

Lit Review for handling numerical ties in PCPs

Denise Bradford

Introduction

In modern data analysis, the complexity of datasets with large n (number of observations) and large p (number of variables) presents unique challenges. High-dimensional data, where $p \geq 4$ and $n \geq 50$, is increasingly common in genomics, finance, and social sciences domains. Analyzing such data often requires methods that enhance interpretability while preserving the intricate relationships within the data. However, the scale of observations and variables can obscure meaningful patterns and impede traditional visualization techniques.

Large p datasets are particularly challenging because visualizing relationships across multiple dimensions often leads to information overload, occlusion, and cognitive strain (Keim 2002). Visual representations struggle to maintain clarity when $p \geq 4$, as patterns become increasingly difficult to discern in higher-dimensional spaces (Velagala et al. 2020). Similarly, large n datasets exacerbate these issues by introducing visual clutter, making it harder to track individual observations (Heer and Bostock 2010). These challenges require innovative visualization techniques that simplify complex data while preserving its structural integrity.

Parallel Coordinate Plots (PCPs) have become a powerful way to examine data with more than one dimension. PCPs show each observation as a polyline on parallel axes, letting analysts follow specific data points in multiple directions (Inselberg 1985). By setting factors up as parallel vertical axes, PCPs make it easier to find patterns, trends, and outliers in large datasets. These plots are especially good at showing relationships and trends between multiple factors. However, seeing ties and overlapping data points can be hard when many numbers are the same or very close. This can cause important information to be lost in the research. This study fixes the problems that regular PCPs have with finding numerical ties by adding a new method that makes it easier to tell the difference between data sets that overlap. We want to create a methodology that makes it easier to tell the difference between ties by using a standard delta difference approach. Our suggested methods should make it easier to understand data, leading to more accurate research and decision-making in large, complex datasets.

Most standard data visualizations work within the Cartesian coordinate system, with variables or functions of variables mapped to the x and y axis. Additional variables can be mapped to different properties of the plotted points, bars, or lines, and analysts can also create small multiples that show subsets of the data; even with these additions, viewers quickly become overwhelmed by the amount of information when more than $p = 3$ or 4 variables are shown (including the x and y coordinates). When it is necessary to understand the relationship between more than four variables, visualizations on a Cartesian

grid no longer work as well; extensions to additional dimensions are also ineffective (Inselberg 1997). Other approaches are necessary when it is necessary to understand $p \geq 4$ dimensions of data. Here, we examine parallel coordinate plots (PCPs) as a solution to $p \geq 4$ dimensional data visualization and assess the impact of different modifications of PCPs on their effectiveness for visualizing high N and/or high p dimensional data. We specifically evaluate the ability of each PCP version to facilitate the identification of overall trends, outliers, and clusters within $N \geq 4$ dimensional data across different magnitudes of N (observations) and p (variables).

Parallel Coordinate Plots (PCPs)

Parallel coordinate plots (PCPs) leverage a projective coordinate system, instead of a Cartesian coordinate system: each line in Cartesian space is a set of points in projective space, and each point in Cartesian space can be represented as a line in projection space (Inselberg 1985). The result is that a single data point is represented as a line that crosses each parallel axis representing a variable; clusters, then, appear as a group of lines which have similar paths.

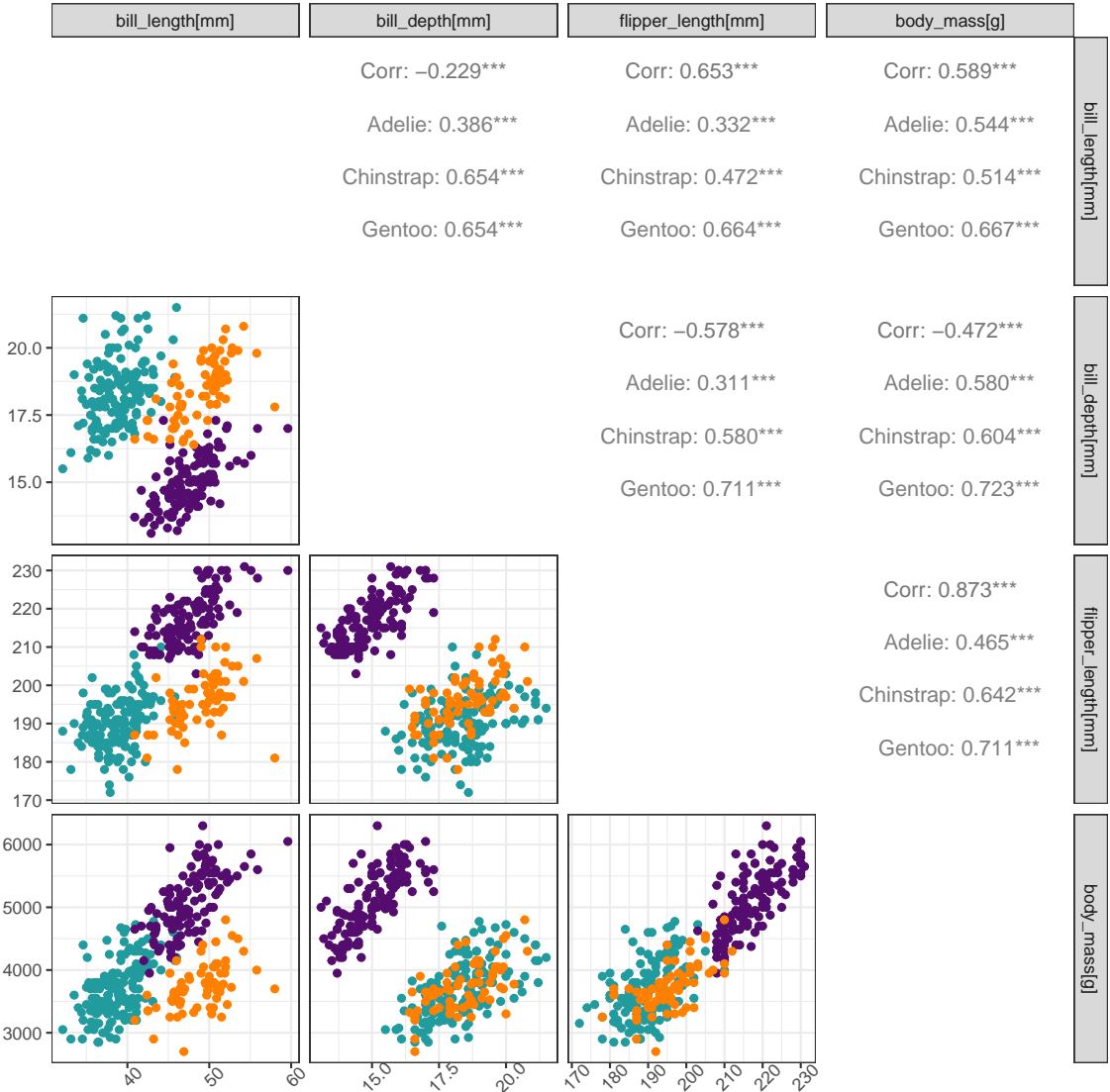


Figure 1: A generalized pairs plot

Figure 1 is a **scatterplot matrix**, which shows pairwise relationships between variables in Cartesian space. It shows how the variables for three kinds of penguins (Adelie, Chinstrap, and Gentoo) are distributed and how they are related.

While some of the information is obscured due to overplotting, the overall relationship between variables is relatively clear and is reinforced by the numerical values displayed in the corresponding pair of variables across the diagonal. Overall, the matrix-style plot does an excellent job of showing the links within and between variables. The major downside to scatterplot matrices is that it is not possible to easily connect a point in one scatterplot to a corresponding point in another; the data representation makes it difficult to get a sense of the multivariate relationships beyond any combination of $p = 2$ variables.

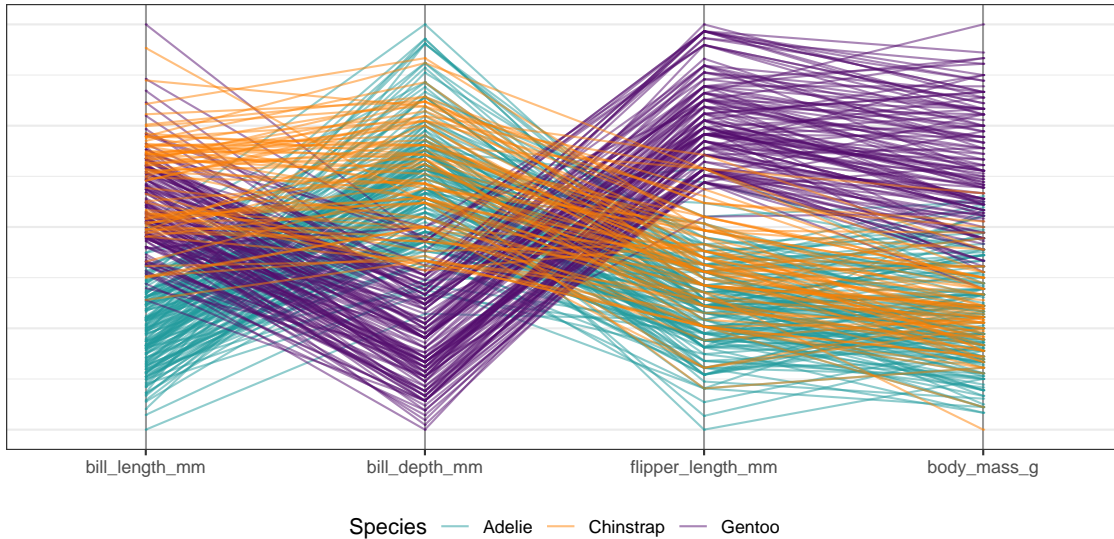


Figure 2: A parallel coordinate plot

In Figure 2, each line shows a different penguin, and the colors of the lines show the species. The Gentoo (blue) tends to have higher body mass and flipper length values, while the Adelie (red) tends to have lower values for these measurements. There is an overall negative relationship between flipper length and bill depth, indicated by the “X” shape of the lines, but within each species, the overall relationship between flipper length and bill depth is positive, as indicated by the largely parallel block of lines of each color. Even in projection space, we can see Simpson’s paradox at work; the second plot in the third row of Figure 1 shows the same basic information. While there is a significant amount of information obscured due to overplotting, it is easier to connect observations across $p > 2$ vertical axes, providing a greater ability to visualize more than two dimensions. At each vertical axis, the plot resembles a rug plot, Cleveland (1993) which can provide some information about the distribution of the variable, but is less direct than a continuous density plot.

As originally proposed, PCPs could be difficult to interpret, in part because PCPs were initially defined only for numeric variables; extensions which treated categorical variables as numeric suffered from overplotting, as different lines converge on a single point and then diverge again, destroying the ability to trace a single observation through the categorical-turned-numerical axis. The ggpcp package (VanderPlas et al. 2023) introduced a new way to handle categorical variables, dividing the axis up into “boxes” and ordering observations within those boxes; for relatively small n , this preserves the ability to follow single observations across the plot, and for larger n , a series of lines moving together converge to form an approximate hammock plot. A demonstration in Figure 3

replicates Figure 2 with the addition of a categorical species axis on each side of the plot. In the InfoVis community, other modifications to parallel coordinate plots have been proposed: smoothed lines, density-based PCPs (J. Heinrich and Weiskopf 2009), bundling of similar points, and other modifications such as interactivity (Johansson and Forsell 2015) may support identification of clusters and outliers in multidimensional space.

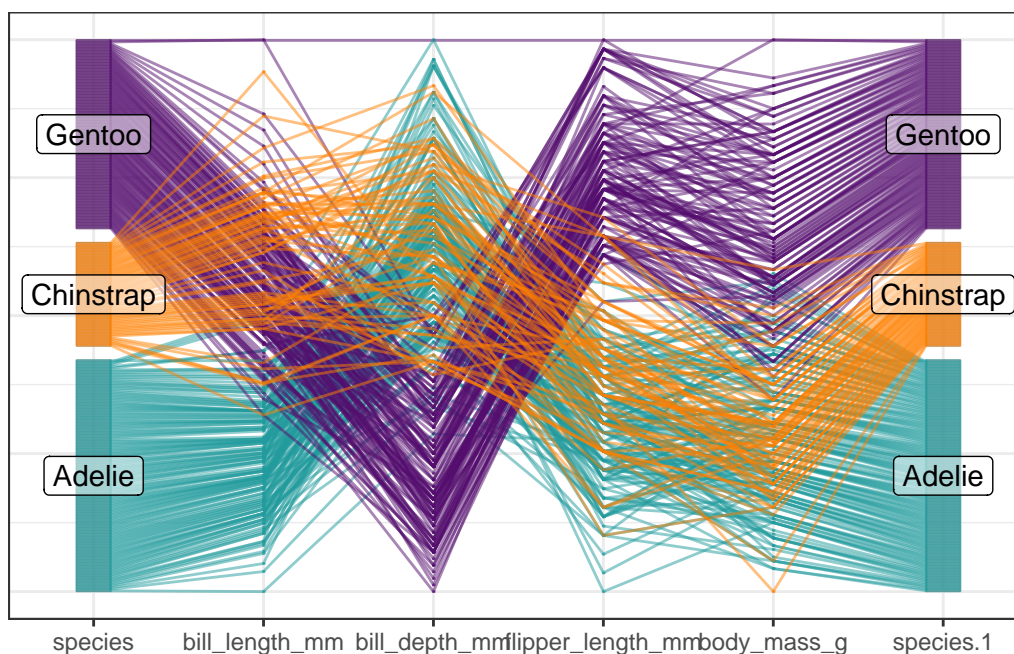


Figure 3: A generalized parallel coordinate plot, with species on the left and right of the plot; observations are ordered on the right side based on the value of body_mass_g, and on the left side based on the value of bill_length_mm. Translucent lines reduce the impact of overplotting and allow perception of the Adelie lines (which are plotted first) even as the Gentoo and Chinstrap lines are plotted on top. With this treatment, we can see that the strong positive relationship between bill depth and flipper length in Gentoo penguins is much less pronounced in Adelie and Chinstrap penguins; there are many more line crossings in both species, which suggests a more moderate relationship.

