DOI: 10.1111/cgf.14923

COMPUTER GRAPHICS *forum*
Volume 42 (2023), number 6, e14923

Smooth Transitions Between Parallel Coordinates and Scatter Plots via Polycurve Star Plots

Dora Kiesel,[1] Patrick Riehmann[2,1] and Bernd Froehlich[1]

[1]Bauhaus-Universität Weimar, Weimar, Germany
{dora.kiesel, bernd.froehlich}@uni-weimar.de
[2]Jönköping University, Jönköping, Sweden
patrick.riehmann@uni-weimar.de

Abstract

This paper presents new techniques for seamlessly transitioning between parallel coordinate plots, star plots, and scatter plots. The star plot serves as a mediator visualization between parallel coordinate plots and scatter plots since it uses lines to represent data items as parallel coordinates do and can arrange axes orthogonally as used for scatter plots. The design of the transitions also motivated a new variant of the star plot, the polycurve star plot, that uses curved lines instead of straight ones and has advantages both in terms of space utilization and the detection of clusters. Furthermore, we developed a geometrically motivated method to embed scatter points from a scatter plot into star plots and parallel coordinate plots to track the transition of structural information such as clusters and correlations between the different plot types. The integration of our techniques into an interactive analysis tool for exploring multivariate data demonstrates the advantages and utility of our approach over a multi-view approach for scatter plots and parallel coordinate plots, which we confirmed in a user study and concrete usage scenarios.

Keywords: visual analytics, visualization, information visualization, interaction, interaction techniques

CCS Concepts: • Human-centred computing → Interaction techniques; Visualization theory concepts and paradigms

1. Introduction

Parallel coordinates and scatter plots are the most popular visualizations for exploring multivariate data. While parallel coordinates shine in showing relations across multiple attributes [HW13], scatter plots are most valuable in assessing the kind and strength of relationships and finding structural clusters [PH21]. The usual way to benefit from both visualizations would be to implement them in a coordinated multi-view framework. However, matching the representations of one data item in separate yet coordinated views requires the selection of its representation in one visualization and finding the corresponding highlighted representation in the other visualization, that is brushing and linking.

In order to enhance the cluster and correlation analysis with parallel coordinates and explain the relationship between data patterns in both visualization types, we developed novel techniques for seamlessly transitioning between parallel coordinates and scatter plots and a new way of embedding additional cluster and correlation information into parallel coordinates. The transition is based on the observation that parallel lines intersect at infinity. Thus, parallel coordinates can be considered as a specific instance of a star plot. By tilting a pair of parallel axes against each other, the intersection point moves closer and forms the centre of a star plot. The lines representing the data items are bent in an elliptic manner resulting in a star plot with curved connectors – the polycurve star plot. The polycurve star plot uses rational cubic Bézier curves instead of the usual lines to improve line tracking and provide a smooth transition to the scatter plot by shifting and the inner Bézier control points towards the scatter point and increasing their weight at the same time (Figure 1). As a result, the Bézier curve converges towards two lines that intersect in the scatter point. We propose to keep the lines and the scatter points during all stages of the transition. This allows us to introduce a new way to project the scatter points during the reverse transition from a scatter plot into star plots and parallel coordinates, thus adding further cluster and correlation information to the respective plots. We implemented a prototypical visualization tool for demonstrating the concept. The tool sets parallel coordinates on a rectangular layout of rails and implements the described transitions on its

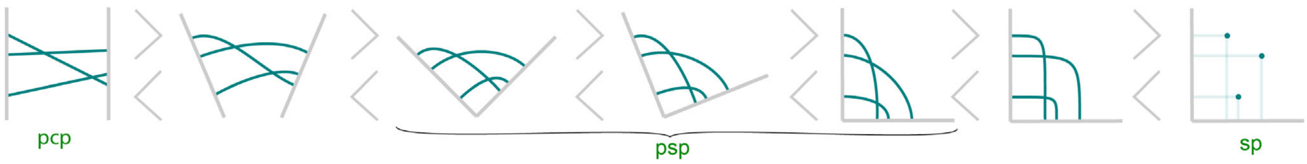


Figure 1: *Schematic representation of the transformation from parallel coordinates (pcp) to scatter plots (sp) via polycurve star plots (psp): While the axes of the parallel coordinates plot rotate, the straight lines representing the data items bend to form the curved segments of the polycurve star plots (pcp - psp). Then rational cubic Bézier curves smoothly transition further into the scatter plot (psp - sp).*

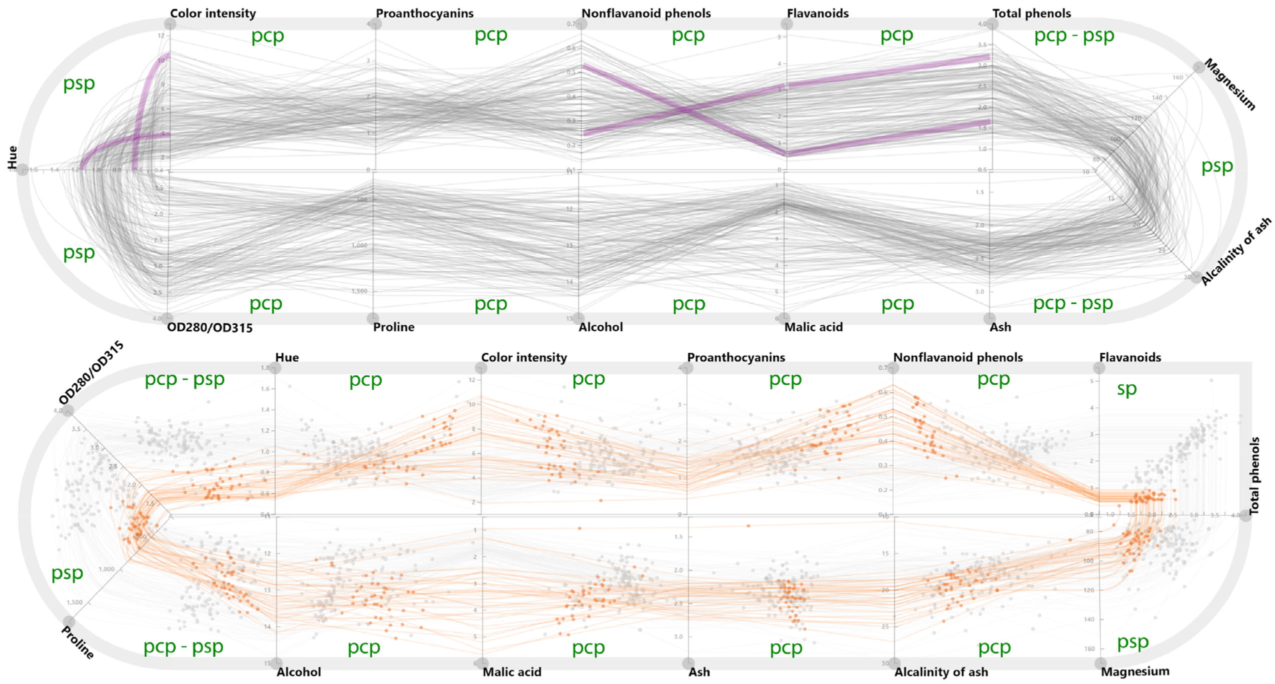


Figure 2: *The prototypical implementation of our transition concept between parallel coordinates and scatter plots from Figure 1: The parallel coordinates (pcp) are set on and can be dragged along rails. At the corners the parallel coordinates smoothly become polycurve star plots (psp). When the axes are perpendicular, the user can grab and drag the corner out to transform the star plot into a scatter plot (sp). The scatter points on the parallel coordinates in the lower image give hints regarding clusters and (cor)relations. The example shows a wine dataset [For98]; the purple lines mark positive and negative correlations. In the lower image, a selection was made using point selection in the scatter plot between the attributes flavanoids and total phenols which highlights the smaller cluster of less steep positive correlation.*

turning points at the left and right end: first to polycurve star plots and on demand to scatter plots and vice versa (Figure 2).

