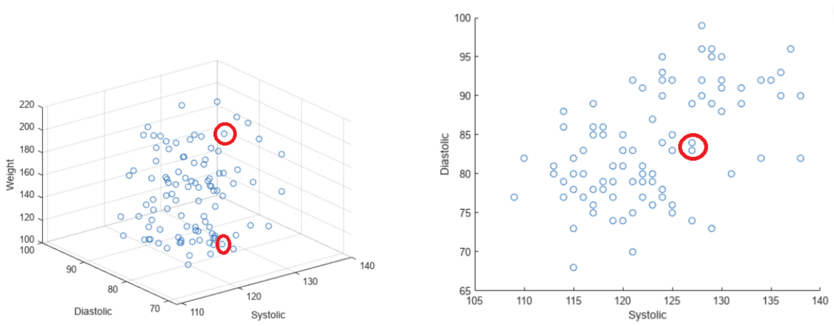

Figure 5: Example of false positive in a 3d scatterplot using PCA

The idea of false positives and false negatives is linked to the distance between two points rather than their absolute position in the plane.
During a projection, such as a Cartesian projection or PCA, false positives are introduced. This refers to the possibility of points appearing close in the lower-dimensional space while being distant from each other in the original dataset. This could result in distorted data, misrepresented pairwise distances, and lead to the creation of clusters or structures that do not accurately reflect the true underlying structure of the data in the original space. As an example, it can be seen, from Figure 5, that the two points circled in red have a different value related to the z axis (Weight), while having a very similar value related to the x and y axes (Diastolic and Systolic): this would mean, if the two point are considered only on the x and y axes, that they would appear very close to each other, but in the real dataset, they are far away from each other since they have a very different value of Weight attribute.
On the other hand, in nonlinear mappings, in addition to false positives, there is also the issue of false negatives. This refers to a situation where two data points are placed farther from each other in the lower-dimensional space than they are in the original high-dimensional space. This could result in distorted data, misrepresented pairwise distances, and loss of crucial information or structures in the data.

3.4.1 Mitigation

A potential solution to the problem of false positives and false negatives could be the establishment of a metric that determines whether a data point falls within one of these categories, as well as to what degree it does so. This metric would allow for more precise and reliable analysis

of data sets.
In this way, a user, looking at the points represented in the visualizations, could at least understand which point is represented with the wrong distance with respect to another point and therefore act accordingly to draw more accurate conclusions and insights from their analysis.
Firstly a function should be defined to calculate how much a point's distances are wrong compared to the "optimal" value, and so how much a point is a false positive or a false negative.
After defining such a function, one method could be to color the points based on a chromatic scale designed to visually signal to the user which points are represented right or wrong distances.
For example a scale ranging from green to purple, passing through orange, as the one proposed by [15], could be defined: a green point would mean that it is represented with a correct distance with respect to others or that it has a negligible amount of error in this distance, orange/red that it has a meaningful amount error in this distance, and purple that this distance is completely wrong.
Furthermore, after defining such a metric, it is also possible to use a range slider to select a "threshold" value: this would allow the user to exclude all points that have a distance error greater than the threshold selected via the slider.
This avoids the users using points with "wrong" distance with respect to others in their analysis through direct manipulation, thus preventing them from making errors in their insights.

3.5 Meaning of the position of a point

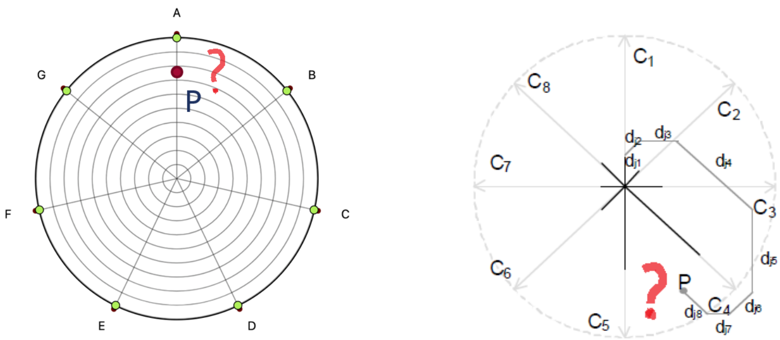

Figure 6: Single point plotted in RadViz and Star Coordinates

It can be difficult to analyze the meaning of the position of a certain data point in detail. In particular, a point, despite being correctly positioned on an axis and having its position governed correctly by that axis, could be difficult to interpret and analyze. For instance, is very challenging to explain precisely what is the meaning of points plotted in Figure 6, where, in particular, are used RadViz and Star Coordinates visualizations?
Inferring the values of the attributes of a data point represented in a high-dimensional visualization such as radial axis visualizations is not straightforward, since transformations are often applied that approximate or lose some information. Furthermore, in many of these transformations, normalization of the values of data points is often applied, which can result in a loss of numerical analysis capabilities.
Moreover, the position of a data-point on the visualization does not provide unique information, as there are multiple ways of representing the same point. This makes it very difficult in practice to estimate high-dimensional values associated with an embedded point by simply visualizing the plot. Additionally, the aforementioned normalization introduces ambiguity in the visualization, as a single data point can correspond to more than one data value.

3.5.1 Mitigation

To answer the above question, the concept of *representative point* should be defined, also taking advantage the error metric defined above. Specifically, the representative point refers to the dimensional values (or dimensional proportions) that a point in a certain position in the visualization should have, with an error equal to 0.
This concept could be combined with an auxiliary visualization, such as a pie chart or a bar chart, allowing the to visualize the dimensional attribute values or proportions that a point in a certain position should have, in such a way that he can understand what's the meaning of a point in a certain position in the plane.