Why?

Plots using Cartesian coordinates are straightforward, in part because they are commonly encountered and taught in grade school. Unfortunately, plots using Cartesian coordinates, such as scatterplots, are limited to two variables displayed using spatial dimensions, which are perceived more accurately than other aesthetic mappings such as color and shape (Cleveland and McGill 1984). Projections of three-dimensional scatterplots can be created, but these charts can be difficult to read and interpret; interactive 3D scatterplots still suffer from the loss of information inherent to 2D projection on a screen, but allow for some sense of the full shape of the data. Data sets with more than three numerical dimensions are extremely common, but cannot be easily shown using Cartesian coordinates; analysts must resort to strategies like tours (Wickham et al. 2011), dimension reduction (van der Maaten and Hinton 2008; Abdi and Williams 2010), or plotting bivariate relationships between variables in order to visualize these data sets. They make it easier to tell the difference between different information points and their exact values by displaying data points in an easily understood manner. Cartesian coordinates are an excellent method to present data clearly because they can handle up

to three dimensions. They also need help with growth because adding more dimensions requires a lot of subplots or complicated three-dimensional plots, which can get boring and challenging to follow. Cartesian graphs additionally can take up a lot of room, and for data with more dimensions, they usually need more than one plot.

Parallel coordinate plots (PCPs) are an excellent way to show data with many dimensions since each is shown on its plane. PCPs are useful for drawing attention to trends, clusters, and outliers and efficiently show many variables in a small area. However, as with their Cartesian equivalents, PCPs are vulnerable to overplotting, becoming difficult to read and interpret when N is large. In addition, PCPs are a much less familiar form of visualization than scatterplots; there is an initial adjustment period required in order to understand which features of a PCP correspond to familiar features of a scatterplot, as shown in ?@fig-features. Comparisons show that whilst PCPs are more suited for exploratory research when dimensions surpass reasonable limits for scatterplot matrices (Munzner 2014; Inselberg 2009), scatterplot matrices clearly show a depiction of interdimensional interactions for small to medium-scale data.

Although PCPs are sometimes criticized for overplotting and difficulty recognizing patterns, such as clusters or non-linear correlations, which scatterplots more efficiently disclose, PCPs provide a compact depiction of high-dimensional data (Johansson et al. 2005; Qu et al. 2007). Studies show, however, that users adjust to PCPs relatively fast following appropriate introduction since interactive techniques and hierarchical approaches significantly increase their usability (Fua, Ward, and Rundensteiner 1999; Claessen and Van Wijk 2011). When combined with user training and interactive visualization techniques, PCPs remain beneficial for exploratory analysis of higher-dimensional data, even if scatterplots provide instantaneous interpretability and clarity—especially for two-dimensional relationships (Inselberg 2009; Julian Heinrich and Weiskopf 2013a).

Variations on Parallel Coordinate Plots

Since Inselberg (1985) introduced parallel coordinate plots (PCPs) in 1985, considerable advances have been made to address the original method’s framework and improve its capacity to depict high-dimensional data effectively. The enhancements include changes to visual representation, handling various kinds of interactive data elements, and incorporating advanced computational approaches to increase the clarity and interpretability of the visualizations. In this section, we examine some of the modifications proposed to enhance PCPs after their original introduction, as well as variations on PCPs which evolved in parallel for e.g. categorical variables.

Categorical and Hybrid PCPs

Traditional PCPs were primarily designed to collect continuous data. However, real-world datasets often contain a mix of continuous, ordinal, and categorical data. To solve this, changes have been made to display mixed-type data within the same PCP simultaneously. Categorical data, for example, can be visually differentiated using unique colors, symbols, or segmented lines, although continuous variables are still represented by standard lines linking the axes. Categorical Parallel Coordinate Plots (CPCPs) were created to handle the challenge of representing categorical data in PCPs, which are not suitable for continuous axis representation. Specific strategies, like adjustments to the ends of links, have been suggested to enhance how category and numerical values

are connected visually. This strategy enhances the interpretability of mixed data by modifying the plotting criteria for axes representing category variables.

CPCPs use discrete axis segments or unique markers for each category, transforming categorical variables into distinguishable visual elements. Siirtola and R  ih   (2006) introduced axis segmentation, using markers and spacings to represent categorical values distinctly, improving categorical data visualization in PCPs. Inselberg (2009) enhanced this representation by introducing specific encodings, which allowed categories to be differentiated visually, preserving the PCP’s interpretative power. Johansson and Forsell (2015) explored alternative segmentation methods to minimize visual clutter, enabling effective handling of multiple categorical values on the same axis.

Designed to accommodate categorical data using methods including axis restructuring, symbols, and color-coding to increase clarity and comprehension, Categorical Parallel Coordinate Plots (CPCPs) are a development of Parallel Coordinate Plots (PCPs). Looking at the given papers, CPCPs come from more general ideas than from them straightforwardly. Although it does not particularly address categorical variables, Wegman (1990) groundbreaking study on PCPs offers the mathematical underpinning for visualizing high-dimensional numerical data. Every PCP invention, including CPCPs, is built on this work. LeBlanc, Mellor-Crummey, and Fowler (1990) investigation of PCPs for assessing concurrent program performance emphasizes PCP adaptability but mainly focuses on numerical or sequential data rather than categorical data.

Holten and Van Wijk (2009)’s work on force-directed edge bundling aims to improve graph visualizations by lowering visual clutter, which guides the CPCP objectives of increasing interpretability. Though it has nothing to do with PCPs or categorical data, its primary goal is to make complex data visualizations more readable. Therefore, even if these papers specifically address CPCPs, they add to the larger conversation on developing multidimensional visualizing techniques.

Beyond categorical data, the CPCP innovations—axis structure and the use of symbols and colors—are relevant to additional PCP versions. These ideas improve the interpretability of interactive PCRs, hybrid PCRs combining data kinds, and continuous PCRs. CPCPs observe both a tool for categorical data but also larger progress in multidimensional data display.

The Generalized Parallel Coordinate Plot (GPCP) is a modified version of the original PCP that introduces nonlinear changes to the axes for more complex data representation. These can show more complicated relationships in the data that might not be visible in a regular PCP. Nonlinear scales such as logarithmic and exponential transformations are possible. Wegman and Luo (1997a) provide a comprehensive discussion of the GPCP idea stating, “The parallel coordinate plot is a geometric device for displaying points in high-dimensional spaces, in particular, for dimensions above three. As such, it is a graphical alternative to the conventional scatterplot.” The authors enable the visualization of data distributions, correlations, clusters, and higher-dimensional structures by connecting Cartesian and parallel coordinate representations. The “grand tour” idea is important for seeing how dynamic high-dimensional structures change over time because it allows data analysts to examine very large and very complex data sets. This method avoids the problems of axis adjacency in PCPs by constantly rotating the data axes. It ensures that every possible combination of variables is explored dynamically. These new ideas greatly improve the ease and usefulness of PCPs, making them more useful in many situations and giving us tools for exploring and analyzing data in the modern world, (Wegman and Luo 1997b).

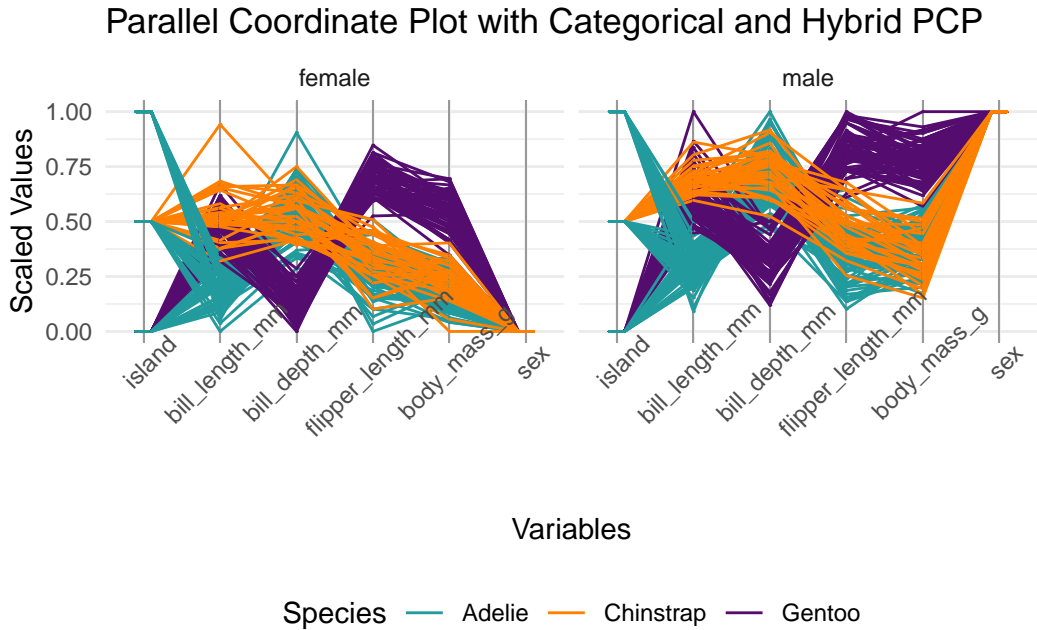


Figure 4: Categorical and Hybrid in PCP

Categorical Variants of PCPs

Parallel sets plots are similar to PCPs, but are designed for categorical variables. Parallel sets are a technique for visualizing categorical data with great promise for assisting visual analytics. Among their various important characteristics and improvements are frequency-based representation, flexibility to mixed data, quality measurements for improved visualization, interactive elements like axis reordering, and dynamic filtering.

For frequency-based representations, parallel sets replace individual data points, abstracting data to emphasize linkages. The SET-STAT-MAP system is flexible for mixed-data situations like environmental and policy studies because it expands parallel sets to include numerical and spatial data, Wang et al. (2022). Quality measures, including overlap, ribbon width variance, and mutual information, optimize dimension and category ordering and improve visualizations' interpretability.

Parallel sets depend on interactivity to allow users to investigate data relationships more naturally. These tools let users create aesthetically pleasing layouts that best fit their application. Across many fields, including meteorology, consumer behavior study, marketing, and resource management, parallel sets find application.

Dennig et al. (2021) introduced "ParSetgnostics," a set of quality metrics to reduce visual clutter in parallel sets. Metrics like overlap, ribbon width variance, and mutual information optimize dimension and category ordering, significantly improving visualizations' interpretability; for instance, applying these metrics reduced clutter by up to 81% in test cases. Parallel sets are also useful for instruction and analysis since their combination of statistical summaries and category flow analysis reveals a deeper understanding. Combining categories and automating axis selections guarantees that parallel sets stay scalable, therefore matching the aim of reducing the number of possible combinations produced on display without compromising data integrity.

Hammock plots are similar to parallel coordinate plots, but were designed to show categorical data (Schonlau 2003). Hammock plots represent a breakthrough in data visualization, particularly for managing numerical and categorical data. While allowing

mixed data types, they are meant to show relationships between categorical variables visually. Especially in health science datasets, key characteristics of Hammock plots include their capacity to simplify complicated relationships and lower visual clutter by using “bandwidth” representations to highlight links between categories.

When it comes to multidimensional categorical data, hammock plots show a more compact representation than mosaic plots, lowering the cognitive load on the observer and surpassing conventional tools like bar charts. These benefits are absolutely vital in disciplines including health research, where datasets can comprise mixed numerical and categorical variables needing simultaneous display.

Symanzik, Friendly, and Onder (2018) showed the Titanic dataset using hammock plots, arguing that “Hammock plots allow for line alignment representing similar or identical values.” This makes understanding the data in parallel coordinate space easier and gives you a better sense of the different data types and their groups. They do say, though, that Hammock plots “may add visual clutter in cases lacking categorical data distinctions,” which is meant to stress the problems that might arise when they are used with continuous variables.

Kavvadias et al. (1996) use a decomposition method called “Hammock-on-ears” to work on bigger graph-theoretic problems, with the goal of enhancing the ability to find ties in large datasets. This shows that Hammock plots can handle small amounts of complex data. However, they warn that “the method may require significant computational resources,” which means it might not work for complex or larger-scale tasks.

Pilhöfer, Gribov, and Unwin (2012a) show that Hammock plots “manage clusters and resolve data ties” well, especially when comparing data clusters. Key strategies comprise clustering visualization, which employs algorithms to reorder rows and columns, obtain a pseudo-diagonal form, and resolve data connections via optimal sorting algorithms like the Weighted Bertin Classification Criteria (WBCC). In graphical displays, these techniques minimize crossing lines or misalignments and maximize concordance. Optimized displays of clustering results, using the “Ecotest dataset,” show how ordering methods enhance cluster identification, Pilhöfer, Gribov, and Unwin (2012b). The paper also offers an evidence-based method, using measures such as the Bertin Classification Criteria (BCC) to evaluate graphical order, supported by computational tests. Both theoretical considerations and actual data support the assertions regarding Hammock plots controlling clusters and tying off conflicts.

Still, Pilhöfer, Gribov, and Unwin (2012a) points out a problem: “The focus on clustering may introduce unnecessary complexity in dispersed data.” They suggest that Hammock plots work best for datasets with clear clusters. Both visual and computational methods help to find definite clustering in the content. The main technique is pseudo-diagonalization, in which rows and columns of a data matrix are rearranged to visually show clusters by combining similar entries. This technique is compared by a top-down partitioning method, which clusters the data using measurements such as Kendall’s τ or the Bertin Classification Criteria (BCC), determining ideal cut spots in rows and columns. The BCC, which computes the difference between conforming and non-conforming pairings to gauge the alignment of matrix entries with a pseudo-diagonal form, is one of the metrics to evaluate clustering clarity. The Weighted Bertin Classification Criteria (WBCC) emphasizes the proximity of data points and improves this by using distance-based weights; the Bertin Classification Index (BCI) scales the BCC to gauge the strength of the link between clusters. BCI provides a consistent approach to evaluating datasets, with values ranging from 0 (perfectly aligned clusters) to 1 (indicating no clustering). The main evaluation criterion for clarity is the degree to which the matrix

resembles a block-diagonal structure; low BCI values imply unambiguous clusters, while high values indicate indistinct or overlapping groupings. The paper also underlines the need for visual inspection, in which visually distinct and well-separated blocks in matrix plots—such as those shown in the Ecotest dataset examples—characterize clear grouping. Still, clarity is very arbitrary since it depends on thresholds selected for BCI and the interpretability of the produced visuals. This method detects and assesses clusters by combining empirical visualization with quantitative measurements.