We were motivated by the limitations of previous attempts to combine scatter plots and parallel coordinates: Overlaying the scatter plot creates occlusion [HYFC14], integrating the intersection points of approximated regression lines produces visual clutter and difficult to interpret point patterns [ZW18], redirecting the lines of the parallel coordinates to go through the points of an associated scatter plot distorts the information from the original parallel coordinates [YGX*09], and projecting the points onto the lines without providing a transition that can explain the emerging patterns would result in additional mental effort for the user[REB*16]. Our work overcomes most of these limitations by the following contributions:

- A mathematically grounded technique to smoothly transition between parallel coordinates, polycurve star plots and scatter plots.
- The discussion of the properties of a polycurve star plot. It represents data items as polycurves approximating an elliptic form for orthogonal axis pairs and improves space utilization and the detection of clusters compared to regular star plots.
- The embedding of scatter points into star plots and parallel coordinate plots that are consistent with our transitions and support the visual tracking of structural information such as clusters and correlations during the transition between plot types and the analysis with parallel coordinates and star plots.
- A prototype implementation that demonstrates the seamless integration of parallel coordinates, star plots, and scatter plots into a single view and smooth, user-controlled, geometrically motivated transitions between all three plot types.

We demonstrate the advantages and benefits of our techniques with case studies using the developed prototype and confirm their usability in a user study (see Supporting Information).

2. Related Work

Scatter plots are one of the most straightforward visualization techniques but are still subject to recent research [FHSW13, SSB*15, WHZ*18, PH21]. Harrison et al. [HYFC14] showed scatter plots to be the most efficient visualization to show correlations. However, scatter plots and scatter plot matrices are less effective when comparing data points across more than two attributes at once. Additionally, they do not scale well with an increasing number of attributes. Star coordinate plots generalize scatter plots to depict more than two dimensions [LT13, ML20] and attempt to give a holistic view of the data set. The axes are individually scaled and rotated to get the best possible view of the clusters. However, the ability to examine relationships between individual attributes is lost.

According to Harrison et al. [HYFC14], parallel coordinates [ID90] are the second most effective visualization for negative correlations. Positive correlations, clusters, and non-linear relations are less efficiently detectable. However, research is being conducted to improve these capabilities [ZCQ*09]. Heinrich and Weiskopf [HW13] give a comprehensive overview of the field. Holten and Wijk [HW10] tested several variants of parallel coordinates regarding their performance in cluster evaluation tasks and showed that only pairing the parallel coordinate plot with a scatter plot has a significant effect on cluster identification. To improve the effect, we propose not only showing scatter plots and parallel coordinates side-by-side but also smoothly integrating the scatter points directly into the parallel coordinate system.

Several attempts have been made for integrating of scatter plot information into the parallel coordinates: Holten and Wijk [HW10] depict a small 45 deg rotated scatter plot between two axes of the parallel coordinate plot. Yuan et al. [YGX*09] directly draw a scatter plot of a third attribute between two axes. Thereby, each line is bent to lead through the corresponding dots to link points and lines with each other. In contrast our point projection places points onto the lines without changing the parallel coordinate itself. Likewise, the point projection of Raidou et al. [REB*16] is applied. Their points, however, are placed based on the slope of the line in parallel coordinates. The positions of our scatter points cover more of the given space and base on the ratio of the respective coordinate values. Related but working towards a different goal are Zhou and Weiskopf [ZW18], who integrated indexed points into the parallel coordinates system to depict local correlation.

Other hybrid visualizations have been proposed. Gruendl et al. [GRPF16] seamlessly integrated time series into parallel coordinates. NodeTrix [HFM07] combines node-link diagrams with adjacency matrices. Claessen and van Wijk [CvW11] systematize axis-based hybrid visualizations with the ARGOI model that shows how axes are connected.

Star plots and their various variants [Saa08] are often used in the medical field to compare patient data [SLF*11, CGDGL20]. Sangli et al. [SKK16] extended the visualization to deal with very high dimensional data. Xie and Karki [XK19] extend parallel coordinates with attached star plots showing context axes to better exploit the available space in a Focus & Context view. Slightly different, Fanea et al. [FCI05] combine star glyphs and parallel coordinates to alleviate visual clutter due to many data items. In this paper, the star plot is a mediator between parallel coordinate and scatter plot.

Animation in data visualization can be beneficial during data analysis, as discussed by Fisher [Fis10]. Heer and Robertson [HR07], for example, showed a significant increase in data point tracking and change estimation using simple staged animations and a strong user preference for animation over direct transitions. Systematic analysis of animations and transitions were performed by Tominski et al. [TAA*21], who coined the term *flexible visual analytics*, by Thompson et al. [TLLS20], who explore the design space of animated graphics, and by the Gemini system [KH21] that proposes a grammar of animation. Ruchikachorn and Mueller [RM15] propose animated transitions for teaching visualization techniques. Our proposed method will follow the notion of *fluid interactions* coined by Elmquist et al. [EMJ*11] and employ smooth user-controlled and naturally emerging transitions to explain the relations between data representations of parallel coordinates, star plots, and scatter plots. Bezerianos et al.[BCD*10] employ these principles in transforming different views on node link diagrams into one another. Like our work, transmogrification [BNP*13] also describes a smooth transformation of different visual representations. Their method is, however, limited to transformation through spatial distortion, while the transformation from parallel coordinates to scatter plots requires the transition from a straight line to a single point as representative of each data item.

3. From Parallel Coordinates to Scatter Plots and the Need for a Mediator: The Polycurve Star Plot

The main aspect of this paper is to define a mathematically sound, intuitively comprehensible and aesthetically appealing transition from parallel coordinates to scatter plots that works in both directions and shows the data at any stage during the transition in an interpretable way. For that the continuity of the polylines connecting the data values within the parallel coordinates has to be guaranteed while seamlessly transforming the lines towards the points of a scatter plot as shown in Figure 1 and described in Section 4. To further facilitate the overall impression of looking at a single visualization in different stages and not at two distinct visualizations shapeshifting into one another we provide a novel method for projecting the points of the scatterplot onto their respective lines in the parallel coordinates display (see Figure 2 and Section 5). This way important phenomena from the scatter plot are directly embedded into the parallel coordinate plot to begin with and all the way during its metamorphosis towards the scatterplot.