3.6 Misplacement of a point

In contrast to the issues of false positives and false negatives, misplacement pertains to the erroneous absolute positioning of a given point, rather than its relative distance to other points. False positives or negatives, in fact, occur where they should be on the coordinate axes within which they are represented and refer to the distance between a point and other points with similar values on the axes within which it is represented, which may differ significantly for other missing axes.
On the other hand, misplacement concerns a singular point and denotes that, with regards to the axes utilized to assign a value to the given point, the point itself yields erroneous information. In other words, the point's position is misleading in relation to the manner in which the axes are interpreted by to infer values for the various dimensions of the point in question, leading to inaccurate conclusions and insights.
Moreover, some visualizations are prone to the issue of misleading data representation due to the order of their axes. Certain axis dispositions may be less effective than others, and if a user wishes to compare two dimensions that are distant from each other, it can be challenging, if not impossible, to do so without the ability to move them next to each other.

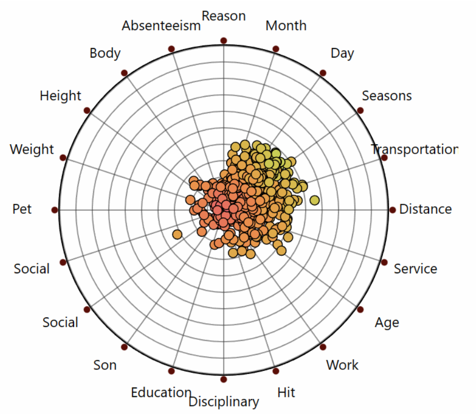

Figure 7: Example of clumping towards the center inside RadViz using a dataset with a high number of dimensions (20)

Another points' misplacement example could be RadViz, that is affected by the problem called *clumping towards the center* (as shown in Figure 7), meaning that, as the number of dimensions increases, it is more likely that points have approximately equal (normalized) values for the dimensions that lie on opposite sides of the circumference; this causes them to be pushed towards the center of the plot, making it difficult to distinguish between points that have similar values on those dimensions.

3.6.1 Mitigation 1 - Error metric

To address the misplacement issue, a potential solution is to define an error metric that quantifies the degree of error associated with the placement of each data point. The error metric would calculate the deviation between the plotted position of a data point and its actual position, based on the given attributes. This would provide a measure of how far off each data point is from where it should be, allowing for the identification of inaccurately represented points.
Once the error metric has been established, color-coding the data points based on a chromatic scale can be an effective way to visually signal to users the degree of error associated with each point.

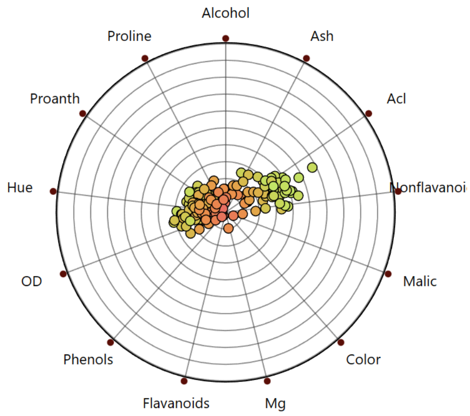

Figure 8: Example of color-coded points' error metric

For example, as can be seen in Figure 8, a scale ranging from green to purple, passing through orange, could be used. The green color could represent accurately placed data points, while the purple color could indicate completely misplaced data points. Orange or red colors could be used to represent data points with a certain degree of error, indicating that they should be approached with caution when interpreting the data.
Furthermore, adding a range slider that allows users to select a threshold for the degree of acceptable error could be helpful. This would enable users to exclude all data points that have an error greater than the selected threshold. By doing this, the visualization becomes less crowded and more manageable, allowing users to focus on the most relevant data points.

3.6.2 Mitigation 2 - Representative Point Comparison

The concept of representative point introduced in 3.5.1 is very powerful because it can also be used to offer a benchmark for displaying the dimensional attribute values that a point should have compared to those it actually has.
This can be done through another visualization that, contemporarily to the representative point values, shows also the actual values of a certain point, allowing for a visual comparison of the dimensional attribute values that a point in a certain position should have to have an error of 0 compared to those it actually has with a certain amount of error.
In this way, in addition to the advantage discussed above obtained through the error metric, which allows users to understand whether a point is in the right or wrong position, this tool also offers the possibility of understanding how much this position is wrong, greatly assisting the user in their analysis.

3.6.3 Mitigation 3 - Automatic Dimension Arrangement

One of the challenges with radial axes visualizations is that the positioning order of the axes representing the dimensions can result in ineffective data-point representations. This is because the positioning order can affect how well the data points are distributed on the visualization.
In order to address this challenge, one solution could be to use the error metric described earlier and develop an algorithm that rearranges the dimensions of the axes to minimize the positioning errors of the data points in the visualization. This algorithm could be activated either by pressing a button or automatically when the visualization is loaded.
The algorithm would work by taking into account the position of each data point and the associated error metric, and then reordering the dimensions of the axes accordingly to optimize the positioning of the data points on the visualization based on an "optimal" value. By doing this, the algorithm would ensure that the data points are more accurately represented, allowing users to make more informed decisions based on reliable information.

3.7 Direct manipulation

In high-dimensional visualizations, it can be useful for users to have a selector to directly interact with the represented data points and select a certain subset to compare them on certain attributes.