The `ggparallel` package Hofmann and Vendettuoli (2013) provided a grammar-based interface for visualizing multivariate categorical data, implementing Hammock plots (Schonlau 2024). This makes them better suited for data that has both category and numerical relationships. They say this grammar-based adaptation retains the established behaviors of parallel coordinate while giving you more options for dealing with edge cases like categorical ties.

Hofmann and Vendettuoli (2013) worked on Common Angle Plots, and Schonlau added Hammock plots to improve categorical data display. However, they approached the problem in different ways. Schonlau (2024) Hammock Plots: Visualizing Categorical and Numerical Variables generalizes parallel coordinate plots by swapping lines with box elements. The width of the boxes shows the number of observations, so they can be used for both categorical and mixed data. By changing traditional Hammock plot designs, Schonlau deals with perceptual problems like the reverse line width illusion. Hofmann and Vendettuoli (2013) also addresses the line width illusion, proposing Common Angle plots that use a constant angle to mitigate the line width illusion while maintaining visual continuity of the categories. Hofmann’s Common Angle Plots are unique because they use consistent angular links to fix problems when line slopes or widths are different. This ensures that the visualizations are accurate. Schonlau stresses the importance of being flexible with different data types, while Hofmann uses a consistent angular structure to clarify interpretations. Together, these approaches show a shared desire to make categorical and mixed data displays more accurate and useful, using different but complementary methods.

Fig 5 is a Hammock plot that shows the relationship between two categorical variables, gear and cylinders (cyl). Each color-coded band represents a different category within these variables. The plot provides frequency but lacks precise counts or numeric values. Overplotting is evident, making it difficult to trace individual pathways. The plot shows the general distribution and relationships between gear and cylinder categories, but lacks numeric relationships or exact proportions.

Bundling and Curving

Bundling and curving techniques have emerged as practical solutions for managing visual clutter in high-dimensional data visualizations. These methods group similar paths, reduce overplotting and make overarching patterns more visible. By organizing the chaotic lines often seen in parallel coordinate plots (PCPs), bundling and curving enhance interpretability. As noted by Holten and Van Wijk (2009), bundling provides a compelling strategy to mitigate the visual complexity frequently encountered in these datasets.

Building on this concept, McDonnell and Mueller (2008) introduced curvilinear PCPs, leveraging curved lines to distinguish intersecting paths and improve visual clarity. Subsequently, Johansson and Forsell (2015) assessed the effectiveness of bundling and curving in PCPs, offering criteria for when these modifications are most beneficial for interpretability.

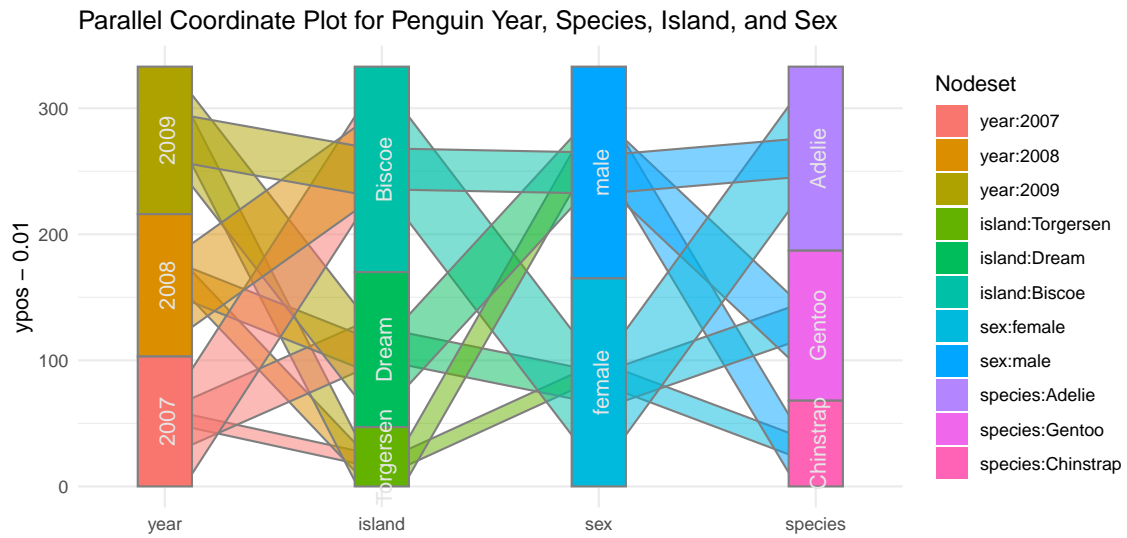


Figure 5: Hammock Plot (Schonlau 2024)

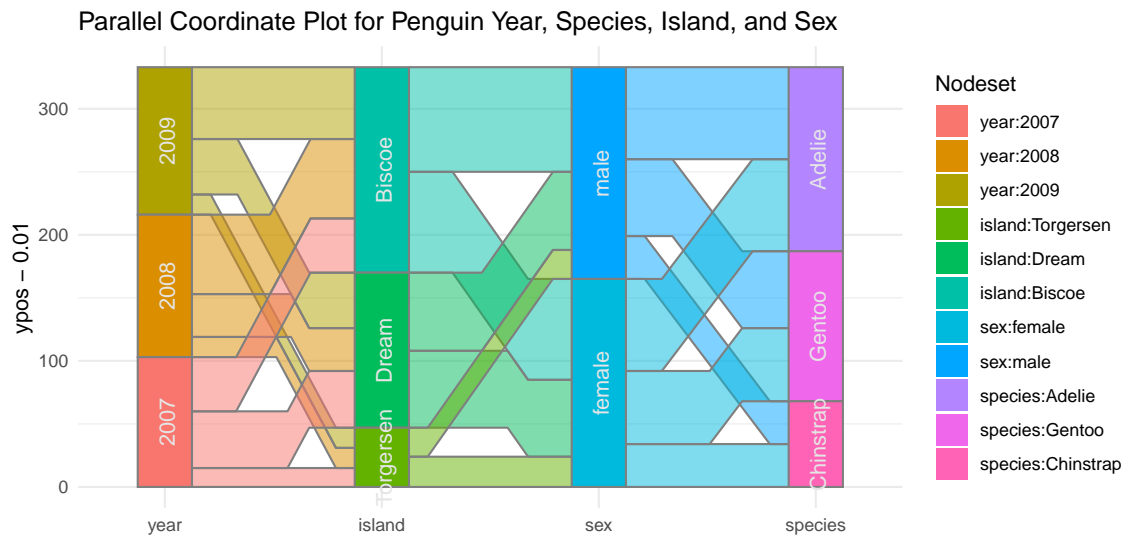


Figure 6: Common Angle Plot (Hofmann and Vendettuoli 2013)

Curving lines in PCPs can be adapted based on data attributes, enabling better visual separation and pattern recognition. This is particularly advantageous when working with strongly correlated dimensions, where overlapping paths obscure relationships. Curved lines visually separate trajectories by reducing intersections, making data relationships easier to discern. Additionally, they enhance aesthetic appeal and reduce cognitive load, improving the overall readability of visualizations. Curved lines offer a more intuitive perception of relative data positions by simulating a three-dimensional effect within a two-dimensional space. Research supports these benefits, with studies showing that curving lines reduces clutter and improves data interpretation Qu et al. (2007). This approach has proven effective in handling complex datasets with overlapping points (Kachway 2013).

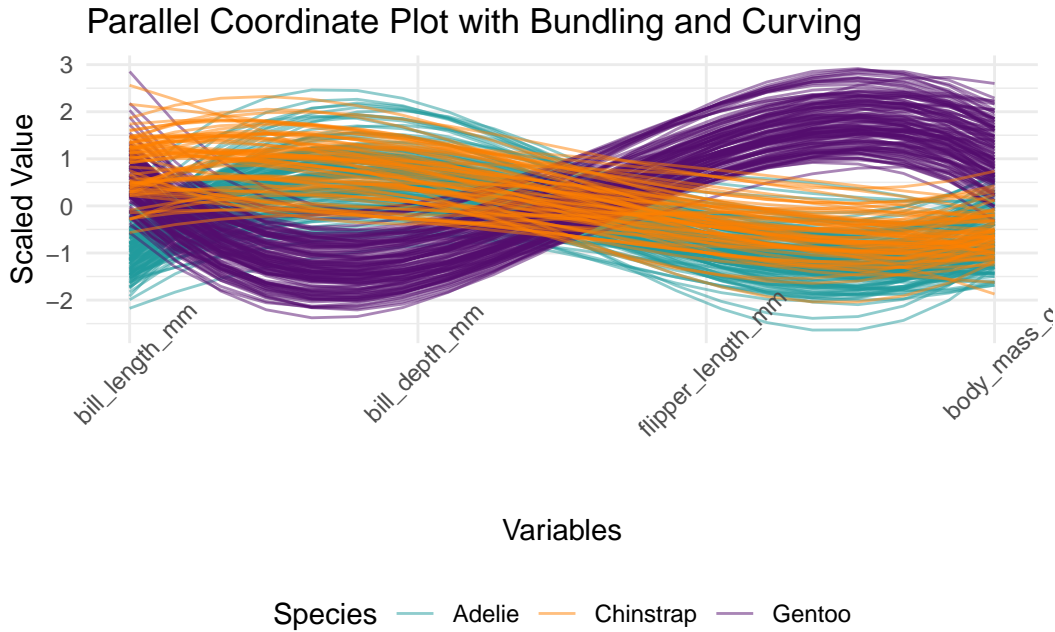


Figure 7: A Bundling and Curving Parallel Coordinate Plot

In Figure 7, a PCP modified with bundling and curving demonstrates how these techniques smooth lines, reduce distractions and highlight patterns across different species. Bundling reveals general trends within each species, but it lacks the detail to show specific distribution shapes, such as those visible in histograms. A notable trade-off of bundling is its tendency to obscure outliers, as extreme values are pulled into the primary flow, making deviations harder to detect. However, curved lines make variable relationships more apparent, indicating that specific measurements rise or fall together depending on the species. Despite this, minor differences between individual points may be harder to discern.

Rida E. Moustafa (2011) expanded on PCPs by integrating them with parallel coordinate density plots (PCDPs) to effectively manage large datasets. By transforming PCP images with density estimation methods, Moustafa’s approach converts polylines into continuous, smooth depictions of data density. This innovation highlights groupings and trends often hidden in traditional PCPs. Moreover, the interactive elements of Moustafa’s methodology allow users to explore different dimensions in real time, making the visualization accessible even to non-experts. This interactivity significantly enhances the analytical capabilities of PCPs.

While bundling simplifies complex datasets and reduces visual clutter, it has limitations. Grouping similar paths can obscure individual data points and outliers, which may be

critical in specific analyses. McDonnell and Mueller caution that the clarity achieved through bundling may compromise data precision, as unique data paths or outliers are often lost in the simplification process (McDonnell and Mueller 2008).

Enhanced Color Encoding and Shading

Visual enhancements such as color coding, opacity modifications, and line thickness adjustments have been implemented to address overplotting, particularly in large datasets. These upgrades allow significant data trends to stand out while less relevant information is minimized. For instance, varying colors or adjusting line opacity based on data density helps users identify trends and outliers that might otherwise remain hidden. Similarly, modifying line thickness to represent variables such as data frequency or confidence intervals adds depth to the visualization. These multi-layered techniques enrich traditional parallel coordinate plots (PCPs) by encoding additional dimensions of information.

In their comprehensive review, *State of the Art of Parallel Coordinates*, Julian Heinrich and Weiskopf (2013a) discusses numerous methods to reduce overplotting and improve data interpretation. Techniques such as alpha blending, which adjusts line transparency to minimize visual clutter, and interactive line manipulation, which allows users to explore data in real time, are highlighted as essential tools for enhancing PCP readability in large datasets.

Building on these foundations, Blaas, Botha, and Post (2008) emphasized the importance of dynamic visual traits like adjustable colors and opacity in their study *Extensions of Parallel Coordinates for Interactive Exploration of Large Multi-Timepoint Data Sets*. Their approach enables real-time modification of visual settings, allowing users to focus on specific data features and control line density. Similarly, Raidou et al. (2015), in *Orientation-Enhanced Parallel Coordinate Plots*, demonstrated how varying opacity levels and color gradients can draw attention to patterns, show line densities, and facilitate interactive exploration. These strategies effectively tackle overplotting challenges in complex datasets by employing techniques such as smoothing and averaging polylines.

The work of Zhou et al. (2008) in *Visual Clustering in Parallel Coordinates* introduced a user-driven approach to highlight data clusters using customizable color and opacity settings. This technique differentiates data groups and reshapes curves to enhance clarity. Likewise, Fua, Ward, and Rundensteiner (1999)’s *Hierarchical Parallel Coordinates for Exploration of Large Datasets* proposed reducing visual clutter through hierarchical representations and variable line thickness. By effectively visualizing data density, these methods reveal trends otherwise obscured in large datasets.

Color gradients and shading have emerged as powerful tools to encode additional data attributes, such as density or frequency, in PCPs. Using these visual encodings, users can identify trends and correlations more effectively. For example, color gradients transform PCPs from purely structural plots into tools for multi-dimensional analysis. As Theus and Urbanek (2008) demonstrated with color-coded PCPs, mapping additional variables through color significantly enhances interpretive power. This approach was further refined by Bertini, Dell’Aquila, and Santucci (2005), who applied density shading techniques to make frequent patterns in large datasets more visible. Novotny and Hauser (2006) expanded on this work with opacity and color-blending techniques, making overlapping patterns easier to interpret intuitively.

However, reliance on color-based encoding poses challenges. Overusing colors or gradients can lead to visual strain and reduce clarity, especially for users with color vision

deficiencies. Bertini et al. noted that excessive color application risks overwhelming the viewer, making the visualization more confusing rather than more insightful.