In order to ensure that smooth transition of the line representation to the scatter points, we need the starplot visualization to meet specific requirements to properly function as a mediator. For orthogonal axes pairs, we require the connecting curve to form a circle segment for pairs of equal data values; ellipse segments for unequal ones. For other angles δ between the axes, the curve segments shall retain the ellipse's property of intersecting with the axes perpendicularly, such that data items are visualized by a continuous smooth shape strengthening the connection between polycurve segments and allowing for easier tracing of polycurves (see Figure 3 left).

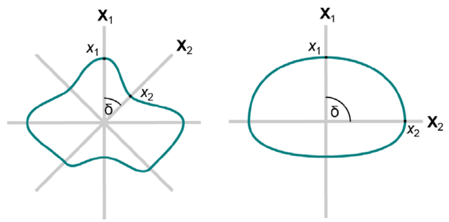
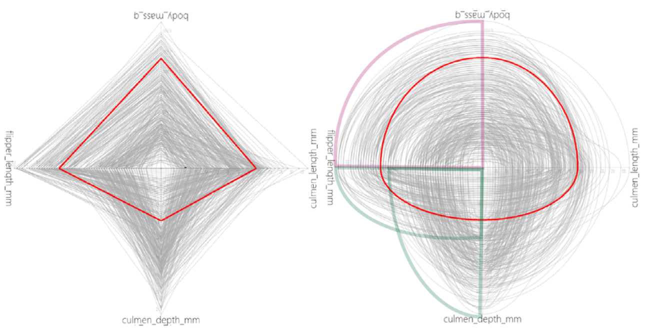


Figure 3: *Schematic structure of the polycurve star plot: the axes $\mathbf{X}_i$ are placed at equal angles δ (left: $\delta = 45$ deg, right: $\delta = 90$ deg) around the center. The coordinates x_i define where the polycurve intersects each axis.*



(a) standard star plot (b) polycurve star plot

Figure 4: *Comparison of standard and polycurve star plots using the penguin data set (345 samples)[HHG20]. The curved connectors in (b) allow for some interesting observations: In the upper left quadrant (purple mark), only few crossings of curves occur, suggesting a positive correlation. In the lower left quadrant, on the other hand, a significant number of crossings is visible. On closer inspection, one can observe that the crossover stems from two clusters of positive correlation (green marks), since within each band, only few crossings are visible. This conclusion could not be drawn from the standard star plot in (a) due to heavy overplotting issues.*

A polycurve star plot using rational cubic Béziers can be constructed to meet these criteria. It employs one axis $\mathbf{X}_i$ per dimension arranged at equal angles δ around the centre as regular star plots, yet with polycurves instead of straight polylines representing a data item (see Figure 3). Those are modelled as rational cubic Béziers, since they are able to perfectly represent ellipses, allowing us to always intersect axes orthogonally and offer a smooth way to transition to the scatter plot by increasing the weights of the inner control points. Figure 4 shows an example of the penguin data set [HHG20]. Though the general concept of polycurved star plots is not entirely new (it is used for some infographics and charting tools, e.g. Microsoft Excel, ggplot, or d3js) it was never thought of as mediator visualization. Hence, our work will provide – to the best of our knowledge – the first scientific consideration of its benefits and usage as a transitional mediator between parallel coordinates and scatterplots.

Although the polycurve star plot started as a concept to smooth the transitions between parallel coordinates and scatter plots, the design has some additional advantages: (1) The design makes better

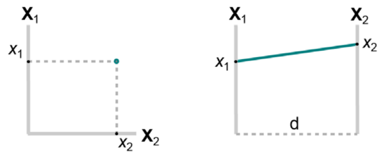


Figure 5: *Schematic structure of scatter plot and parallel coordinates denoting the convention of variables used in the paper: coordinates x_i are situated on the axes $\mathbf{X}_i$ to construct the point (scatter plot), line (parallel coordinate) or curve (polycurve star plot). d describes the distance between two axes in parallel coordinates.*

use of the available space compared to a regular star plot, thereby reducing some of the overplotting issues. (2) The curves conform with the Gestalt Principles of Continuity and Closure to foster the traceability of individual data items and the Gestalt Principles of Proximity and Similarity on groups of curves to highlight patterns and clusters (Figure 4b).

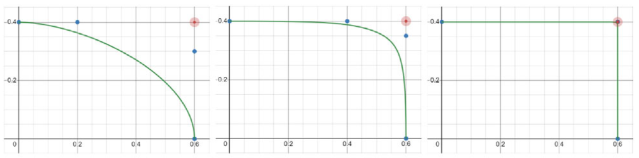
4. Smooth Transition Between Parallel Coordinates and Scatter Plots: Line Representation

The transition between parallel coordinates and scatter plots depends on two mathematically grounded considerations: (1) parallel coordinates can be seen as special instance of star plots with axes crossing at infinity, thus one form can be transitioned into the other by moving the intersection point of the axes (Figure 1 pcp-psp). (2) shifting and increasing the weights of a rational cubic Bézier curve's inner control points smoothly transforms an ellipse segment to a rectangle segment with its corner being the scatterpoint (Figure 1 psp-sp). The idea for this transformation was motivated by superellipses which can also transition from an elliptic segment to a rectangular segment; however, Bézier curves are easier to use since superellipses require the use of large exponents to converge towards a rectangular shape.

We uniformly denote all axes – the infinite lines defining the spatial position and angle of an attribute – as $\mathbf{X}_i$ and coordinates x_i (see Figure 5 and Figure 3). Additionally, d denotes the distance between axes for parallel coordinates, δ is the angle between axes in the polycurve star plot.

4.1. From parallel coordinates to polycurve star plots

The transition from parallel coordinates to polycurve star plots (Figure 1 (pcp – psp)) follows geometrical considerations: the axes of the parallel coordinates intersect at infinity. By rotating two adjacent axes left and right, respectively, the intersection point of the axes moves into the finite range. Thereby, the range of data values is also moved to remain in the visible area. Since the transformation distorts the space between two axes, the lines start to bend while the axes are rotated, resulting in ellipse segments for perpendicular axes. Thus, we reinterpret the star plot's radial layout to conform with polar coordinates: The coordinates are interpreted as radii instead of lengths connected via ellipse segments rather than straight lines.