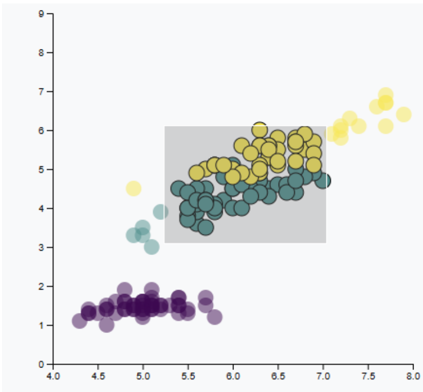

Figure 9: Rectangular selector in a scatterplot

For example, the classic rectangular/square selector used in scatterplots (as shown in Figure 9).
However, this selector, with its sides parallel to the orthogonal axes of the scatterplot, does not lend itself well to analyzing and gathering meaningful information on the spatial position of data points in other visualizations besides those with orthogonal axes.
This could lead the user to select points incorrectly, leading him to believe that some points may be correlated when they are not or for other reasons.

3.7.1 Mitigation 1 - Selector of a specific shape

When it comes to high-dimensional cases such as radial axes visualizations, it is immediately apparent that a selector of the classical squared shape is meaningless. In fact, these types of visualizations have the axes arranged radially around a central point representing the origin, each spaced from the others by an angle that varies based on the number of represented dimensions, and so they need a selector able to represent the planar properties of the point inside them.
One solution to obtaining a selector suitable for these visualizations could be to use selectors of different shapes, depending on the arrangement of the axes in the visualization under analysis. Moreover, unlike the base case of the orthogonal axes visualizations such as the Cartesian plane

where only one selector is used since there are only two axes, in the presence of radial axes visualizations where all the axes corresponding to all n dimensions are present, it would be useful to have the option of using multiple selectors, perhaps simultaneously, one for each axis pair, in order to compare multiple points across various dimension pairs at once.

3.7.2 Mitigation 2 - Manual dimension arrangement

It is common for a user to have the desire to compare data-points, for explorative purposes, using two attributes that are situated on opposite sides of the radial axis visualization.
To address this challenge, one potential solution could be to provide the user with the ability to manipulate the axes and rearrange them as they see fit. As the user moves the axes, the visualization will dynamically update to reflect the new arrangement.
By utilizing this functionality, the user can effectively compare data-points on attributes that were originally separated by a significant distance on the visualization.

3.8 Comparison matrices

Table 1: Problems matrix

Problems \ Visualizations	RadViz	Star Coordinates	Principal Component Biplot	Adaptable Radial Axes Plot	Dimensionality Reduction techniques
False positives / false negatives	No	No	No	No	Yes
Misplacement of a point	Yes	Yes	No	No	No
Difficult meaning of a point	Yes	Yes	No	No	Yes
Curse of dimensionality	Yes	Yes	Yes	Yes	No
Distortion	Yes	Yes	Yes	Yes	No
Missing original dimensions	No	No	No	No	Yes
Weak direct manipulation	Yes	No	Yes	No	No

Table 2: Solutions matrix

Problems \ Solutions	Error metric	Representative point	Meaningful selector	Dimension arrangement	Auxiliary visualization with original values
False positives / false negatives	Yes	No	No	No	No
Misplacement of a point	Yes	Yes	No	Yes	Yes
Difficult meaning of a point	No	Yes	No	No	No
Curse of dimensionality	No	No	No	No	Yes
Distortion	Yes	No	No	No	No
Missing original dimensions	No	No	No	No	Yes
Weak direct manipulation	No	No	Yes	Yes	No

After having analyzed the problems that affect high-dimensional visualizations, in particular radial axes ones, two matrix have been created, one (Table 1) showing problems related to visualizations and one (Table 2) showing solutions related to each problem, all of that in a systematized way.
Thanks to these matrices, after systematizing the existing problems in high-dimensional visualizations, and in particular in radial axes visualizations, it can be noted that the RadViz is the visualization with more problems.

For this reason, in order to demonstrate the utility of some of the proposed mitigations, the goal of this work is to tackle, through the case study of RadViz, the problems of misplacing points and selectors made in a certain way to maintain the spatial properties of the points on the representation.

4 Case Study

In the context of the high-dimensional visualizations, and in particular radial axes ones, as described in the previous Section, there are certain issues related to the accurate positioning of a point p. Moreover, it could be useful to have a selector of a meaningful shape in order to help users with direct manipulation and analysis of the data points.