These studies underscore the importance of adapting visual attributes in PCPs to address overplotting and improve interpretability. By incorporating techniques like color coding, opacity adjustments, and line thickness modifications, PCPs become more effective tools for analyzing multi-dimensional datasets. However, balancing these enhancements to maintain their usability is essential, particularly for diverse audiences. Thoughtful application of these techniques ensures that PCPs remain accessible, clear, and insightful for exploring complex datasets.

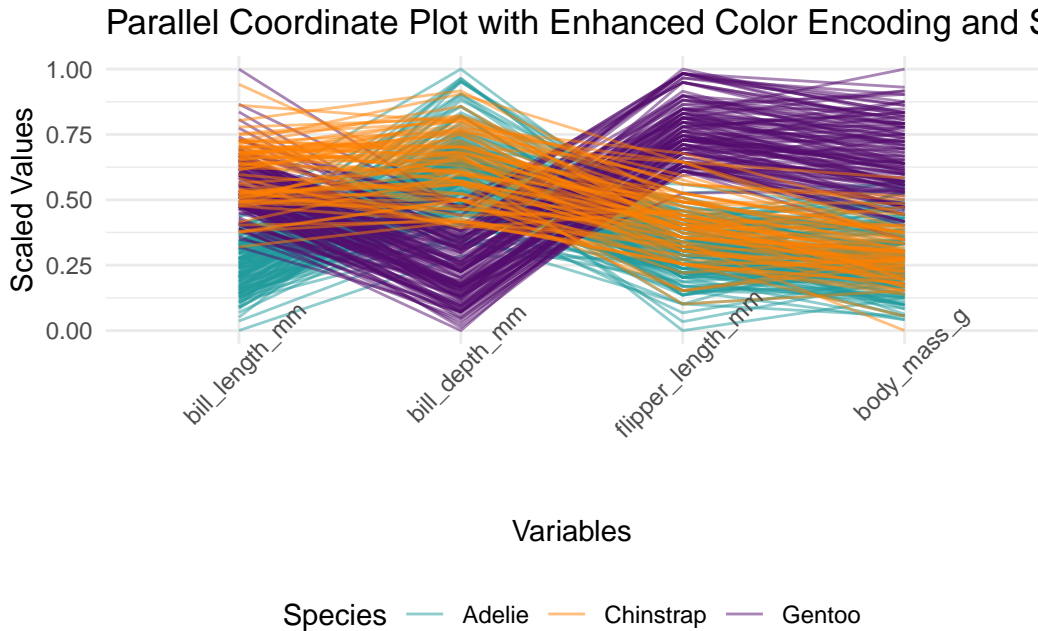


Figure 8: Enhanced Color Encoding and Shading in PCP

The parallel coordinate plot illustrated in Figure 8 employs enhanced color coding to distinguish between species effectively. Each species is represented by a distinct color: red for Adelie, green for Chinstrap, and blue for Gentoo. Adjustments to line density further highlight the differences between species across key variables such as bill length, bill depth, flipper length, and body mass. This approach successfully conveys general trends for each species, such as the Gentoo’s higher flipper length and body mass values than the Adelie, which typically exhibits lower values.

Despite these enhancements, overplotting remains an issue, particularly around the middle ranges of each axis. This visual congestion obscures slight differences and makes identifying outliers challenging. While the plot does not explicitly show the distribution shapes of individual variables, the grouping of lines along each axis hints at central tendencies for each species. For instance, clustered lines suggest shared characteristics within species, though these clusters lack precision in depicting variability.

The general trends revealed in the plot suggest relationships between numerical factors, but the overlapping lines make it difficult to isolate individual correlations. While this visualization provides a broad overview of trends across species, its ability to pinpoint specific data points or detect rare observations is limited. To enhance its utility, further modifications—such as interactive features or additional encoding dimensions—could help overcome these limitations and improve the analysis of individual patterns within the data.

Dimension Reduction Techniques

Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA), are widely used to improve the readability of parallel coordinate plots (PCPs) before visualization. These methods reduce the number of dimensions while retaining the most significant aspects of the data, simplifying analysis without losing critical information. PCPs of reduced dimensions precisely depict high-dimensional interactions, making patterns and relationships easier to discern.

In addition to dimensionality reduction, axis ordering is crucial in enhancing PCPs. The relationships between variables become more apparent by arranging axes based on correlation, mutual information, or other metrics. Automated axis arrangement techniques have been developed to reduce line crossings, significantly improving the identification of patterns. These optimizations make PCPs a more effective tool for exploring complex datasets.

The Orientation-Enhanced Parallel Coordinate Plots technique, introduced by Raidou et al. (2015), represents a significant innovation in making PCPs more accessible and informative. This method addresses common challenges such as clutter and overlapping lines by dynamically adjusting plot axis orientations. By applying both automatic and user-interactive adjustments, this approach reduces visual noise and highlights meaningful data patterns and correlations. Modifying axis orientation on the fly offers users greater control over the visualization, enabling more precise analysis of large datasets.

Orientation-enhanced PCPs incorporate preprocessing steps like PCA and multi-dimensional scaling (MDS) to optimize axis arrangements. These techniques ensure that data clusters are well-separated and that overlapping lines are minimized. Interactive features further empower users to customize the orientation based on their specific research needs, making these plots adaptable to various analytical scenarios (Raidou et al. 2015).

Reducing dimensions has become essential for managing the complexity of PCPs when visualizing high-dimensional data. Techniques such as PCA and clustering allow analysts to focus on the most critical features of the data while reducing visual overload. PCA, for example, condenses the data into components that capture the majority of variance, ensuring essential information is retained. Similarly, clustering simplifies the dataset by grouping similar features, making trends more discernible. Wegman and Luo (1997b) demonstrated the value of PCA within PCPs by enabling users to represent key data features while discarding less relevant dimensions. Further advancements by H. Guo, Xiao, and Yuan (2012) introduced hierarchical dimension reduction and clustering techniques, streamlining the analysis of high-dimensional datasets while preserving usability. As explored by Yang et al. (2017), non-linear methods further enhanced PCPs by capturing complex, non-linear relationships that traditional techniques might overlook.

While dimension reduction offers significant advantages, it has trade-offs. Techniques like PCA or clustering can inadvertently remove subtle nuances or minor patterns, potentially masking relationships that exist only in higher-dimensional spaces. H. Guo, Xiao, and Yuan (2012) cautioned that these beneficial methods might lead to information loss, particularly for niche analyses where fine-grained details are crucial.

By balancing the need for simplification with careful consideration of the data's nuances, PCPs can effectively represent high-dimensional datasets while maintaining their interpretive power. Combining dimensionality reduction, axis optimization, and user-interactive features ensures that PCPs remain robust tools for exploring and analyzing complex data.

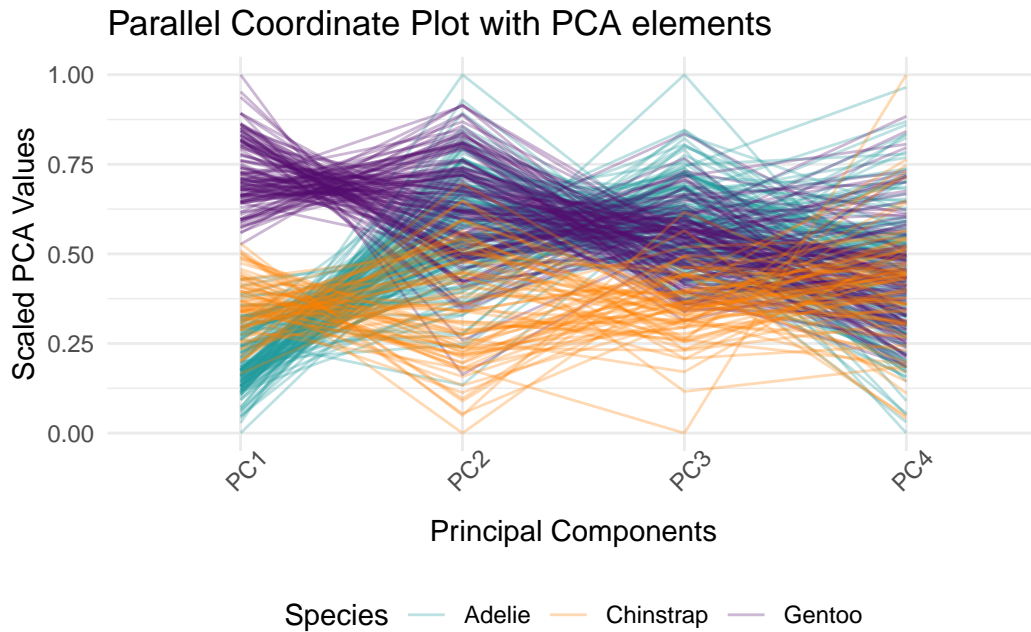


Figure 9: A PCA in Parallel Coordinate Plot

In Figure 9, the parallel coordinate plot leverages principal component analysis (PCA) to simplify high-dimensional data. The x-axis represents PCA components (PC1 to PC4), while the y-axis displays scaled PCA values, color-coded by species: Adelie (red), Chinstrap (green), and Gentoo (purple). PCA encapsulates complex relationships into fewer dimensions by reducing the original variables into principal components, allowing for a more precise depiction of overall trends.

The lines connecting the components highlight how each penguin transitions across these derived features, making species-level patterns more discernible in the reduced dimensional space. Despite this, overplotting persists due to the numerous overlapping lines, particularly in densely populated regions. However, distinct separations between species emerge along specific principal components, showcasing PCA's utility in revealing broad group-level trends.

Clusters around the PC axes indicate each species' general concentration of values, hinting at their central tendencies in the reduced space. While explicit distribution shapes are not depicted, these clusters offer insight into species-specific patterns. Color-coded lines and groupings emphasize the relationships between PCs and species, suggesting that specific principal components contribute to species separation. However, the bundling effect of overlapping lines makes detecting outliers and understanding individual variability more challenging.

This visualization demonstrates the potential of PCA in reducing dimensional complexity and enabling the identification of trends in large datasets. While PCA's abstraction simplifies relationships, trade-offs remain in preserving fine-grained details and outlier visibility, underscoring the importance of complementary techniques for in-depth analysis.

Cluster-Based and Hierarchical PCPs

Clustering techniques, such as spectral clustering, have proven instrumental in improving the interpretability of parallel coordinate plots (PCPs) by grouping related data points

and revealing underlying structures in high-dimensional datasets. These methods allow analysts to uncover patterns and relationships not immediately evident in raw data, providing insights into the natural grouping of data points (Zhao and Kaufman 2012). Visualizing clusters in PCPs makes connections between data points and cluster features more apparent, facilitating a deeper understanding of complex datasets.

Before generating PCPs, data reduction techniques such as hierarchical clustering and principal component analysis (PCA) are often employed. These strategies focus on the most significant components that explain the majority of the variance in the data, limiting the number of dimensions visualized and reducing visual clutter. This preprocessing step enhances the clarity and utility of PCPs, mainly when dealing with large, multidimensional datasets.

The study *Structure Revealing Techniques Based on Parallel Coordinates Plot* by Zhao and Kaufman (2012) highlights the limitations of traditional PCPs in uncovering important patterns within high-dimensional data, particularly when faced with challenges like overplotting. To address these issues, the authors propose clustering and sorting techniques designed explicitly for PCPs. By leveraging spectrum theory, algorithms are developed to group similar polylines, reducing visual clutter and exposing hidden trends and correlations.

Additionally, the study introduces a correlation-based sorting method to arrange axes in a way that highlights relationships between variables. This systematic approach enables the visualization of trends across dimensions, making it easier to interpret complex data. Including view-range metrics further improves data clarity by aggregating and limiting noisy datasets. Experimental results demonstrate that these enhancements significantly improve PCPs' ability to reveal valuable patterns, trends, and correlations, enabling more efficient data analysis.

For multidimensional datasets, clustering, and hierarchical PCPs provide structured and organized visualizations, making comparing clusters and identifying trends easier. Hierarchical clustering enhances the narrative of PCPs by presenting high-level overviews and detailed views, as argued by Fua, Ward, and Rundensteiner (1999). This dual perspective allows analysts to observe the big picture while examining finer details within subsets of the data.

Fua, Ward, and Rundensteiner (1999) pioneered hierarchical PCPs with a zoomable interface, enabling users to explore data clusters and delve into specific subsets for detailed analysis. Geng, Deng, and Ali (2005) expanded on this concept by applying clustering to group similar data points, highlighting patterns, and making general trends more accessible. Further advancements by Poco et al. (2011) incorporated hierarchical clustering with user-defined granularity levels, providing flexible and interactive visualizations tailored to analytical needs.

While clustering simplifies data visualization and enhances the accessibility of patterns in PCPs, it comes with trade-offs. The grouping of data points can obscure individual data, particularly outliers. Geng, Deng, and Ali (2005) observed that clustering effectively identifies general patterns but may mask unique or significant findings hidden within homogenous groups. To strike a balance, analysts must carefully weigh the benefits of clustering against the potential loss of detail, ensuring that the visualization supports broad overviews and detailed investigations.

By integrating clustering techniques, hierarchical approaches, and correlation-based sorting methods, PCPs have evolved into more powerful tools for exploring high-dimensional

datasets. These advancements enhance the visualization’s ability to expose hidden structures while maintaining flexibility for detailed analysis, making PCPs indispensable in modern data analysis workflows.