(a) initial state (b) control points move; the weight increases. (c) final rectangle

Figure 6: *The interplay between linearly shifting the control points (blue) towards the actual scatter point (red) and exponentially increasing the weight of the control points transforms an ellipse into a rectangle segment.*

4.2. From polycurve star plots to scatter plots

The design of a transition between polycurve star plots and scatter plots relies on a rational Bézier curve's ability to represent ellipse segments and converge towards rectangle segments (see Figure 1 (psp - sp)). A rational Bézier curve in general is defined as:

$$B(t) = \frac{\sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} P_i w_i}{\sum_{i=0}^{n} \binom{n}{i} t^i (1-t)^{n-i} w_i} \tag{1}$$

were P_i are control points and w_i are weights. The higher the weight, the closer the resulting curve will pass by the respective control point. While a rational quadratic Bézier curve – so $n = 2$ – suffices to perfectly represent an ellipse segment, we use rational cubic Bézier curves which obviously can also represent rational quadratic Bézier curves but additionally allow us to represent S-shaped curves as required for general polycurve starplots with tangential transitions across axes as shown in Figure 3. In order to achieve the desired effect of turning the ellipse segment into a rectangle, we not only need to increase the weights of the inner control points P_1 and P_2, but also shift them gradually towards the scatter point (Figure 6). For the whole transition to feel linear – as required later in our implementation – the control points are shifted linearly towards the scatter point while the weights are increased exponentially towards infinity to pull the curve into the rectangular corner.

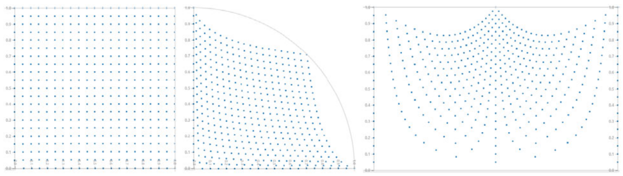
5. Smooth Transition Between Parallel Coordinates and Scatter Plots: Point Representation

we introduce a new projection of scatter points onto parallel coordinates and star plots to continuously show the scatter point on the curves and lines throughout the whole transition which avoids blending in the actual scatter point of the scatter plot after the transition of the line representation has concluded. Additionally, some of the structural information that allow the scatter plot to be so effective in cluster and correlation analysis become accessible in the PCP.

The projection is based on the definition of a regular scatter plot's point: A coordinate pair (x_i, x_{i+1}) places the point in the scatter plot such that it has a distance of x_i from $\mathbf{X}_{i+1}$ and a distance of x_{i+1} from $\mathbf{X}_i$. In Figure 5, we marked these distances with light grey dashed lines. The grey dashed lines form the rectangle segment that results from our transition described in Section 4. When the range of data values is normalized to [0, 1], the position of the point on this rectangle can be described by the ratio of the coordinate x_{i+1} (the length of the horizontal gray dashed line in Figure 5) to the total arc length of the rectangle (both gray dashed lines in Figure 5), so by the fraction f:

$$f = \frac{x_{i+1}}{x_i + x_{i+1}} \tag{2}$$

The points of the resulting embedding move smoothly during transitions and can be used for the analysis in polycurve star plots and parallel coordinates (Figure 7).



(a) scatter plot (b) star plot (c) pcp

Figure 7: *A regular grid of scatter points in (a) scatter plot, (b) polycurve star plot, and (c) parallel coordinates (PCP): each transformation warps the original grid further while still maintaining structural information, especially along the main diagonal. (0,0) is missing in the PCP, since it is spread across the lower end.*

Since x_i and x_{i+1} need to be normalized to range [0,1], the fraction f for (0,0) results in a division by zero. Thus, the scatter point representing (0,0) is not defined and will not be displayed in the PCP. The nature of the transition explains the peculiarity: As the axes are pulled apart in the transition from star plot to parallel coordinates, (0,0) representing the point of contact of the two axes is also pulled apart and degenerates to a line.

Please note that these projected points have an entirely different meaning than those proposed by Inselberg himself [ID90]. Inselberg's points are constructed from a line in a scatter plot and could be used for example to visualize a regression line as a summary of many single coordinates. In our point projection, each line is enhanced by one point that encodes the same information in another way, thereby enabling the user to draw refined conclusions.

5.1. Scatter point patterns in a parallel coordinate plot

Geometrically, fraction f projecting a scatter point onto a parallel coordinate describes the normalized angle of the polar coordinate:

$$f = \frac{\arctan \frac{x_{i+1}}{x_i}}{\frac{\pi}{2}} \tag{3}$$

All points with the same ratio between x_{i+1} and x_i, so the points on a straight positively sloped line through (0,0) will have the same fraction f. Transforming a straight positively sloped line through (0,0), thus, results in a straight line parallel to the axes in parallel coordinates. Corresponding correlation patterns – global strong positive correlations – are, therefore, also vertically oriented. Since in our model of transforming the parallel coordinates to polycurve star plots (see Section 4.1), we moved the intersection point from infinity to a finite position, it makes sense that all lines through (0,0) in a scatter plot are parallel to the axes in parallel coordinates. Our positioning of the scatter points is consistent with this.

Figure 7 shows the distortion in an originally regular grid when transformed to parallel coordinates. Points near (0,0) are dragged apart, while points near the end of the diagonal are compressed. The embedding transforms the grid's straight columns and rows in the scatter plot to curves in the parallel coordinate system. Still, there are three straight lines in the distorted grid, all parallel to the axes: one in the centre between the two axes that constitutes the diagonal and one directly on each axis. These result from grid lines originally forming straight lines through the origin that will transform into parallels to the axes as described above. The projection of the main diagonal onto a straight centred axis-parallel line is particularly useful since the projected point cloud of a positive correlation will arrange itself close to it, providing additional visual cues to an otherwise hardly detectable PCP line pattern. While the embedding does not preserve the regularity of the original grid, structural relations like neighbourhoods are largely preserved so that the patterns can give valuable cues regarding the estimation of relations between attributes and the detection of clusters.

Strong positive relations translate into either parallels to the axes if the regression line intersects (0,0) or into curves that tend towards parallelism on one end and intersect with one axes on the other forming smoothed L-like shapes (first row of Figure 8). Strong negative correlations show as U-like shapes intersecting with one or even both axes (second row of Figure 8).

The lower half of Figure 8 shows two different linear separable cluster configurations. Even though the shapes of the clusters are distorted due to the embedding, the clusters are still clearly separable since structural relationships are preserved. Although densities are distorted during the transformation, even the clusters in the bottom row can be easily told apart. So, the embedding of the scatter plot into the PCP results in an informative visual aid in analysing cluster and correlation patters in parallel coordinates.