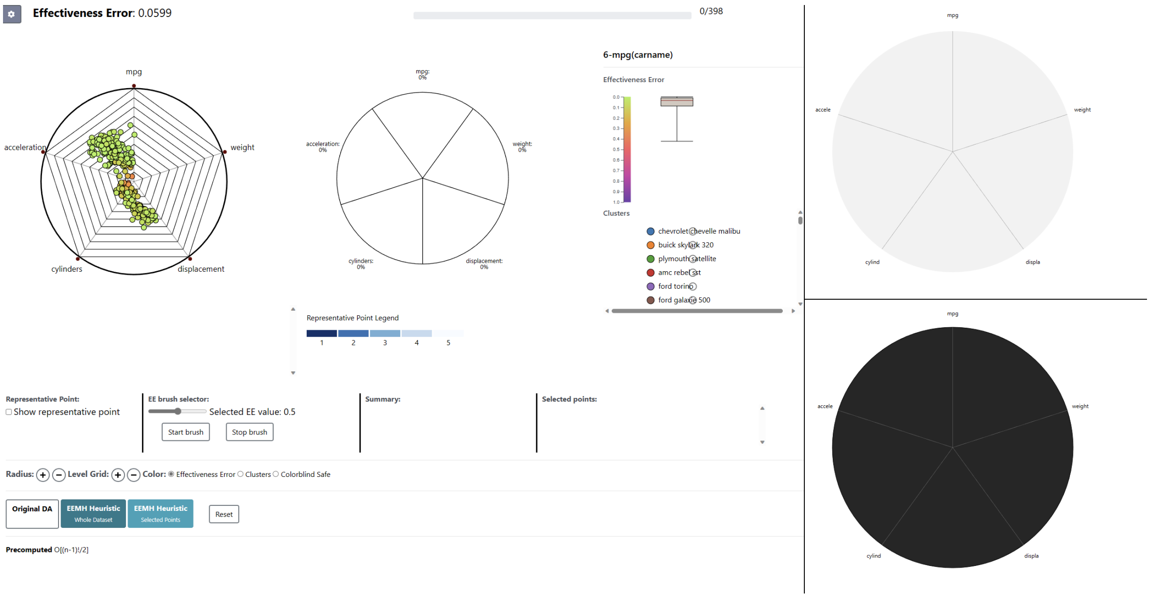

Figure 10: New RadViz Prototype

The present chapter will apply some general solutions proposed in the systematization from previous Section in the specific RadViz radial axes visualization, in order to mitigate some of its problems.
In particular, the different solutions and experiments are proposed by utilizing an experimental prototype, as illustrated in the Figure 10. This prototype was developed based on the work carried out by [1, 2]. Furthermore, from the same authors, it has been possible to adopt color coding of points within RadViz based on the error metric presented in [1].
Starting from this prototype, some significant additions and modifications have been proposed. More in details, as can be seen from the figure, even the RadViz itself has been slightly modified. More specifically the internal grid has been altered from a concentric circular structure to a reticular structure formed by a regular polygon inscribed in the circumference with vertices coinciding with the anchors, and concentric polygons with decreasing sides' length. The idea stems from the following reasoning: first of all, there is a external polygon inscribed in the circumference with vertices coinciding with the anchors because it delimits the area where points can be plotted inside the RadViz; moreover, just as a user on a Cartesian plane can have a grid of squares to facilitate the reading of the coordinates of a point, a grid made in the above mentioned way has been used here to help the user understand the proportion between dimensions of a certain point.
Regarding the dataset used in this case study, the MPG car dataset[1] has been chosen.

[1] http://archive.ics.uci.edu/ml

4.1 Representative Point

In RadViz, accurate positioning of a point p is challenging due to the non-linear transformation used to map n-dimensional points to a 2D space. This transformation introduces a certain degree of uncertainty in the position of p and may result in an inaccurate representation of its true location in the original n-dimensional space, meaning that the values representing dimensions' influence on each attribute value could be not correct.
Given a set of n anchors $A = (a_1, a_2, \ldots, a_n)$, the main goal of the representative point is to provide a tool for obtaining the correct proportion values for each anchor a_i for a certain selected point, indicating how much that anchor a_i influences each attribute of the point. This allows the viewer to understand the actual predominance proportions between the various attributes for that point.
The representative point is used to discern the proportional ratio between attributes for a given data point, assuming an *Effectiveness Error* that approaches zero. In other words, the representative point concept answers the following question: "in these coordinates, which point should be located there?"
To achieve this, as can be seen in Figure 11, the solution presented in this work also includes an ancillary visualization, positioned conveniently next to the RadViz chart.

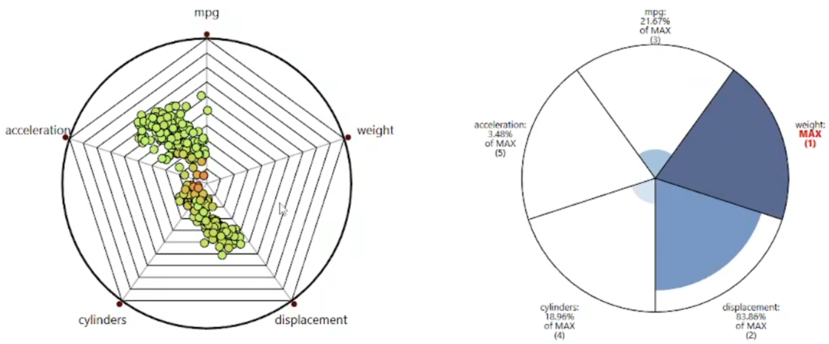

Figure 11: Pie chart auxiliary visualization

This is a pie chart that displays the percentage value of how much each anchor a_i, with respect to the most predominant anchor (which has the MAX value in red color to help users identifying it), influences each attribute of the point when the mouse is hovered over it, based on the values calculated by the representative point for that point.
Moreover, as depicted in the Figure 12, orange lines have been implemented to represent the current attribute predominance values of the mouse-selected point in the RadViz visualization, with a certain degree of effectiveness error.