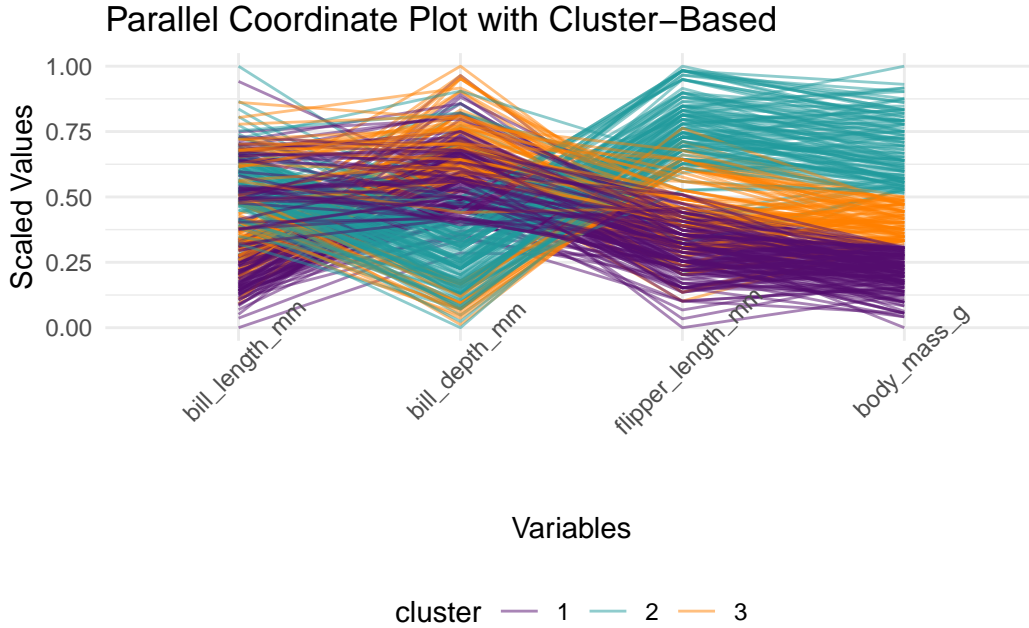


Figure 10: Cluster-Based PCP

Optimized Layout Algorithms

Layout algorithms for parallel coordinate plots (PCPs) have been refined to improve clarity by minimizing line crossings, which is critical when visualizing high-dimensional data. An optimized layout enhances the readability of PCPs by reducing visual clutter and facilitating smoother navigation. Effective design often determines whether a PCP is comprehensible or chaotic (Ankerst, Berchtold, and Keim 1998). In an empirical analysis using three visualization techniques, including PCPs, researchers compared conventional sequential layouts with similarity-based arrangements. Their findings revealed that similarity-based configurations significantly enhanced readability by grouping comparable dimensions, making patterns, correlations, and functional connections more apparent. This alignment of similar dimensions allows for more precise identification of trends, underscoring the impact of thoughtful axis arrangement.

Heuristic layout algorithms in PCPs often generate results misaligned with specific user-defined analytical needs (Johansson et al. 2005). To address this limitation, Johansson et al. (2005) introduced an optimized axis positioning algorithm to maximize spacing between overlapping lines, thereby improving data distinction. In this context, data distinction refers to the ability to expose significant structures within dense PCPs, such as cluster formations, local outliers, and intensity fluctuations.

Johansson’s work incorporates clustering, high-precision texturing, and transfer functions (TFs) to highlight multiple data properties. These innovations allow users to discern cluster structures, detect local outliers, and examine intensity variations in large datasets. Techniques like logarithmic TFs reveal dense structures and subtle patterns, while variance-based feature animation highlights areas requiring closer analysis. These methods, evaluated on datasets with up to 100,000 data points, demonstrate the capacity of PCPs to visualize complex trends and sub-clusters efficiently.

Two notable approaches for optimizing PCPs include the Similarity Clustering Approach (Ankerst, Berchtold, and Keim 1998) and the Clustered Parallel Coordinates with High-Precision Textures (Johansson et al. 2005). Both methods enhance interpretability by focusing on dimensional relationships, albeit with differing priorities.

- Similarity Clustering Approach (Ankerst, Berchtold, and Keim (1998)): This method emphasizes arranging dimensions based on functional similarities and correlations rather than explicitly reducing line crossings. Grouping related variables facilitates the discovery of patterns and connections, enhancing interpretability in datasets with strong inter-variable relationships. However, it is less effective for categorical data where reducing line crossings is critical.
- Clustered PCPs with High-Precision Textures (Johansson et al. (2005)): This approach addresses visual clutter by aggregating data points into clusters and representing them with high-precision textures. Transfer functions and outlier detection reduce overplotting and highlight significant data features without altering dimension arrangements. While effective for trend and structure visualization, it does not tackle dimension reordering to minimize crossings.

The **ggpcp** method minimizes line crossings in PCPs, particularly for categorical variables. It provides a direct solution for visual clutter by optimizing variable order, making it ideal for datasets where reducing overlap is a priority. However, its focus on categorical data limits its applicability to broader multidimensional contexts.

In contrast, the methods proposed by Ankerst, Berchtold, and Keim (1998) and Johansson et al. (2005) address broader interpretive challenges in PCPs. The former prioritizes grouping similar dimensions to highlight functional correlations, while the latter enhances visual representation through clustering and advanced texturing techniques. These approaches suit numerical and high-dimensional datasets where trends and relationships require detailed exploration.

Each method offers unique strengths and is best suited for specific use cases. The Similarity Clustering Approach and Clustered PCPs excel in revealing trends and relationships in numerical data, while **ggpcp** is better equipped to handle categorical data by reducing line crossings. These methods illustrate the diverse strategies for optimizing PCPs, ensuring their adaptability to various analytical needs and datasets. While they share a common goal of improving interpretability, their distinct methodologies cater to different aspects of the visualization challenge, from structural clarity to categorical simplification.

Interactive and Dynamic

The integration of interactive features has significantly improved the utility and flexibility of parallel coordinate plots (PCPs). These enhancements allow users to dynamically reorder axes, filter data based on specified criteria, and invert axes to view data from different perspectives. Brushing and connecting highlight specific data points across axes while filtering reduces clutter by hiding irrelevant lines. These capabilities facilitate real-time data analysis, enabling users to uncover patterns and correlations intuitively. Interactive PCPs are particularly valuable for exploratory data analysis, especially when dealing with large and complex datasets.

Julian Heinrich and Weiskopf (2013a)’s State of the Art of Parallel Coordinates provides a comprehensive review of PCP visualization techniques, categorizing them into a taxonomy of approaches. The study highlights modeling methods, visualizing, and interacting

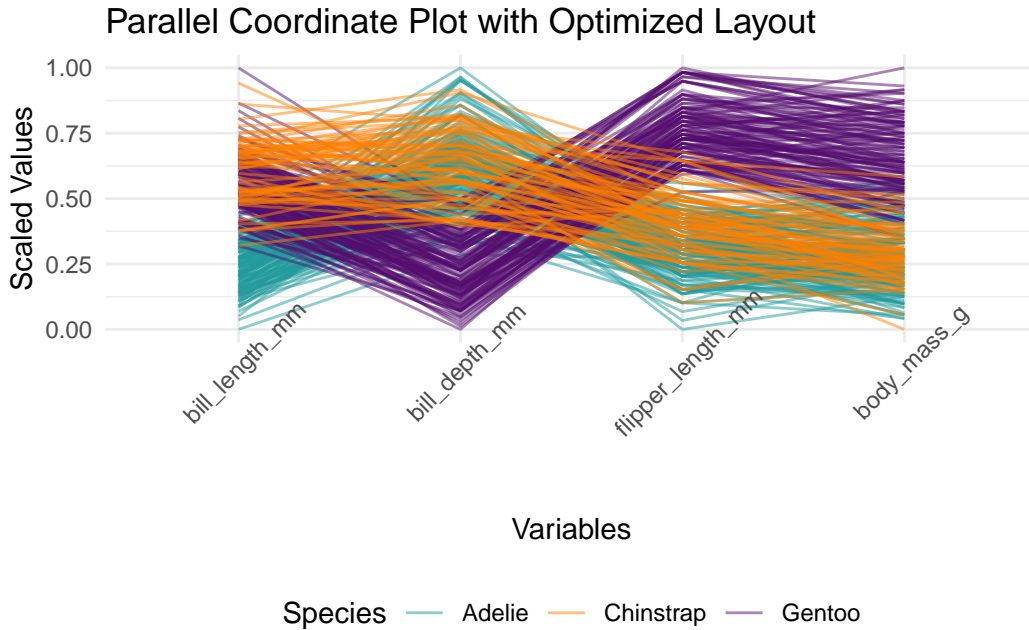


Figure 11: While the PCP employs an optimized layout, its clarity is compromised by overplotting, lack of clustering or feature enhancement, and minimal reduction in line crossings. Clustering, high-precision textures, and transfer functions could greatly improve readability by simplifying data representation and directing attention to key features.

with PCPs, emphasizing their applications in knowledge discovery tasks like sorting, clustering, and regression. Key advancements include geometric modeling, interpolation techniques, point-line duality, and foundational principles of PCPs. Further innovations discussed include density-based visualizations, improved axis ordering, and interactive techniques such as brushing and bundling, which address challenges like overplotting and poor axis arrangements.

The study also explores practical applications of PCPs in engineering and life sciences, underscoring their adaptability for high-dimensional data analysis. Techniques like clustering, density estimation, and interactive filtering enhance data exploration, enabling users to map patterns through curves, shapes, and density plots.

Interactive features in PCPs have revolutionized how users interact with high-dimensional datasets. Tools like brushing, linking, and filtering allow precise and focused analysis. These capabilities help users engage intuitively with data, selectively highlighting areas of interest to identify patterns and generate hypotheses. Real-time interaction, such as reordering axes or filtering data points, makes large datasets more manageable and supports discovery.

The early groundwork for interactive PCPs was laid by Wegman (1990), who introduced fundamental techniques for multidimensional data interaction. Building on this foundation, Siirtola and R  ih   (2006) implemented brushing and linking to improve dimensional analysis, allowing for easier comparisons across attributes in large datasets. Inselberg (2009) further expanded these tools, introducing dynamic filtering and axis rearrangement, enabling users to tailor PCP visualizations to their needs.

While interactivity enhances insight, it can also introduce challenges. Increased complexity can overwhelm novice users and demand higher computational resources, potentially slowing analysis in high-dimensional datasets. As Inselberg (2009) noted, the cognitive

load of multi-step filtering and interaction can lead to user fatigue. Addressing this balance requires careful interface design to ensure that interactive features remain accessible and intuitive.

Modern implementations, such as the live parallel coordinate plots on Plotly’s ggplot2 platform, demonstrate how interactivity can streamline data analysis. Users can filter and separate data ranges by moving their mouse across parallel axes, with tooltips providing detailed information about individual data points. Drag-and-drop axis reordering allows users to adjust dimensions to uncover hidden patterns and correlations. Brushing options dynamically emphasize relevant data paths, making trends and outliers visible while downplaying less critical information (“Parallel Coordinates Plot in Ggplot2,” n.d.). [Interactive Parallel Coordinate Plot](#)

Interactive PCPs continue to evolve, offering increasingly intuitive and adaptable methods for high-dimensional data exploration. By combining dynamic features like axis reordering, filtering, and brushing, PCPs enable users to extract meaningful insights from complex datasets. However, balancing usability with computational demands and minimizing user fatigue remain ongoing challenges. Future developments will likely focus on enhancing user interfaces and optimizing algorithms to make interactive PCPs more efficient and accessible for a broader range of users and applications.

Reordering and Axis Flipping

Adaptive reordering and axis flipping are essential techniques for simplifying the analysis of multidimensional relationships in parallel coordinate plots (PCPs). Reordering aligns highly correlated or thematically related dimensions, minimizing intersections and reducing visual clutter. When implemented effectively, it transforms PCPs from a tangled mass of lines into a clear roadmap of relationships.

Axis flipping complements reordering by addressing negative correlations between adjacent dimensions. Inverting the scale of one or more axes eliminates unnecessary line crossings, revealing patterns that might otherwise remain hidden. Together, these methods improve the interpretability of complex datasets, making relationships and trends across dimensions more discernible.

Early works, such as Inselberg and Dimsdale (1990), proposed axis reordering to enhance the interpretability of PCPs, especially for datasets with highly interrelated variables. Building on this foundation, LeBlanc, Mellor-Crummey, and Fowler (1990) introduced correlation-based reordering, which positions axes to align more related dimensions and reduce visual noise. Later advancements by Peng et al. (2004) incorporated automatic reordering algorithms that adjust axis positions and flip orientations based on user-defined weights, providing greater flexibility for tailoring PCPs to specific analytical needs.

Dynamic reordering techniques optimize the alignment of correlated variables, enabling more precise identification of associations and trends (Yuan et al. 2009). By grouping linked variables and separating unrelated ones, reordering reduces the visual noise caused by crossings between weakly connected variables. Researchers Johansson et al. (2005) and Wegman and Luo (1997b) found that well-designed reordering algorithms reveal patterns, clusters, and outliers more effectively while mitigating overplotting. These methods are particularly valuable for analyzing datasets with numerous dimensions, where clear visual grouping can facilitate discovery.

Axis flipping addresses challenges associated with negative correlations by inverting the scale of one or more axes, reducing unnecessary line crossings, and improving clarity. This adjustment is beneficial when adjacent axes contain negatively correlated factors, revealing trends and relationships that standard PCPs might obscure. Automated axis flipping algorithms, such as those explored by Dasgupta and Kosara (2010), dynamically adapt plots based on data properties, reducing the cognitive burden on users while improving interpretability.

Flipping axes enhance the visual representation of negative correlations and help uncover subtle patterns in datasets with unrelated or oppositely related variables. By turning “problematic” relationships into familiar shapes, axis flipping clarifies connections and aids in exploring complex datasets. However, frequent flipping can disrupt the analytical flow, as users may need help maintaining a stable mental model of the data structure (Julian Heinrich and Weiskopf 2013b).

While automatic reordering and axis flipping greatly enhance PCP readability, they can sometimes misalign dimensions critical to specific analytical questions. LeBlanc et al. cautioned that automated ordering might sacrifice user intent by prioritizing generic readability over context-specific relevance. Continuous reconfiguration of axes can also introduce inconsistency, challenging users to track relationships as layouts change dynamically. These methods offer limited benefits for datasets with minimal correlations, as reordering may not significantly clarify insights.

Reordering and axis flipping should balance automation with user-defined parameters to address these concerns. Allowing users to influence axis arrangement ensures the visualization aligns with their analytical goals. Tools like user-adjustable weights for reordering algorithms and toggle options for axis flipping provide this flexibility, enabling users to tailor PCPs to their needs without compromising clarity or interpretability.