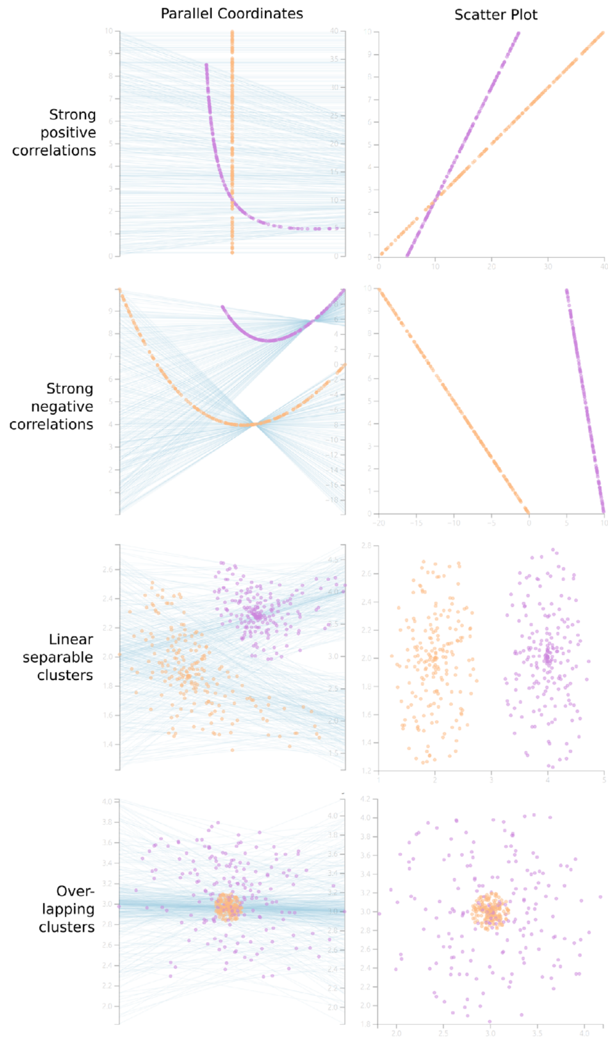


Figure 8: *Typical correlation and cluster patterns formed by the scatter points in the parallel coordinate plot.*

5.2. Comparison to the point definition of Raidou et al.

Raidou et al. [REB*16] designed a slope-based embedding: horizontal lines get a point in the centre; the points of positive-sloped lines are closer to the axis $\mathbf{X}_{i+1}$; the points of negative-sloped lines are closer to axis $\mathbf{X}_i$ ($|\mathbf{X}|$ denotes the axis length, d is assumed to be 1, Δ represents the slope):

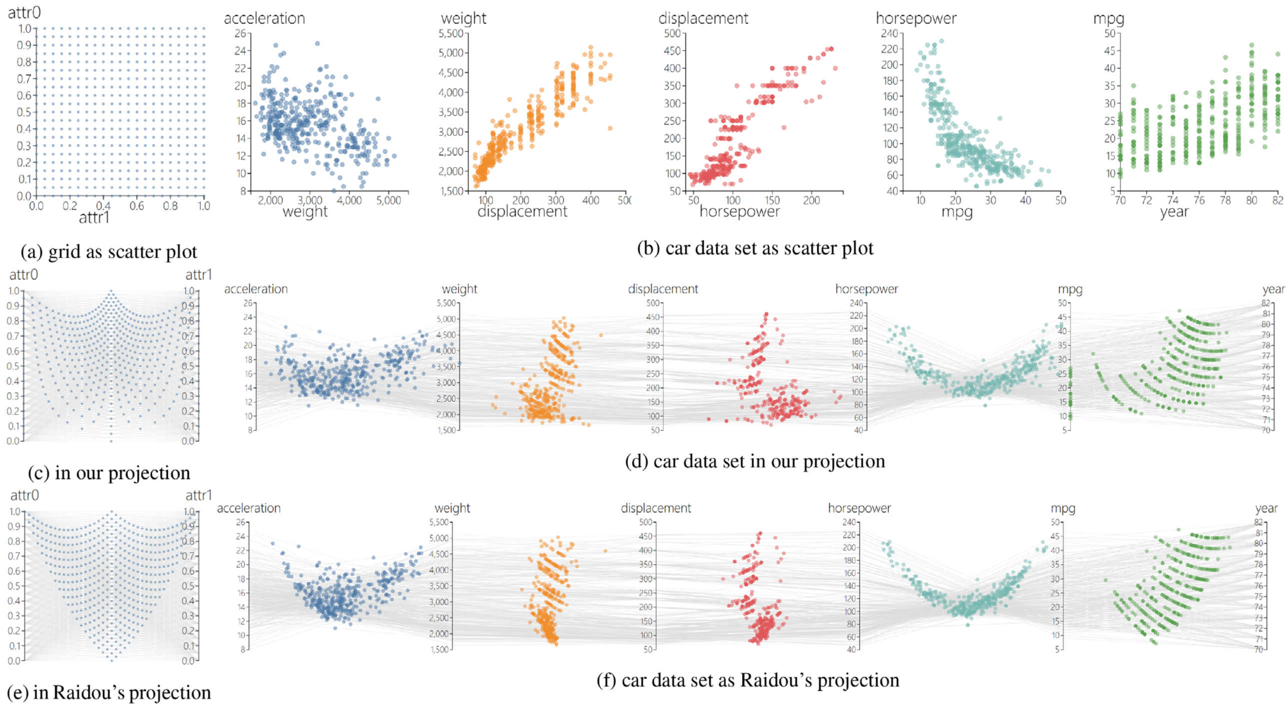
$$\begin{aligned} x_p &= \frac{\Delta}{2|\mathbf{X}|} + \frac{1}{2} \\ y_p &= x_i + \Delta \cdot x_p \\ \Delta = x_{i+1} - x_i&;\ -|\mathbf{X}| \leq \Delta \leq |\mathbf{X}| \end{aligned} \tag{4}$$

Raidou's method generates a warped embedding of a scatter plot in the parallel coordinates plot. Their grid is more compressed but also more evenly spaced than the scheme introduced in this paper, as can be seen in Figure 9c and e. The tighter layout in Raidou et al. [REB*16] introduces fewer distortions, especially in the lower half of the plot. Assuming a normalized range of values of [0,1], (0,0) is represented by one point in Raidou et al. [REB*16]; in our scheme, (0,0) is mathematically undefined but can be interpreted as being spread across the line between the lower ends of the axes. The behaviour of points on a scatter plot's axes also differs: Our scheme projects the points on the axes as one would expect; Raidou et al. [REB*16] place the points on a curved line between the centre of the line between the lower ends of the axes and the upper end of the respective axis (green points in Figure 9b, d, and f).

6. Interactive Analysis Tool

We designed a web-based, interactive analysis tool to demonstrate the natural interplay between parallel coordinates, polycurve star plots, and scatter plots. It implements the smooth, user-controlled transitions from parallel coordinates to scatter plots via polycurve star plots and the scatter points, the embedding of the scatter plot into the PCP.

The application sets the axes of the regular parallel coordinates plot on rails, thus defining a closed path along which the whole



(a) grid as scatter plot (b) car data set as scatter plot

(c) in our projection (d) car data set in our projection

(e) in Raidou's projection (f) car data set as Raidou's projection

Figure 9: *Comparison of the scatter points from this paper with those from Raidou et al.[REB*16]: Both methods transport indicators for positive (I-shaped point clouds, orange and red) and negative correlations (U-shaped point clouds, blue and cyan). The lower parts of our projections are more spread out than those of Raidou et al. [REB*16]. Points on the axes (green) are only detectable in our scheme.*

plot can be dragged (Figure 2). Arcs connect each adjacent pair of straight rails. When dragging the parallel coordinate plot over an arc, the parallel coordinate representation is seamlessly transformed into a polycurve star plot, thereby revealing groupings of similar data items and simplifying the interpretation of the projected scatter points (see Figure 2). The inner arcs have configurable radii to reduce overplotting of the polycurve star plot near the centre. In order to transform the polycurve star plot into a scatter plot, the user simply pulls out the outer border, which then becomes square (Figure 2, sp on the right turn). We decided to give the control over the transition to the scatter plot explicitly to the user instead of doing it automatically, so that the user can stop at any in-between state, better understands what happens during the transition, and to avoid introducing the interaction delay an animation introduces.