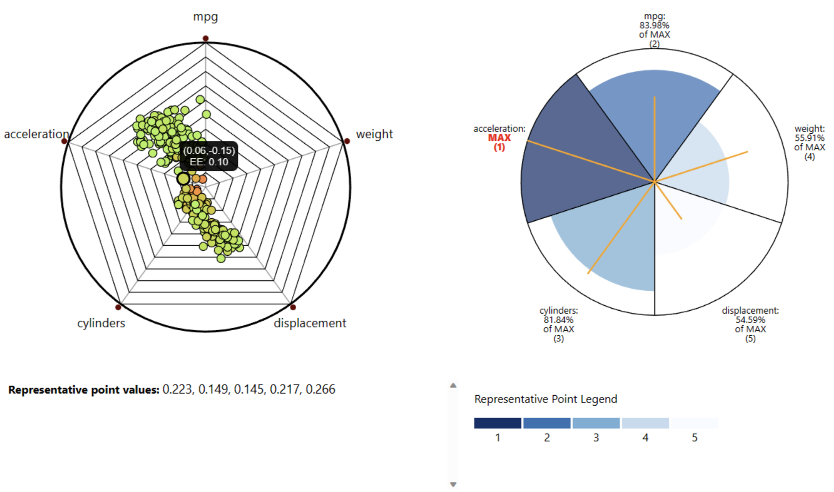

Figure 12: Representative Point comparison

This allows for the determination of the extent to which the attribute predominance values of the selected point deviate from the correct predominance values obtained through the representative point with an effectiveness error of 0. In other words, the orange lines assist the user in determining the degree to which a point is misleadingly positioned on the RadViz visualization. Also, as can be seen in the same figure, there is also a section in the prototype in which are shown representative point values numerically, normalized between 0 and 1.

4.1.1 Formalization

To begin with, it is important to recall the concept of *Effectiveness Error* (EE). In essence, the EE is considered zero when the order of distances from the anchors of a given point is the exact opposite of the order of values of its representative point. Mathematically, given the Euclidean distances $\Delta_1^{a,p}, \ldots, \Delta_n^{a,p}$ between a point $p = (v_1^p, \ldots, v_n^p)$ and the n Dimension Anchors, the whole thing can be translated in the following way:

$$\Delta_i^{a,p} < \Delta_j^{a,p} \implies v_i^p > v_j^p \tag{1}$$

In general, to calculate the representative point values, matrices are used to solve a system of linear equations.
First of all, a $3 \times n$ matrix A is created: given the number of Dimension Anchors n, a number of columns equal to n and 3 rows are created, where the first is x, the second is y, and the third is made up of all 1s, meaning that the sum of all values must be 1 (normalization). In particular, in the x row, all the sines of the various angles indicating the positions of the Dimension Anchors on the circumference are inserted, cell by cell; in the y row, all the cosines of the various angles indicating the positions of the Dimension Anchors on the circumference are inserted, cell by cell.
Moreover, a 1×3 matrix B is filled where, given a point in Cartesian coordinates $pxy = (x, y)$ whose representative point must be calculated, the x coordinate is in the first cell, the y coordinate is in the second cell, and the third cell contains the value 1 for normalization.
These matrices are used to solve a system of linear equations by multiplication $Ax = B$.
As an illustration, consider a point in a 4-dimensional space whose 2-dimensional coordinates are $(x, y) = (0.2, 0.4)$. In this scenario, the initial system can be formulated as follows:

$$\begin{cases} sin(0 \cdot p_a) + sin(\frac{\pi}{2} \cdot p_b) + sin(\pi \cdot p_c) + sin(\frac{3\pi}{2} \cdot p_d) = 0.2 \\ cos(0 \cdot p_a) + cos(\frac{\pi}{2} \cdot p_b) + cos(\pi \cdot p_c) + cos(\frac{3\pi}{2} \cdot p_d) = 0.4 \\ p_a + p_b + p_c + p_d = 1 \end{cases} \tag{2}$$

Then through the above mentioned matrices A and B the whole system can be rewritten as follows:

$$A = \begin{bmatrix} sin(0) & sin(\frac{\pi}{2}) & sin(\pi) & sin(\frac{3\pi}{2}) \\ cos(0) & cos(\frac{\pi}{2}) & cos(\pi) & cos(\frac{3\pi}{2}) \\ 1 & 1 & 1 & 1 \end{bmatrix},\ B = \begin{bmatrix} 0.2 \\ 0.4 \\ 1 \end{bmatrix} \tag{3}$$

In this case, A is a rectangular matrix. When A is a rectangular matrix (i.e., it has more columns than rows or more rows than columns), the equation $Ax = B$ may not have a unique solution. This system is called *underdetermined*. For underdetermined systems of linear equations, i.e., systems with fewer equations than unknowns, there are infinitely many solutions that satisfy the equations. In this case, there is a whole set of solutions that satisfy it. To obtain a unique solution for underdetermined systems, typically can be used methods that introduce additional constraints or assumptions. One common approach is to find the solution with the smallest norm, such as the *Euclidean norm*.

In particular, have been used the *least-squares method* to solve a linear matrix equation. Specifically, it solves the equation $Ax = B$ where A is the *coefficient matrix* and B is the *constant vector*, by finding the values of x that minimize the Euclidean norm $\|B - Ax\|_2$.

In this case, in the presence of a rectangular matrix, this method returns the solution that has the *smallest norm* (i.e., the smallest length).

However, the resulting solution may not be unique, and additional constraints or assumptions may be necessary to obtain a unique solution with EE 0.

To cope with this, optimization was used, specifically through minimization. In particular, first of all, a set of constraints has been defined: these constraints verify the order of distances between values of an n-dimensional point and the dimension anchors; moreover an objective function that needs to be minimized have been defined. In this case, it is the norm of the difference between Ax and B. Then, a bounds list have been generated to define the lower and upper bounds for each element of the solution. In this case, each element must be between 0 and 1. Finally the minimization function is employed with the objective function, the initial guess (which represents the result of the least squares method mentioned earlier), the A and B matrices, the bounds list, and the constraints. The function returns the minimum value of the objective function, subject to the constraints.

4.2 Trapezoidal selector

Effective data visualization involves not only the selection of appropriate chart types and visual elements but also the use of selectors, which allow viewers to interact with the data and explore it in greater depth. Rather than presenting a static view of the data, selectors allow users to interact with it in real-time, enabling them to compare different subsets of the data and investigate relationships between variables. This can lead to the discovery of new patterns or insights that might not have been apparent in a static visualization.