Reordering and axis flipping are complementary techniques that address overplotting in PCPs. Reordering aligns correlated variables to reduce crossings while flipping axes minimizes interference in negatively correlated dimensions. When used together, these methods create a cleaner, more interpretable visualization, making it easier to detect patterns and explore relationships in large datasets (Healey and Enns 1999).

Studies by Julian Heinrich and Weiskopf (2013a) show that combining reordering and flipping significantly enhances the accuracy and usability of PCPs, improving their ability to reveal groups, trends, and outliers. These advancements are critical for data analysts, enabling more informed decision-making based on visual insights. However, the potential for disruption caused by continuous reordering highlights the importance of thoughtful implementation. Ensuring stability and clarity in PCPs requires balancing dynamic adjustments with consistency, helping users build a reliable mental model of the data.

By integrating these techniques, PCPs become powerful tools for exploring and understanding multidimensional data, providing analysts the flexibility and precision to uncover meaningful insights.

In Figure 12, the parallel coordinate map changes the variables by rearranging them and flipping the axes to show patterns and connections between the penguin species more clearly. By flipping some axes and putting related factors closer together (like body mass and flipper length), the plot makes species-based trends stand out more: The Gentoo (purple) tends to have a shallower bill but a bigger body and longer flippers, while the Adelie (pale blue) tends to have the opposite pattern. Overplotting is still a problem, especially when values are close together in the middle ranges, making it hard to tell the difference between lines or peaks. You can’t see the distributions’ exact forms, but

Parallel Coordinate Plot with Reordering and Axis Flipping

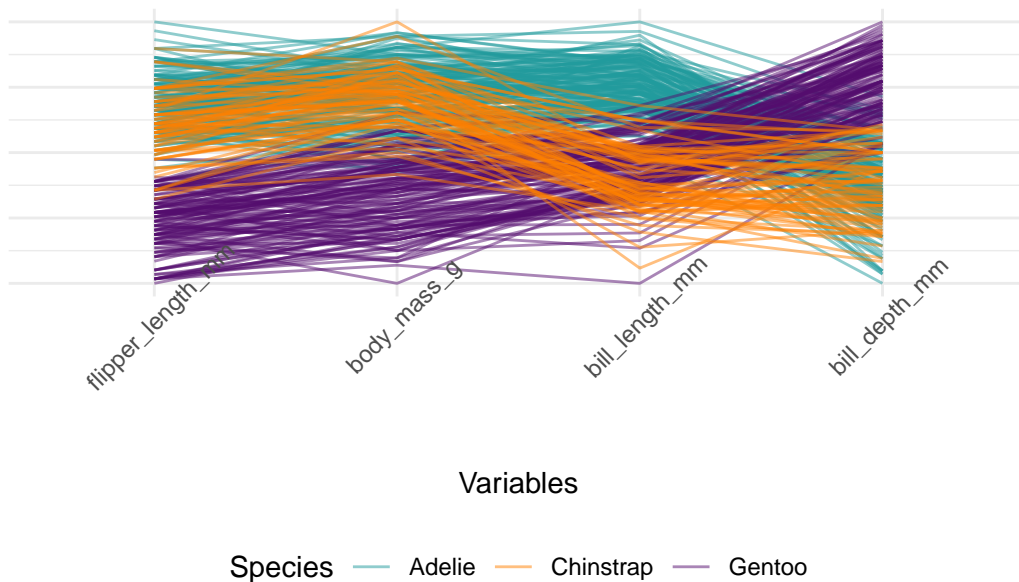


Figure 12: Reordering and Axis Flipping in PCP

the lines' density and spread along each axis give you a good idea of the range of values for each species. The new order makes relationships between variables stand out more because the changes between the axes make species-based correlations in measurement trends more apparent. However, individual relationships are still hard to see because lines cross.

Handling Numerical Ties

Getting rid of ties in statistical data analysis has been hard for a long time for experts who want to make visualizations more clear. Previously, statistical methods focused on rank-based approaches and binary data handling. They often needed help with tied observations, which threw off distributions or made them more complex. Earlier methods for dealing with ties depended on changing ranking systems or adding rules to break ties, which could make results less accurate or easier to understand. As graphics data analysis grew, pioneers like Tukey, Chambers, and Cleveland created new ways to deal with ties visually. There are now several ways to deal with ties, such as slowly increasing the space between values. This makes the visualization more accurate and easier to understand without changing the structure of the data itself.

Tufte (2001) work, particularly in “The Visual Display of Quantitative Information”, is grounded in empirical evidence that shows how visual noise, any element that complicates or crowds the visual field, can impede the interpretation of data. He advocates for “graphical excellence,” a design philosophy that presents data most clearly and efficiently. According to Tufte, separating tied values can reduce overlapping lines and decrease visual noise, making it easier for users to see patterns and accurately compare variables. Distancing tied values in complex visualizations can reduce interpretative difficulty by reducing visual noise, aiding rapid perception of quantitative relationships.

In complicated graphs like parallel coordinate plots, tied numbers often cause lines to overlap and hide other data points. Tufte calls this “chartjunk.” He says that even minor cuts in visual noise can make it much easier for people to process information quickly

and correctly. For example, if an extensive dataset has many similar or similar values, separating the tied values can help the viewer see trends and relationships without having to think too hard.

Jittering Points

Random Jittering of Data Points

Random jittering of data points is a standard method that adds a small random value to each tied observation. This is especially helpful in scatter plots and dot plots to keep them from being too dense. This approach, discussed in “Graphical Methods for Data Analysis,” is easy to use and keeps the overall distribution, making things clearer without changing the data (Chambers 1983). One problem with random jittering is that it can hide exact numbers, making it hard to understand if the jittering distance is too big. Experiments show that jittering makes visualizations much clearer when there are small overlaps, especially when used rarely.

In random jittering, we add a small random value (usually drawn from a uniform or normal distribution) to each tied observation to differentiate them.

For a set of tied values x_1, x_2, \dots, x_n , we define:

$$x'_i = x_i + \epsilon_i$$

where:

- x'_i is the jittered value of x_i ,
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$,
- σ is the standard deviation for the normal jitter.

Parallel Coordinate Plot with Random Jittering

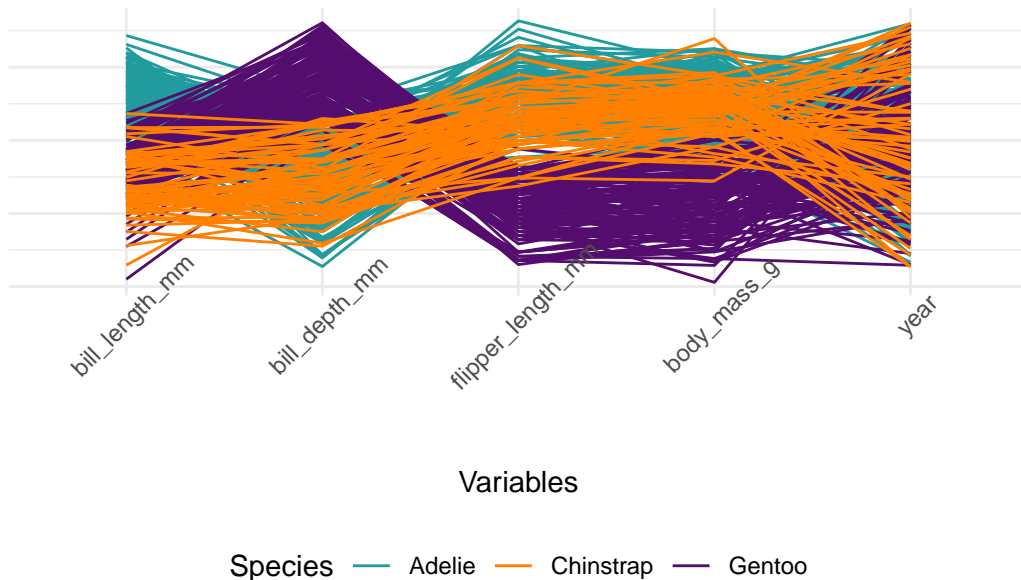


Figure 13: Random Jittering of Data Points in PCP

Rank Jittering (Rank Adjustment)

Rank jittering is another useful method for non-parametric analyses. It changes the ranks of tied values instead of the real data values. This makes it easier to see the ranks in non-parametric tests like the Wilcoxon signed-rank test. Conover (1999) “Practical Nonparametric Statistics” shows that rank jittering keeps the ordinal form of data while not changing the distribution much. The best thing about rank jittering is that it keeps rank-based readings while making ties less noticeable. However, the method can be hard to run on computers when dealing with big datasets, and making too many rank changes can accidentally change the test significance levels. For these reasons, it is important to be careful when using rank jittering.

In rank jittering, we adjust the ranks of tied observations by adding a small random or systematic jitter to each tied rank. This is usually applied to data that will undergo rank-based testing.

for tied ranks R_1, R_2, \dots, R_n :

$$R'_i = R_i + \delta_i$$

where:

- R'_i is the jittered rank of R_i ,
- $\delta_i \sim \mathcal{N}(0, \sigma_{rank}^2)$
- σ_{rank} are small values ensuring that the adjustment is minor.

Parallel Coordinate Plot with Rank Jittering

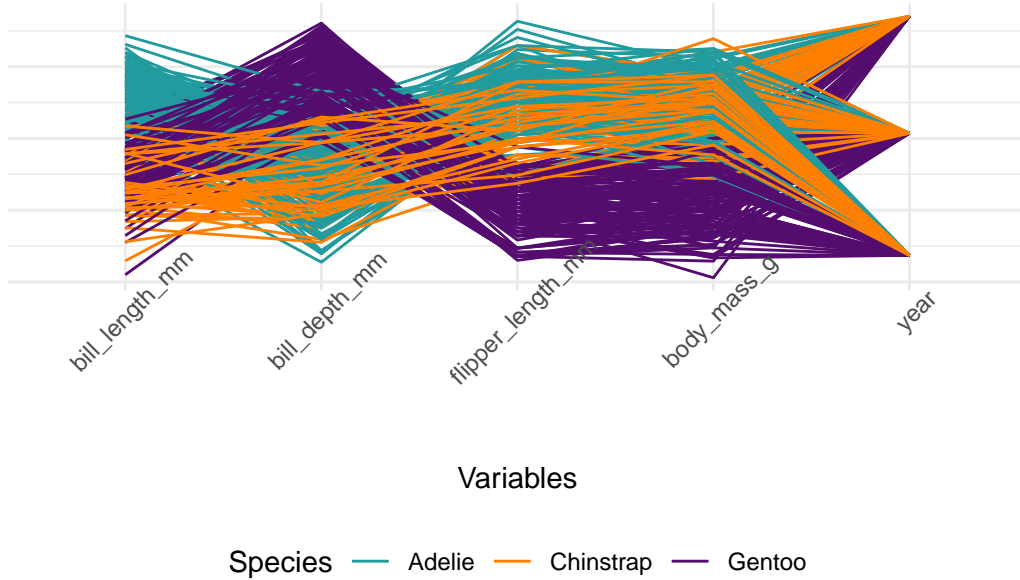


Figure 14: Rank Jittering of Data Points in PCP

A systematic way to use binary relationships in the data is to use ranking differences to break up ties. Ipkovich, Héberger, and Abonyi (2021) writes that rank-based separation ensures that each tied observation is positioned proportionally to its relative rank, maintaining interpretability and visual clarity. This is one way to get rid of visual clutter in PCPs. This rank-based method assigns spacing factors based on rank differences to ensure that overlapped lines are spread evenly across the plot. The method works well to reduce overlap and maintain the order of the items, but it can be very time-consuming

to run on big datasets with many ties. In any case, it works incredibly well for showing categorical or ordinal data where order is important.

Deterministic Jittering (Fixed Perturbation)

When you need to be clear, deterministic jitter gives you a structured way to do it by adding a small fixed value (epsilon) to each tied observation instead of depending on chance. This method explained in more detail in David and Tukey (1977) “Exploratory Data Analysis”, is useful when random noise is not wanted because it ensures that each observation is moved in a way that can be predicted and repeated. Deterministic jitter is easier to understand than random jitter, but it needs to be done carefully so that it doesn’t introduce fake patterns, especially when scaling is involved. Deterministic jittering is a good choice for random noise that doesn’t affect the integrity of the distribution when used carefully.

In deterministic jittering, a small, fixed value ϵ is added to each tied value to systematically spread them apart without introducing randomness.

For tied values x_1, x_2, \dots, x_n , we use:

$$x'_i = x_i + i\epsilon$$

where:

- x'_i is the adjusted value of x_i ,
- i is the index of the tied observation (e.g., $i = 1, 2, \dots, n$),
- ϵ is a small fixed distance chosen based on the data scale.