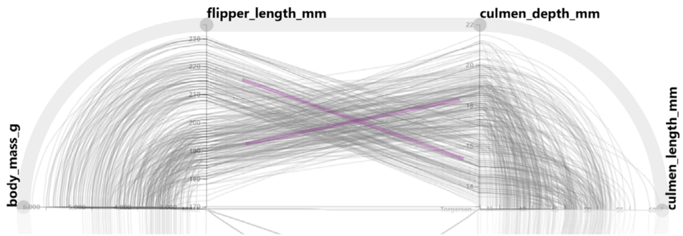
For our examples (Figure 2 and Figure 10), we chose a slim rectangular shape to optimize screen space usage. This layout creates four 90 deg angled spaces for polycurve star and scatter plots, which is the ideal configuration for both visualizations as established in Section 4.2. In order to optimally exploit the available space, the parallel coordinate plot is spread around the whole rail and closed to form a band of axis pairs (or a closed path of axes according to the ARGOI model [CvW11]). Axes can be dragged to be swapped allowing different axis configurations. Other layouts or angles are possible and might be more useful in special scenarios.

The application offers two brush types: one to select lines, the well-known standard that can be easily used to define value ranges along the axes; one to select scatter points to simplify cluster selection. Moreover, the user is able to adjust the opacities for lines and points separately. Points can be faded out when the lines should be scrutinized for correlations, or the lines can be faded into the background when analysing the scatter plots.

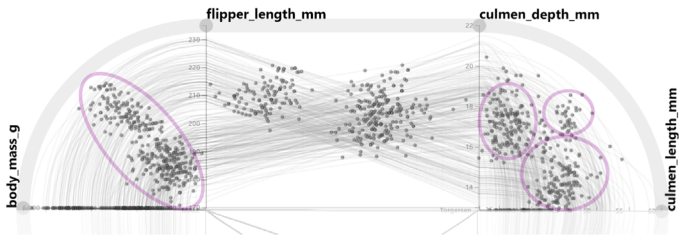
Other PCP implementations use cubic curves instead of straight lines [GK03, ROF12] for connecting the axes. Our own experiments show that cubic curves for the parallel coordinate plot make the transition appear even smoother and strengthen the traceability of the polycurves through the plot, we still decided to use the standard implementation of the parallel coordinate plot with straight lines because they convey important information for judging local or global correlation in the data which might not be as apparent if cubic curves would be used.

7. Analytic Scenarios

The first use case analyses a dataset of 179 wines with 13 continuous wine quality indicators (see Figure 2)[For98]. Looking at the polycurves, there seem to be several negative correlations (e.g. *flavanoids – nonflavanoid phenols or hue – colour intensity*, purple crosses in Figure 2) as well as a positive one between *flavanoids* and *total phenols* (Figure 2, purple parallels). To check whether this is indeed a positive correlation, we first add points to the plot, which gives us another hint towards a positive correlation due to the point cloud staying within a corridor parallel to the axes. Moreover, we can even see two clusters of wines that are separable on this attribute pair. Brushing these points shows that the clusters also has very distinct values on other attributes. They relate, for example to many *nonflavanoid phenols* and, for the most part, *low hue* and *OD280/OD315* values. The transition of the axis pair *flavanoids* and



(a) without scatter points



(b) with scatter points

Figure 10: *Two screens of the penguin dataset: (a) Without scatter points it seems to indicate a negative correlation between* flipper length *and* culmen depth *(purple cross). (b) The points reveal that the pattern actually stems from two positively correlated clusters. The top right quadrant even exhibits a third cluster (purple ellispes).*

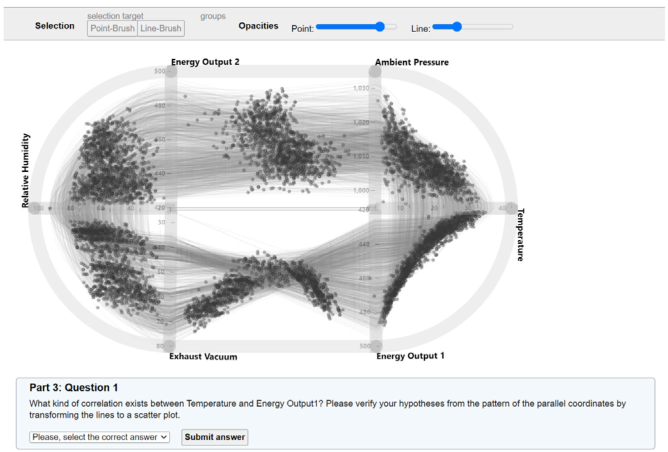


Figure 11: *Web application with the main study's first question on the combined cycle power plant dataset [KTG12]. The form of the point cloud between* temperature *and* energy output 1 *indicates a strong negative correlation. We increased the width of the axis representations in comparison to the other depictions of the system to facilitate selecting and dragging the axes during the study.*

total phenols to a scatter plot confirms the assumption: two different positively correlated patterns become visible.

The second use case bases on a dataset of 345 observations of penguins on three islands [HHG20]. It consists of only two continuous and two categorical attributes regarding different properties of these penguins. The dataset was chosen due to its well known and clear data patterns some of which can already be observed by just analysing the polycurves of our implementation in Figure 10 (top). The top centre parallel coordinates view between the attributes *flipper length* and *culmen depth* seems to show a typical crossing pattern indicating a negative correlation. But once our novel point projection is turned on (Figure 10 (bottom)) it becomes very clear that there are actually two clusters each showing a positive correlation – as can be quickly verified by transforming it smoothly into a scatter plot. The right star plot even reveals a third cluster and the left one shows the apparently strong positive correlation between *body mass* and *flipper length* in the upper left quadrant as already described in Section 3 (both marked in purple in Figure 10).

The dataset of a combined cycle power plant's measurements and output energy (about 9500 data, five attributes, see Figure 11)[KTG12]. The *output energy* has been duplicated to better fit the requirements of our application and to offer quicker analysis between input and output variables. It shows a strong negative correlations between *energy output* and *temperature* and *energy output* and *exhaust vacuum*. Also, the corridor of points between *energy output* and *ambient pressure* hints at a positive correlation between these attributes. The lines also show a distinct cluster of low *exhaust vacuum* values. Highlighting the cluster reveals barely visible clusters in other attribute pairs, for example *ambient pressure* and *energy output*.

These cases show how the implementation of our concepts – the combination of straight parallel coordinates, polycurve star plot and point projection – supports in analysing even obscured patterns. The ability to seamlessly transition to a scatter plot eliminates the additional analysis step and mental effort to switch context and getting oriented in a separate scatterplot for verifying found patterns.