All the above mentioned benefits can be especially useful when dealing with large or complex datasets that may contain many variables or dimensions. By using selectors, users can quickly focus their attention on the data that is most relevant to their analysis.

The use of selectors in RadViz is not as immediate as it happens in the two dimensional Cartesian coordinates system. In contrast to this reference system, where the selection operation involves selecting points within a value range on two orthogonal axes, the x and y axes respectively, the RadViz visualization presents a multidimensional representation with a number of radial axes equal to the number of dimensions for each data point, bigger than 2. Consequently, designing a selector for RadViz is a challenging task due to the high number of dimensions and attributes. Traditional rectangular or squared selectors, commonly used in two dimensional scatterplots or Cartesian coordinates systems as their sides are parallel to the orthogonal axes and can express the correlation between the position of the points within the selector, may not

be suitable for RadViz as they fail to capture the positional information of the data points on that visualization.
In this regard, in this paper is proposed a novel *isosceles trapezoidal selector* (Figure 13) with oblique sides parallel to the axes that enclose it.

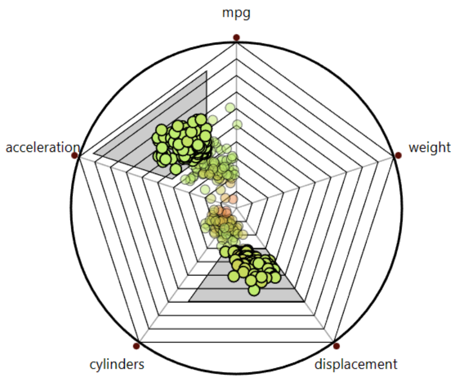

Figure 13: RadViz trapezoidal selector

This approach allows capturing the predominance of two dimensions over a point located within the selector, within the space delimited by those dimensions.
This solution also allows for multiple selectors to be active simultaneously, enabling the selection of points across multiple attribute pairs, as RadViz is a multidimensional visualization with more than two attributes.
Moreover, in the experimental prototype, are also indicated the number of points belonging to different clusters (if present in the dataset) based on the selected points from various attribute pairs. Additionally, the total number of selected points and their average effectiveness error is calculated. This allows users to better understand the selected points compared to the situation without a selector, enabling them to gain more insights from the data.

Furthermore, is proposed an additional slider for selecting the effectiveness-error objective threshold. By selecting specific data points, the user can better identify those with a lower effectiveness error threshold. This feature provides an intuitive means for understanding which data points occupy the correct position in relation to the selected attributes, thereby eliminating virtually all points that are not in the appropriate position, improving readability of the visualization and isolating only points that occupy the correct position related to the selected desired effectiveness error.

4.2.1 Formalization

The trapezoidal selector, situated between two anchors a and b, facilitates the identification of points that exhibit a greater degree of association with these two anchors in comparison to others. To compute the degree of predominance, one can utilize the representative point, specifically by calculating the ratio between the representative point's value for a given anchor and the values associated with other anchors. This ratio provides an estimate of the anchor's predominance in comparison to others.
Specifically, the trapezoidal selector spans an *exponential scale* between 1 (a ratio equal to 1 denotes that all anchors hold equal predominance at the origin of the axes) and ∞ (a ratio equal to ∞ indicates that the two anchors a and b exert complete predominance over others, and this takes place at the chord of the circumference connecting the two anchors, called *infinite predominance line*).

A very important property of this selector is the fact that it has the oblique sides that vary in inclination according to the axes that contain them, implying that maintaining the same inclination while varying the pair of axes would render the selector meaningless, despite its initial effectiveness. Therefore the oblique sides of the selector should be aligned with straight lines that are parallel to each pair of axes.
In Figure 14, 15 it's possible to verify, by utilizing the concept of representative point, the utility of shifting the selector throughout the quadrant, so that points with varying degrees of influence from the two attributes (Acceleration and MPG), within the selector is contained, can be selected.