Parallel Coordinate Plot with Deterministic Jittering

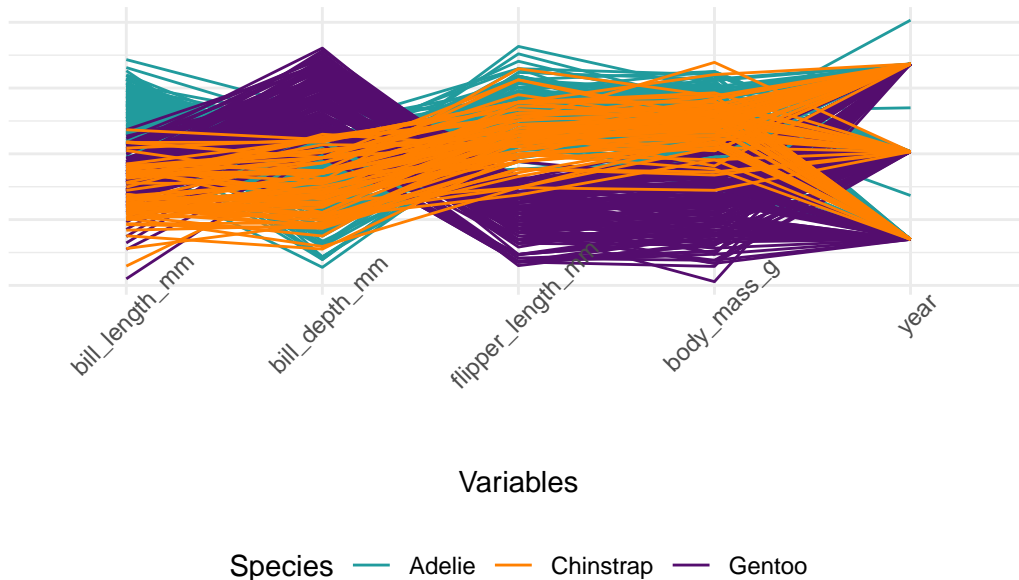


Figure 15: Deterministic Jittering of Data Points in PCP

Jittering can be a common way to deal with ties in parallel coordinate plots (PCPs). It involves adding small amounts of random or predictable noise to tied observations. Swayne et al. (2003) handle ties by jittering as an intuitive way to separate tied values in 2-D correlation tours without distorting their statistical relationships. While jittering

makes things easier to see, its main flaw is that it needs to be easier to figure out the right amount of noise. Similarly, Few and Edge (2008) calls jittering a practical solution to the perceptual challenges caused by overplotting. They point out that it can tell the difference between lines that overlap while keeping the structure of the whole picture. Their review of the current solution stresses that jittering can improve visual clarity in dense datasets, mainly when carefully regulated.

Mean or Median Splitting

One more popular method is mean or median splitting, which gives values tied to their mean or median values. This method works especially well for symmetric distributions and is perfect for images that need to show things centrally, like box plots. Wilcox (2017) “Modern Statistics for the Social and Behavioral Sciences” shows that splitting the mean or median gives a fair view, especially for boxplot data. This method works best with datasets with few ties because it can hide the range and variability within tied numbers if used too much.

In mean or median splitting, we replace each tied value with the mean or median of the tied group. This technique centers tied observations on their central tendency.

For set of tied values x_1, x_2, \dots, x_n let:

$$x_{mean} = \frac{1}{n} \sum_{i=1}^n x_i \text{ or } x_{median} = median(x_1, x_2, \dots, x_n)$$

Then:

$$x'_i = x_{mean} \text{ or } x'_i = x_{median}$$

This replaces each tied value x_i with the computed mean or median, clustering them at a central point.

Parallel Coordinate Plot with Mean or Median Splitting

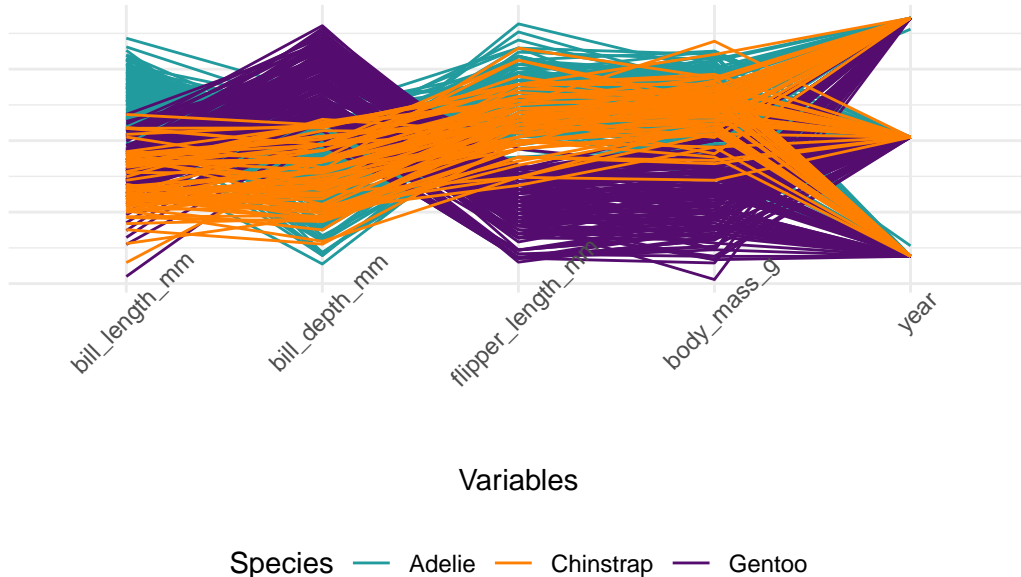


Figure 16: Mean or Median Splitting of Data Points in PCP

Kernel Density Estimation (KDE) with Bandwidth Adjustment

Kernel density estimation (KDE) with bandwidth adjustment is a powerful method for density visualizations. It spreads out overlapping observations by giving each tied value a slightly different weight. This method works especially well in density plots because it evens data distributions without changing the original numbers. Silverman (2018) work in “Density Estimation for Statistics and Data Analysis” shows how KDE with changed bandwidths correctly shows multimodal distributions, reducing errors caused by tied values. KDE works, but you must be careful when choosing the bandwidth because making the wrong choice can lead to misleading visual representations, especially when files are very concentrated.

For KDE with bandwidth adjustment, each data point is used to estimate a probability density function (PDF) by smoothing over a specified bandwidth h , which can be adjusted to account for ties.

The KDE for a set of observations x_1, x_2, \dots, x_n is:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where:

- $\hat{f}(x)$ is the estimated density at x ,
- $K(\cdot)$ is a kernel function
- h is the bandwidth, chosen to spread tied values slightly by selecting a larger h for clusters of tied observations.

The choice of bandwidth h controls the amount of smoothing, which helps visually differentiate tied values.

Parallel Coordinate Plot with KDE with Bandwidth Adjustment

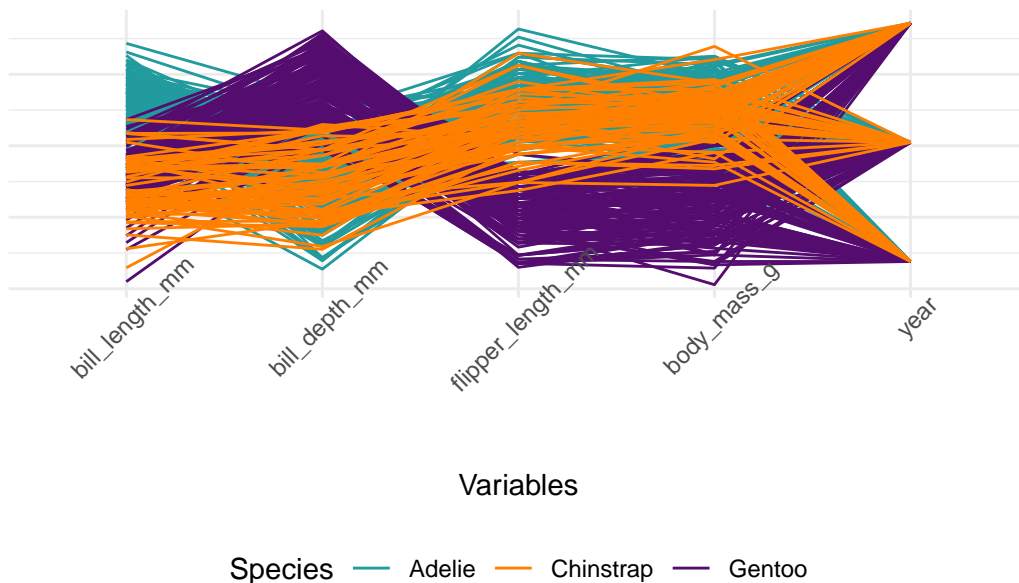


Figure 17: KDE with Bandwidth Adjustment of Data Points in PCP

Advanced visual methods, such as depth cues and density adjustments, are used to deal with ties by making PCPs look like they are in three dimensions.

When combined with line transparency, Depth Cue Parallel Coordinates (DCPC) provide an effective means of visually distinguishing overlapping trends with teemporal binning that employs perception-based coloring and concepts, (Johansson, Ljung, and Cooper 2007). This method was created to deal with the problem of having a lot of temporal data. Their primary focus is on using transparency and brightness changes to layer connected information, which lets users see data as if it were in three-dimensional space. This method keeps the original data and ensures each line can remain identified. Although the technique works, it requires advanced rendering tools that might not be easy to find in all visualization settings.

Each way to show data with ties has unique benefits depending on the type of display and the type of data. Random and predictable jittering, for example, works well in scatterplots because they stop the plot from getting too big and show hidden patterns. KDE and rank-based methods, on the other hand, work well in density plots and ordinal visualizations. Researchers can improve visual clarity by using only some of these methods, ensuring the representation is correct and easy to understand. However, when parallel coordinate plots are used, where multiple directions are plotted simultaneously, the old ways of dealing with ties only sometimes work. In this case, adding the distance between values tied across multiple axes could show patterns and connections that are hard to see because the lines overlap. By creating a new method for parallel coordinate plots that adds space between ties in a planned way, we could see data trends and links across dimensions more clearly while keeping the dataset’s integrity. This new idea would make it easier to understand complex, multidimensional data.

Handling Ties in PCPs

The idea of adding distance to visual ties has changed over time as methods for visualizing data have improved to find a balance between detail and clarity in large datasets. In the early days of visualization, especially in network analysis and multidimensional scaling, closeness and alignment were the main ways to show patterns and ties between things. However, as data grew, visual clutter became a big problem. This was especially true in parallel coordinates plots (PCPs), where lines and intersections that overlapped made it hard to see the underlying data structure. To fix this problem, researchers started looking into ways to give representational parts of these graphs more “visual distance.” They did this using methods like line bundling and changing the spacing between lines to avoid overplotting. “In PCP, participants had to re-order the axes to examine all dimensions, whereas in SPLOM it is straightforward to identify correlation. This result confirms earlier studies showing PCP participants struggle to identify correlation due to the complexity of cross-line patterns,” (Chang, Dwyer, and Marriott 2018a) This shows how separating visual elements can help with cognitive grouping without breaking up the continuity of the data. In PCPs, early studies showed that visual distance could make intersections clearer and improve the overall shape of individual paths. D. Guo et al. (2005) say, “...nested-means approach reduces overlapping problems and ensures the mean value of each variable is always at the midpoint on each axis.” This theory has shaped later developments, showing visual distance as a spatial tool and a key cognitive help in data interpretation needed to get around today’s more complex visualizations.

Axis Reordering

As mentioned in the section on modifications, axis reordering changes the order of dimensions to reduce the number of times lines cross between variables in PCPs, a type of data display. This method makes clusters easier to see by lining up visually linked variables, cutting down on clutter, and making patterns stand out more without overlapping. One benefit is that it works well for datasets where certain factors are strongly related, making clusters and trends easier to read without changing the data. Changing the variable order for clarity rather than data continuity could make it hard to tell if the data is in order. Blumenschein et al. (2020) say that reordering based on correlation minimizes unnecessary line intersections, allowing natural clusters to emerge. This shows how important changing the arrangement can be for seeing clearly. Studies in this area have shown that changing the order of dimensions makes it much easier to find patterns in datasets where dimensions are linked. However, it only works well on datasets with dimensions tied together.

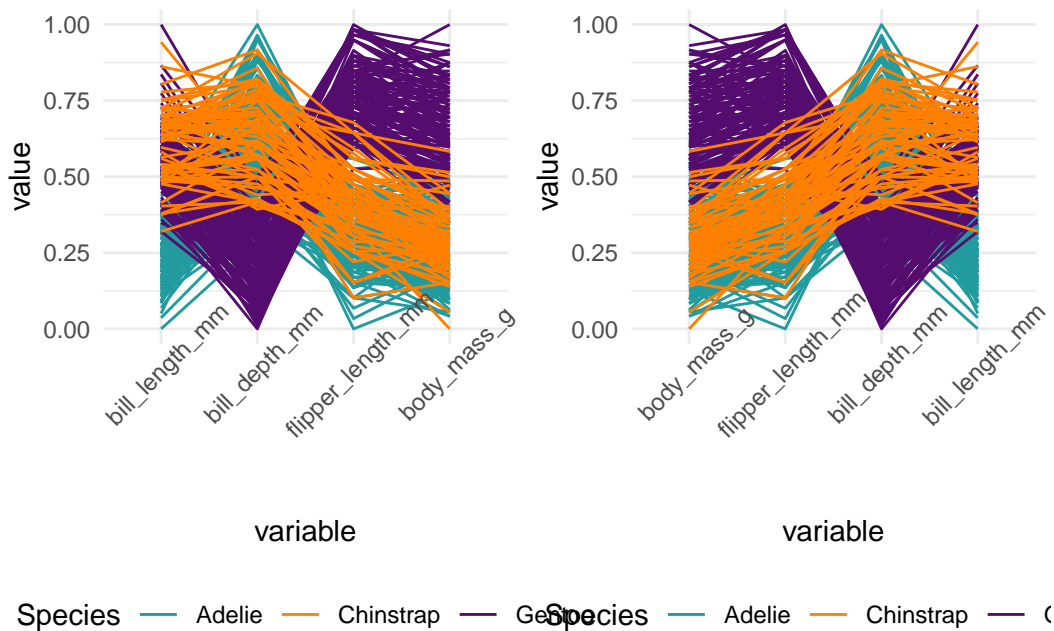


Figure 18: PCP Without Axis Reordering

Line Bundling

By putting similar lines into “bundles,” line bundling cuts down on unnecessary crossings and clutter. This method works well in dense PCPs, where many lines cross and make it hard to see individual patterns. Heinrich et al. say that “bundling lines based on value similarity reduces cognitive strain and enhances visual coherence in complex datasets.” This shows how bundling can make complex data easier to understand without losing larger patterns. Experiments have shown that bundling makes large datasets easier to read, which makes it easier to find general trends even when the data is heavily packed (Julian Heinrich et al. 2011). It makes things easier to understand by collecting similar data lines; this works especially well for numerical ties where values are close but not exactly the same. Bundling, on the other hand, can hide individual lines within a group, making it hard to tell one data point from another.