8. User Study

We conducted a user study to evaluate how the draggable layout, the smooth transformation between parallel coordinates, star and scatter plots and the projected points support the users. Nineteen participants (five female, 14 male) with basic to profound experience with visualization and a background in computer science took part. All participants successfully participated in an introductory visualization course and thus were familiar with the concepts of parallel coordinates, star plots and scatter plots.

The study consisted of five parts: (1) a short introduction to the visualization types, how correlation can be deduced from visual patterns, how the transformation and the point projection works and an introduction to the web application used for the study, (2) a test phase with eight test questions to make sure the participant is familiar with correlation and cluster patterns in parallel coordinates and scatter plots, (3) a playground where the participant can play around with the application to gain confidence using it, (4) the main study consisting of seven tasks that lead the participants through the analytic steps described in Section 7, and (5) an SUS questionnaire and particular questions about the usefulness of single features.

The main study consisted of seven tasks with increasing difficulty. Three tasks required the users to estimate and find a certain type of correlation (see Figure 11 for an example). Another two tasks were about finding and selecting specific clusters. The remaining two tasks asked about complex multi-attribute relations. The easier tasks could be solved correctly by most participants (80%–95%). For the more complex tasks the correct answers were in the range of 45%- -65%. Here participants identified cluster centres correctly in most cases but had issues with the specification of the outline of clusters which was somewhat cumbersome. Accordingly, some participants remarked that they would appreciate more fine-grained selection control to improve the precision of cluster selections.

The overall average SUS score for our system is 74.7 (SD = 13.5), which translates to a good usability. Particularly noteworthy are the good results regarding integration (question 5), consistency (question 6) and learnability (question 7) with average scores of 4.47 (SD = 0.6), 4.37 (SD = 0.9, inverted) and 4.26 (SD = 1.1). They show how much the integration of the scatter plot into the parallel coordinate system was appreciated and that the points were helpful without obscuring other information.

Eleven participants found the smooth transitions to be particularly useful. One of them especially appreciated the continuous control over the transition; another remarked that being able to transition to the scatter plot greatly improved his confidence in interpreting the PCP patterns. Also, the selection tools (10 participants), the point projection (two participants) and the adjustable layout (two participants) were mentioned as greatly supporting the given tasks. Nine participants missed the ability to switch axes which was not made available to simplify the interface. Additional features like a more fine-grained selection control (three participants), snapping the axes to a grid (two participants) and flexible spacing of axes (one participant) have been proposed after the study. Two participants reported that they found the technique fun to use and found themselves playing around with the data on their own accord. Another participant even felt '[…] like visual analytic experts would really appreciate this system', another agreed in a short oral recap after the study.

Another five questions surveyed the usefulness of specific parts of the design to be answered with a Likert scale from 1 to 5. Through the participants reported that dragging the axes was easy to do (M = 4.5, SD = 0.6) and useful (M = 4.8, SD = 0.5), transitioning from parallel coordinates to the scatter plot was simple to follow and understand (M = 4.5, SD = 0.5) and the scatter points supported the participants in figuring out clusters (M = 4.2, SD = 1.0) and correlations (M = 4.2, SD = 0.9).

Overall, the results from the tasks and the participants' reports regarding the application's support in correlation and cluster analysis show the great potential of the smooth integration of scatter plots and the point embedding to enhance parallel coordinate views.

9. Scalability and Appearance

The 9500 data items of the power plant dataset [KTG12] used in the user study (see Figure 11) push the technique to its limit in terms of overplotting (interaction still works well). The application's scalability depends essentially on its base visualizations' scalability. Thus, advanced techniques that enable scatter plots, star plots, and parallel coordinates to cope with a larger number of data items such as continuous parallel coordinates and scatter plots also improve the scalability of our approach.

The 13 attributes of the wine dataset (see Figure 2) [For98] cap the number of attributes that could be analysed at the same time. For a larger number of attributes, other configurations of rails – like a serpentine line or some off-screen space – might do a better job. The six attributes of the penguin dataset (see Figure 10) are the recommended minimum for the closed rectangular form. Five attributes would work but reduce the number of simultaneously available orthogonal axis pairs to two but using an axis more than once is a sensible option in many cases.

10. Conclusion and Future Work

We presented seamless and geometrically grounded transitions between parallel coordinates and scatter plots via polycurve star plots. The polycurve star plot serves as a mediator visualization to allow for smooth transitions between parallel coordinates and scatter plots. It also improves the traceability of individual data items and cluster detection over regular star plots. The embedding of the scatter points facilitates the assessment of clusters and relations by adding valuable information to the usual line representation of parallel coordinate plots and star plots and provides another precise way to select data items from the data set. The smooth transitions in combination with the scatter points allow for visual tracking of various data configurations across plot types and therefore support the process of verifying clusters and correlations in the different plots. This was confirmed by the results of our user study, in particular the SUS score and also by the participants' comments especially emphasizing the smoothness of the transitions.

The flexible layout of the visualization based on rails allows for extending the technique in various ways. For example, some sections of the rails could be squeezed or expanded to provide a focus and context view. One could also imagine a four-sided layout with a larger rectangle in the middle which could host a control centre or a semantically abstracted view of the data. While the added value of such enhancements depends on the actual dataset, the usability of the techniques needs to be evaluated in an application context where it is just one of many steps in the data analysis process.

Our next step addresses the scalability to larger datasets. To this end, we plan to investigate the applicability of our approach to cluster-based parallel coordinates and continuous parallel coordinates and scatter plots. However, even in its current form, our work contributes already to the portfolio of comprehensible and intuitive transitions between different visualization techniques.

Acknowledgements

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 16DHBKI083 (project SKILL).

Open access funding enabled and organized by Projekt DEAL.

References

[BCD*10] Bezerianos A., Chevalier F., Dragicevic P., Elmqvist N., Fekete J.-D.: Graphdice: A system for exploring multivariate social networks. *Computer Graphics Forum 29* (2010), 863–872.

[BNP*13] Brosz J., Nacenta M. A., Pusch R., Carpendale S., Hurter C.: Transmogrification: Causal manipulation of visualizations. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2013), UIST '13, Association for Computing Machinery, pp. 97–106. https://doi.org/10.1145/2501988.2502046.

[CGDGL20] Caporaso T., Grazioso S., Di Gironimo G., Lanzotti A.: Biomechanical indices represented on radar chart for assessment of performance and infringements in elite race-walkers. *Sports Engineering 23*, 1 (2020), 1–8.

[CvW11] Claessen J. H., van Wijk J. J.: Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2310–2316. https://doi.org/10.1109/TVCG.2011.201.

[EMJ*11] Elmqvist N., Moere A. V., Jetter H.-C., Cernea D., Reiterer H., Jankun-Kelly T.: Fluid interaction for information visualization. *Information Visualization 10*, 4 (2011), 327–340. https://doi.org/10.1177/1473871611413180.