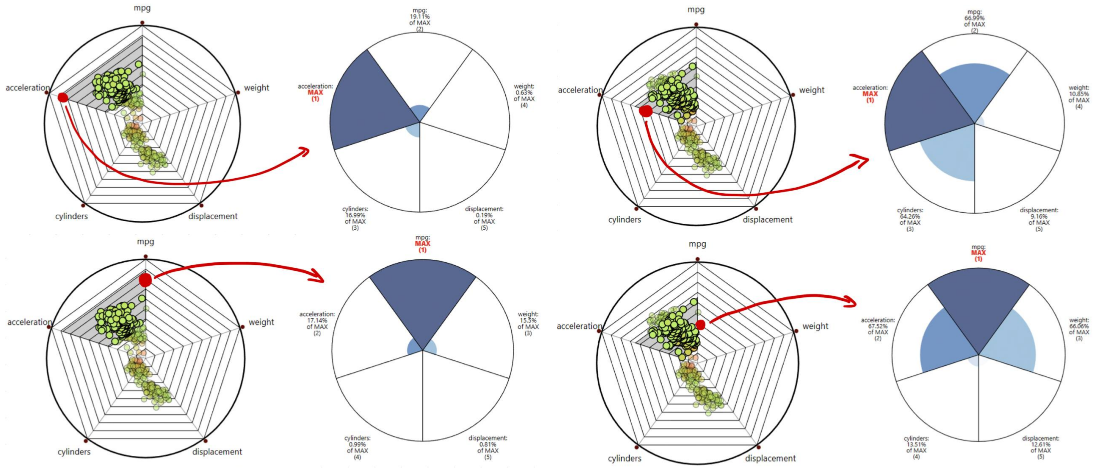

Figure 14: Trapezoidal brush vertically resized

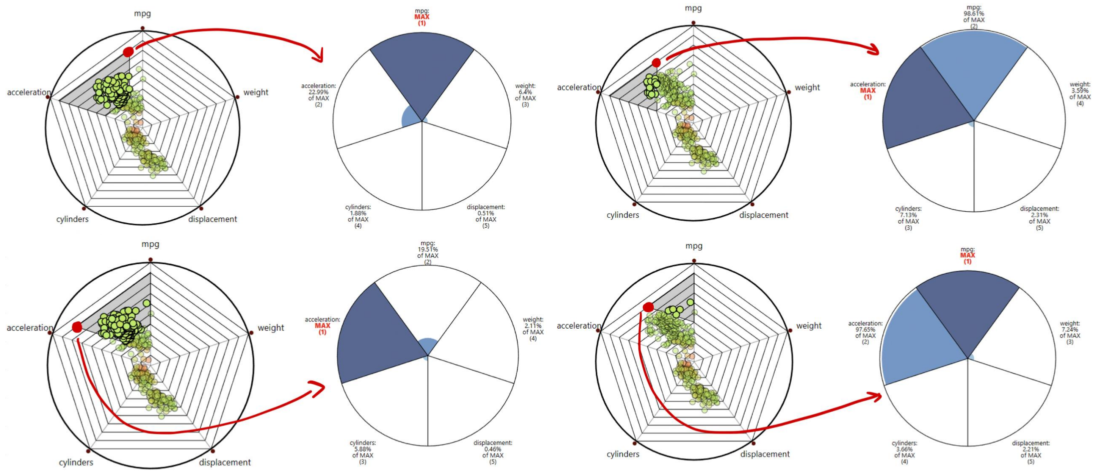

Figure 15: Trapezoidal brush horizontally resized

4.2.2 Validation

This subsection is dedicated to validating the results obtained in the previous one regarding the choice of the trapezoidal selector.
In particular, using the MPG Auto dataset, we can hypothesize a query that selects all points between the acceleration attribute and the mpg attribute, which, in terms of radviz, means the points with a predominance of these two attributes over the others.

In particular, with the squared/rectangular selector, a total of 157 points were selected between the acceleration attribute and the mpg attribute.

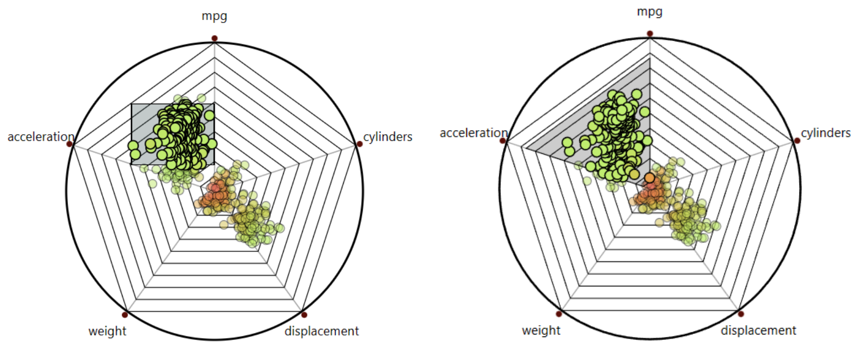

Figure 16: Squared vs Trapezoidal selector

Conversely, with the trapezoidal selector, it was possible to select a total of 462 points. This large difference is due to the fact that, through the trapezoidal selector, it is possible, thanks to the sides parallel to the axes that contain it, to extend it to the maximum from top to bottom in the sector, as shown in Figure 16.
On the other hand, the squared/rectangular selector reaches its maximum extension inside that sector as shown in the Figure 16; otherwise, it would end up in the lower sector and select points that do not fall within the original query, thus creating misleading results for the end user.

From a mathematical point of view, using the Jaccard Distance [17], it can be demonstrated that the trapezoidal selector selects at least the same number of points as the squared one in the case of the aforementioned query within RadViz.
The Jaccard distance is defined as:

$$Jd(A,B) = 1 - J(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

For trapezoidal selector, let A be the set of points selected by it (462) and B the set of points selected by squared selector (157).
The Jaccard distance is $1 - \frac{157}{462}$.
In the same way, for the squared selector, let A be the set of points selected by it (157) and B the set of points selected by trapezoidal selector (462).
Also in this case, the Jaccard distance is $1 - \frac{157}{462}$.
Since the selector with 462 points contains all the points of the selector with 157, then the Jaccard distance between the two selectors would be 0, indicating that the two sets of points are identical. This is because the intersection of the two sets would be equal to the set of points selected by the squared selector with 157 points, since all of those points are also selected by the trapezoidal selector with 462 points. And the union of the two sets would be equal to the set of points selected by the trapezoidal selector with 462 points, since it contains all the points selected by the squared selector with 157 points.
At this point can be noticed that trapezoidal selector selects more points and includes all the points selected by the other selector.

Moreover, an analysis was conducted on the *average* and *standard deviation* of the selected data points pertaining to the attributes of MPG and acceleration. The outcomes of this analysis are shown in the Table 3. Regarding the statistics concerning the acceleration and MPG attributes of the points contained inside the selectors, the trapezoidal selector exhibits a higher average and a larger standard deviation than the squared selector for both attributes, meaning that the squared selector "looses" some points with high values for both attributes.

In general, we can conclude that the trapezoidal selector performs a wider selection of points

Table 3: Average and standard deviation comparison

	Trapezoidal selector	**Squared selector**
MPG average	0.67	0.63
Acceleration average	0.23	0.21
MPG standard deviation	0.13	0.1
Acceleration standard deviation	0.12	0.09

than the squared selector, and the points selected by the trapezoidal selector have higher average values and greater variation compared to the points selected by the squared selector.

5 Conclusions

This paper, after having analyzed the state of the art of high-dimensional data visualizations, specifically radial axes visualizations, proposed an in-depth analysis of their problems. Thanks to this analysis, it allowed to identify issues in common between multiple visualizations and systematized them in a more general manner.
Furthermore, this work attempted to address these problems by proposing and systematizing potential general solutions to help users in their analysis of high-dimensional data.
Moreover, to demonstrate the effectiveness of certain proposed solutions, this work has focused on a specific case study based on RadViz.

More specifically, to mitigate the problem of incorrect positioning of data points in RadViz, the concept of a *representative point* has been adopted, thanks to which a user, after moving the mouse over a point in RadViz, with the help of an auxiliary visualization such as a pie chart, can intuitively and visually understand the proportional relationships between the various dimensions of an ideal point in the same position but with an effectiveness error of 0, i.e., without error. Moreover, they can compare these proportional relationships in real-time with the proportional relationships of the point actually drawn in RadViz, thus having the possibility of getting an idea of which dimensions are represented incorrectly by the point within RadViz and, therefore, in other words, how much a point is misleadingly positioned.

Furthermore, to assist users interactively, a particular type of trapezoidal selector has been used with the RadViz. Multiple selectors can be placed simultaneously within each sector delimited by two axes. This allows users to freely select multiple points among various attribute pairs. An effectiveness error threshold slider has also been added to enable users to “prune” selected points by temporarily removing those that do not meet the threshold, thereby making the selection clearer and facilitating more accurate data analysis.

References

[1] Marco Angelini, Graziano Blasilli, Simone Lenti, Alessia Palleschi, and Giuseppe Santucci. Towards enhancing radviz analysis and interpretation. In *2019 IEEE Visualization Conference (VIS)*, pages 226–230. IEEE, 2019.

[2] Marco Angelini, Graziano Blasilli, Simone Lenti, Alessia Palleschi, and Giuseppe Santucci. Effectiveness error: Measuring and improving radviz visual effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4770–4786, 2021.

[3] RICHARD Bellman. Dynamic programming, princeton univ. *Press Princeton, New Jersey*, 1957.

[4] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer Series in Statistics. Springer New York, 2005.

[5] Chun-houh Chen, Wolfgang Hardle, and Antony Unwin, editors. *Handbook of Data Visualization*. Springer, 2008.

[6] T.F. Cox and M.A.A. Cox. *Multidimensional Scaling, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2000.

[7] Karen Daniels, Georges Grinstein, Adam Russell, and Mason Glidden. Properties of normalized radial visualizations. *Information Visualization*, 11(4):273–300, 2012.

[8] Stephan Diehl, Fabian Beck, and Michael Burch. Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):935–942, 2010.

[9] Geoffrey M Draper, Yarden Livnat, and Richard F Riesenfeld. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, 15(5):759–776, 2009.

[10] Niklas Elmqvist, Pierre Dragicevic, and Jean-Daniel Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE transactions on Visualization and Computer Graphics*, 14(6):1539–1148, 2008.

[11] K. Ruben Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58:453–467, 1971.

[12] JC Gower and David J Hand. Biplots (chapman & hall/crc monographs on statistics & applied probability), 1995.

[13] John C Gower, Sugnet Gardner Lubbe, and Niel J Le Roux. *Understanding biplots*. John Wiley & Sons, 2011.

[14] M Greenacre. Biplots in practice. bbva foundation, madrid, spain, 2010.

[15] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. Dna visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pages 437–441, 1997.

[16] Patrick Hoffman, Georges Grinstein, and David Pinkney. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management*, pages 9–16, 1999.

[17] Paul Jaccard. étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.

[18] Ian Jolliffe. Principal components analysis. 01 2002.

[19] Eser Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. 2000.

[20] Eser Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 107–116, 2001.

[21] Dirk J Lehmann, Fritz Kemmler, Tatsiana Zhyhalava, Marco Kirschke, and Holger Theisel. Visualnostics: Visual guidance pictograms for analyzing projections of high-dimensional data. In *Computer Graphics Forum*, volume 34, pages 291–300. Wiley Online Library, 2015.

[22] Dirk J Lehmann and Holger Theisel. General projective maps for multidimensional data projection. In *Computer Graphics Forum*, volume 35, pages 443–453. Wiley Online Library, 2016.

[23] Manuel Rubio-Sánchez, Laura Raya, Francisco Diaz, and Alberto Sanchez. A comparative study between radviz and star coordinates. *IEEE transactions on visualization and computer graphics*, 22(1):619–628, 2015.

[24] Manuel Rubio-Sánchez and Alberto Sanchez. Axis calibration for improving data attribute estimation in star coordinates plots. *IEEE transactions on visualization and computer graphics*, 20(12):2013–2022, 2014.

[25] Jonathon Shlens. A tutorial on principal component analysis, 2014.

[26] Manuel Sánchez, Alberto Sanchez, and Dirk Lehmann. Adaptable radial axes plots for improved multivariate data visualization. *Computer Graphics Forum*, 36, 06 2017.

[27] Laurens van der Maaten and Geoffrey Hinton. Viualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 11 2008.

[28] Martin Wattenberg, Fernanda Viégas, and Ian Johnson. How to use t-sne effectively. *Distill*, 2016.

[29] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.