[FCI05] Fanea E., Carpendale S., Isenberg T.: An interactive 3d integration of parallel coordinates and star glyphs. IEEE Symposium on Information Visualization, 2005. INFOVIS 2005. (2005), 149–156.

[FHSW13] Fink M., Haunert J.-H., Spoerhase J., Wolff A.: Selecting the aspect ratio of a scatter plot based on its delaunay triangulation. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2326–2335. https://doi.org/10.1109/TVCG.2013.187.

[Fis10] Fisher D.: Animation for Visualization: Opportunities and Drawbacks. In *Beautiful Visualization ed. O'Reilly Media*. April 2010.

[For98] Forina M., Leardi R., Armanino C., Lanteri S.: An extendible package for data exploration, classification and correlation. Elsevier Science, 01 1998.

[GK03] Graham M., Kennedy J.: Using curves to enhance parallel coordinate visualisations. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* (08 2003), vol. *2003*, pp. 10–16. https://doi.org/10.1109/IV.2003.1217950.

[GRPF16] Gruendl H., Riehmann P., Pausch Y., Froehlich B.: Time-series plots integrated in parallel-coordinates displays. *Computer Graphics Forum 35*, 3 (2016), 321–330. https://doi.org/10.1111/cgf.12908.

[HFM07] Henry N., Fekete J.-D., McGuffin M. J.: Nodetrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1302–1309. https://doi.org/10.1109/TVCG.2007.70582.

[HHG20] Horst A. M., Hill A. P., Gorman K. B.: *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*, 2020. R package version 0.1.0. URL: https://allisonhorst.github.io/palmerpenguins/, https://doi.org/10.5281/zenodo.3960218.

[HR07] Heer J., Robertson G.: Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics 13* (2007), 1240–1247.

[HW10] Holten D., Wijk J. V.: Evaluation of cluster identification performance for different pcp variants. *Computer Graphics Forum 29*, 3 (2010), 793-802. https://doi.org/10.1111/j.1467-8659.2009.01666.x.

[HW13] Heinrich J., Weiskopf D.: State of the art of parallel coordinates. In *Eurographics (STARs)* (2013), pp. 95–116.

[HYFC14] Harrison L., Yang F., Franconeri S., Chang R.: Ranking visualizations of correlation using weber's law. *IEEE Transactions on Visualization and Computer Graphics 20* (2014), 1943–1952.

[ID90] Inselberg A., Dimsdale B.: Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the First IEEE Conference on Visualization: Visualization90* (1990), IEEE, pp. 361–378.

[KH21] Kim Y., Heer J.: Gemini: A grammar and recommender system for animated transitions in statistical graphics. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 485–494. https://doi.org/10.1109/TVCG.2020.3030360.

[KTG12] Kaya H., Tüfekci P., Gürgen F. S.: Local and global learning methods for predicting power of a combined gas & steam turbine. In *Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering ICETCEE* (2012), pp. 13–18.

[LT13] Lehmann D. J., Theisel H.: Orthographic star coordinates. *IEEE Transactions on Visualization and Computer Graphics 19* (2013), 2615–2624.

[ML20] Matute J., Linsen L.: Hinted star coordinates for mixed data. *Computer Graphics Forum 39*, 1 (2020), 117–133. https://doi.org/10.1111/cgf.13666.

[PH21] Perkhofer L., Hofer P.: Optimizing scatterplot-matrices for decision-support. In *Information Systems and Neuroscience* (Cham, 2021), Davis F. D., vom Brocke J., Léger P.-M., Randolph A. B., Müller-Putz G., (Eds.), Springer International Publishing, pp. 63–76.

[REB*16] Raidou R. G., Eisemann M., Breeuwer M., Eisemann E., Vilanova A.: Orientation-enhanced parallel coordinate plots. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 589–598. https://doi.org/10.1109/TVCG.2015.2467872.

[RM15] Ruchikachorn P., Mueller K.: Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics 21* (2015), 1028–1044.

[ROF12] Riehmann P., Opolka J., Froehlich B.: The product explorer: Decision making with ease. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2012), pp. 423–432.

[Saa08] Saary M. J.: Radar plots: A useful way for presenting multivariate health care data. *Journal of Clinical Epidemiology61*, *4* (2008), 311–317.

[SKK16] Sangli S., Kaur G., Karki B. B.: Star plot visualization of ultrahigh dimensional multivariate data. In *Proceedings of International Conference on Advances in Big Data Analytics* (2016), pp. 91–97.

[SLF*11] Stafoggia M., Lallo A., Fusco D., Barone A. P., D'Ovidio M., Sorge C., Perucci C. A.: Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *Journal of Clinical Epidemiology 64*, 7 (2011), 770–778. https://doi.org/10.1016/j.jclinepi.2010.10.009.

[SSB*15] Shao L., Schleicher T., Behrisch M., Schreck T., Sipiran I., Keim D. A.: Guiding the exploration of scatter plot data using motif-based interest measures. In *2015 Big Data Visual Analytics (BDVA)* (2015), pp. 1–8. https://doi.org/10.1109/BDVA.2015.7314294.

[TAA*21] Tominski C., Andrienko G., Andrienko N., Bleisch S., Fabrikant S. I., Mayr E., Miksch S., Pohl M., Skupin A.: Toward flexible visual analytics augmented through smooth display transitions. *Visual Informatics 5*, 3 (2021), 28–38. https://doi.org/10.1016/j.visinf.2021.06.004.

[TLLS20] Thompson J., Liu Z., Li W., Stasko J.: Understanding the design space and authoring paradigms for animated data graphics. *Computer Graphics Forum 39*, 3 (2020), 207–218. https://doi.org/10.1111/cgf.13974.

[WHZ*18] Wang Y., Han F., Zhu L., Deussen O., Chen B.: Line graph or scatter plot? Automatic selection of methods for visualizing trends in time series. *IEEE Transactions on Visualization and Computer Graphics 24*, 2 (2018), 1141–1154. https://doi.org/10.1109/TVCG.2017.2653106.

[XK19] Xie K., Karki B.: A hybrid approach based on parallel coordinates and star plot. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – IVAPP*, (2019), INSTICC, SciTePress, pp. 267–274. https://doi.org/10.5220/0007375502670274.

[YGX*09] Yuan X., Guo P., Xiao H., Zhou H., Qu H.: Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1001–1008. https://doi.org/10.1109/TVCG.2009.179.

[ZCQ*09] Zhou H., Cui W., Qu H., Wu Y., Yuan X., Zhuo W.: Splatting the lines in parallel coordinates. In *Proceedings of the 11th Eurographics/IEEE - VGTC Conference on Visualization* (Chichester, GBR, 2009), EuroVis'09, The Eurographs Association & John Wiley & Sons, Ltd., pp. 759–766. https://doi.org/10.1111/j.1467-8659.2009.01476.x.

[ZW18] Zhou L., Weiskopf D.: Indexed-points parallel coordinates visualization of multivariate correlations. *IEEE Transactions on Visualization and Computer Graphics 24* (2018), 1997–2010.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Video S1

Study Data