Quantization and Aggregation

Quantization, also known as Quantized Generalized Parallel Coordinate Plots or QGPCP, groups close-together values into bins. This makes it easier to see patterns in tightly packed data. This method cuts down on the number of lines that cross each other, which makes it great for big datasets. Moustafa states, “By quantizing close data points, we can minimize overplotting and improve discernment of overall trends.” This shows that removing some details can make data easier to read. Quantization can keep important patterns while lowering noise, as tests on real-world data show. However, it may hide small but important differences in the data (Rida EA Moustafa 2009). It makes the image easier to understand by removing small details. This is helpful for high-dimensional datasets with a lot of line overlap. However, detail can be lost during aggregation, making this method less useful for situations where exact data representation is needed.

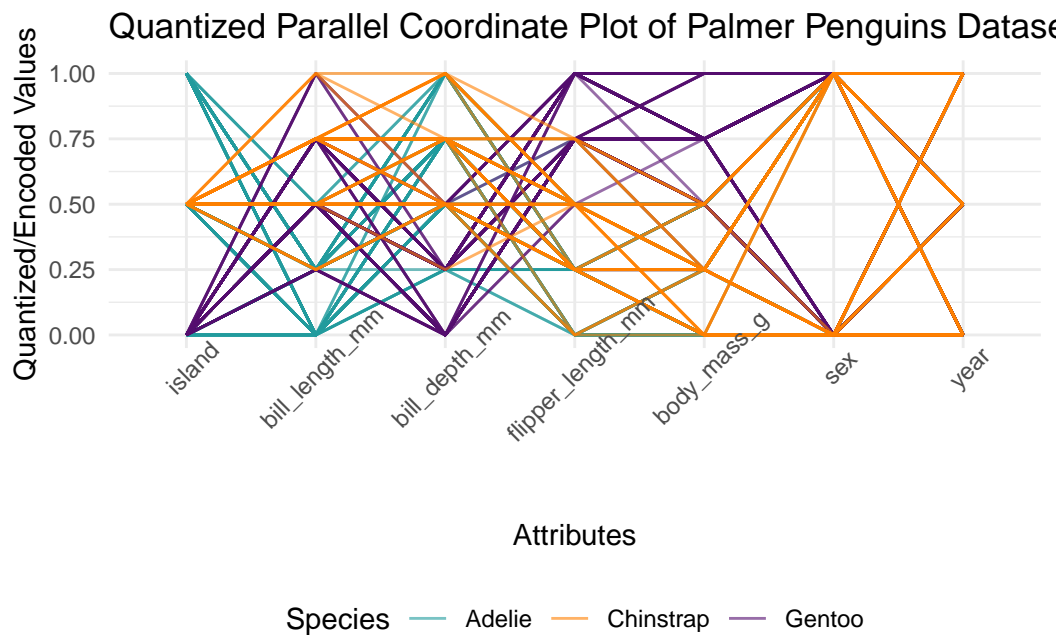


Figure 19: Quantized Generalized Parallel Coordinate Plots

In Figure 19, the overlapping of the Quantized Parallel Coordinate Plot makes it harder to find specific individuals. Quantization makes the visualization even easier by putting numbers into groups. However, this needs to improve some of the accuracy of the original data. The plot can help you see big patterns, like how some species may follow similar paths across many traits, but it doesn’t show exactly how they are related or where they live. Using jittering or interactive plots to cut down on overplotting could make things clearer and make figuring out how factors relate easier.

Color-Coding and Transparency

By changing the color and transparency of lines, you can tell them apart based on density, categories, or values. This makes it easier to tell the difference between PCP lines that meet. Firat et al. say, “Line brushing and transparency adjustments improve the gestalt of the data clusters.” This means that changes in how we see things help us group data cognitively without changing the structure. Firat et al. say that user tests have shown that PCPs with color-coded and clear lines are easier to understand and recognize patterns, especially for differentiating categorical data (Firat, Swallow, and Laramee 2023). This tool can be used to handle both numerical and categorical ties.

The color and transparency options make it easy to visually separate clusters without changing where the lines are placed. However, this method might only work for plots with little information because even changes in color can't fully fix heavy overplotting.

Edge Bundling for Categorical Data

Edge bundling, which comes from network visualization, groups paths for category values in PCPs. This method works best with categorical data because it lets you visually break up ties without changing other factors. Palmas et al. state that “bundling lines with similar orientations prevents overplotting, allowing clearer identification of dominant trends within complex data.” This means bundling can help find trends even in spaces with many dimensions. Bundling made finding categorical clusters in complex PCPs easier and decreased users' work (Palmas et al. 2014). Edge bundling is an excellent way to clear up clutter and make categorical groups easier to see. It lets users see categorical differences without messing up numerical relationships. On the other hand, each bundle may hide small details, making them less useful for datasets that need a lot of precision.

Stacked Histograms and Density Plots

This change adds histograms or density plots on top of PCPs to show how category data is distributed in terms of frequency. This lets users see how distributions change without adding extra line clutter. According to Bok, Kim, and Seo (2020), adding histogram overlays gives categorical context, making variable distributions easier to see without affecting line continuity. This means that frequency-based cues can help you understand categorical distributions without adding more visual clutter. Clear frequency-based differentiation in categorical data without too much line overlap improves the visual understanding of distributions when used with PCPs. Real-world category dataset tests showed that histogram overlays made it easier to differentiate between data sets and see what they meant. Adding histograms makes things look more complicated, which could be too much for some people if used with other PCP changes.

Perceptual Cues (Line Thickness, Texture, etc.)

By changing the lines' width, texture, or other visual properties, you can tell them apart from overlapping lines. This helps people see patterns and tell values apart without separating them physically. According to Chang, Dwyer, and Marriott (2018b), perceptual variations in lines not only differentiate data but also enhance recognition of broader patterns. These cues help with cognitive grouping and pattern differentiation, even in complicated datasets. They work well in areas with a lot of visual information because they let users see trends without changing the structure or order of the data. They are helpful for both numerical and categorical ties. Studies with users showed that changes in how people saw things made finding patterns and trends in large, complex datasets easier. This made it easier for PCPs to look around. In areas with many people, it could be useful because overlapped lines can make it harder to see.

References

- Abdi, Hervé, and Lynne J. Williams. 2010. “Principal Component Analysis.” *WIREs Computational Statistics* 2 (4): 433–59. <https://doi.org/10.1002/wics.101>.
- Ankerst, Mihael, Stefan Berchtold, and Daniel A Keim. 1998. “Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data.” In *Proceedings IEEE Symposium on Information Visualization (Cat. No. 98TB100258)*, 52–60. IEEE.
- Bertini, Enrico, Luigi Dell’Aquila, and Giuseppe Santucci. 2005. “Springview: Cooperation of Radviz and Parallel Coordinates for View Optimization and Clutter Reduction.” In *Coordinated and Multiple Views in Exploratory Visualization (CMV’05)*, 22–29. IEEE.
- Blaas, Jorik, Charl Botha, and Frits Post. 2008. “Extensions of Parallel Coordinates for Interactive Exploration of Large Multi-Timepoint Data Sets.” *IEEE Transactions on Visualization and Computer Graphics* 14 (6): 1436–51.
- Blumenschein, Michael, Xuan Zhang, David Pomerence, Daniel A Keim, and Johannes Fuchs. 2020. “Evaluating Reordering Strategies for Cluster Identification in Parallel Coordinates.” In *Computer Graphics Forum*, 39:537–49. 3. Wiley Online Library.
- Bok, Jinwook, Bohyoung Kim, and Jinwook Seo. 2020. “Augmenting Parallel Coordinates Plots with Color-Coded Stacked Histograms.” *IEEE Transactions on Visualization and Computer Graphics* 28 (7): 2563–76.
- Chambers, John M. 1983. *Graphical Methods for Data Analysis*. Chapman; Hall/CRC.
- Chang, Chunlei, Tim Dwyer, and Kim Marriott. 2018b. “An Evaluation of Perceptually Complementary Views for Multivariate Data.” In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 195–204. IEEE.
- . 2018a. “An Evaluation of Perceptually Complementary Views for Multivariate Data.” In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 201. IEEE.
- Claessen, Jarry HT, and Jarke J Van Wijk. 2011. “Flexible Linked Axes for Multivariate Data Visualization.” *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2310–16.
- Cleveland, William S. 1993. *Visualizing Data*. Hobart press.
- Cleveland, William S, and Robert McGill. 1984. “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.” *Journal of the American Statistical Association* 79 (387): 531–54.
- Conover, William Jay. 1999. *Practical Nonparametric Statistics*. Vol. 350. John Wiley & Sons.
- Dasgupta, Aritra, and Robert Kosara. 2010. “Pargnostics: Screen-Space Metrics for Parallel Coordinates.” *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1017–26.
- David, FN, and JW Tukey. 1977. “Exploratory Data Analysis.” *Biometrics* 33 (4): 768.
- Dennig, Frederik L, Maximilian T Fischer, Michael Blumenschein, Johannes Fuchs, Daniel A Keim, and Evanthia Dimara. 2021. “Parsetgnostics: Quality Metrics for Parallel Sets.” In *Computer Graphics Forum*, 40:375–86. 3. Wiley Online Library.
- Few, Stephen, and Perceptual Edge. 2008. “Solutions to the Problem of over-Plotting in Graphs.” *Visual Business Intelligence Newsletter*.
- Firat, Elif E, Ben Swallow, and Robert S Laramee. 2023. “Pcp-Ed: Parallel Coordinate Plots for Ensemble Data.” *Visual Informatics* 7 (1): 56–65.
- Fua, Ying-Huey, Matthew O Ward, and Elke A Rundensteiner. 1999. *Hierarchical Parallel Coordinates for Exploration of Large Datasets*. IEEE.
- Geng, Huimin, Xutao Deng, and Hesham Ali. 2005. “A New Clustering Algorithm Using Message Passing and Its Applications in Analyzing Microarray Data.” In *Fourth International Conference on Machine Learning and Applications (ICMLA’05)*, 6–pp.

IEEE.

- Guo, Diansheng, Mark Gahegan, Alan M MacEachren, and Biliang Zhou. 2005. "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach." *Cartography and Geographic Information Science* 32 (2): 122.
- Guo, Hanqi, He Xiao, and Xiaoru Yuan. 2012. "Scalable Multivariate Volume Visualization and Analysis Based on Dimension Projection and Parallel Coordinates." *IEEE Transactions on Visualization and Computer Graphics* 18 (9): 1397–1410.
- Healey, Christopher G, and James T Enns. 1999. "Large Datasets at a Glance: Combining Textures and Colors in Scientific Visualization." *IEEE Transactions on Visualization and Computer Graphics* 5 (2): 145–67.
- Heer, Jeffrey, and Michael Bostock. 2010. "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203–12.
- Heinrich, Julian, Yuan Luo, Arthur E Kirkpatrick, Hao Zhang, and Daniel Weiskopf. 2011. "Evaluation of a Bundling Technique for Parallel Coordinates." *arXiv Preprint arXiv:1109.6073*.
- Heinrich, Julian, and Daniel Weiskopf. 2013a. "State of the Art of Parallel Coordinates." *Eurographics (State of the Art Reports)*, 95–116.
- . 2013b. "State of the Art of Parallel Coordinates." In *Eurographics 2013 - State of the Art Reports*, edited by M. Sbert and L. Szirmay-Kalos. The Eurographics Association. <https://doi.org/10.2312/conf/EG2013/stars/095-116>.
- Heinrich, J, and D Weiskopf. 2009. "Continuous Parallel Coordinates." *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1531–38. <https://doi.org/10.1109/TVCG.2009.131>.
- Hofmann, Heike, and Marie Vendettuoli. 2013. "Common Angle Plots as Perception-True Visualizations of Categorical Associations." *IEEE Transactions on Visualization and Computer Graphics* 19 (12): 2297–2305. <https://doi.org/10.1109/TVCG.2013.140>.
- Holten, Danny, and Jarke J Van Wijk. 2009. "Force-Directed Edge Bundling for Graph Visualization." In *Computer Graphics Forum*, 28:983–90. 3. Wiley Online Library.
- Inselberg, Alfred. 1985. "The plane with parallel coordinates." *The Visual Computer* 1 (2): 69–91. <https://doi.org/10.1007/BF01898350>.
- . 1997. "Multidimensional Detective." In *Proceedings of VIZ'97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, 100–107. IEEE.
- . 2009. "Parallel Coordinates: Interactive Visualisation for High Dimensions." *Trends in Interactive Visualization: State-of-the-Art Survey*, 49–78.
- Inselberg, Alfred, and Bernard Dimsdale. 1990. "Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry." In *Proceedings of the First IEEE Conference on Visualization: Visualization90*, 361–78. IEEE.
- Ipkovich, Ádám, Károly Héberger, and János Abonyi. 2021. "Comprehensible Visualization of Multidimensional Data: Sum of Ranking Differences-Based Parallel Coordinates." *Mathematics* 9 (24): 3203.
- Johansson, Jimmy, and Camilla Forsell. 2015. "Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research." *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 579–88.
- Johansson, Jimmy, Patric Ljung, and Matthew Cooper. 2007. "Depth Cues and Density in Temporal Parallel Coordinates." In *EuroVis*, 7:35–42.
- Johansson, Jimmy, Patric Ljung, Mikael Jern, and Matthew Cooper. 2005. "Revealing Structure Within Clustered Parallel Coordinates Displays." In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 125–32. IEEE.
- Kachhway, Inder Singh. 2013. "Enhancement in Visualization of Parallel Coordinates

- Using Curves.”
- Kavvadias, Dimitris J, Grammati E Pantziou, Paul G Spirakis, and Christos D Zaroliagis. 1996. “Hammock-on-Ears Decomposition: A Technique for the Efficient Parallel Solution of Shortest Paths and Other Problems.” *Theoretical Computer Science* 168 (1): 121–54.
- Keim, Daniel A. 2002. “Information Visualization and Visual Data Mining.” *IEEE Transactions on Visualization and Computer Graphics* 8 (1): 1–8.
- LeBlanc, Thomas J, John M Mellor-Crummey, and Robert J Fowler. 1990. “Analyzing Parallel Program Executions Using Multiple Views.” *Journal of Parallel and Distributed Computing* 9 (2): 203–17.
- McDonnell, Kevin T, and Klaus Mueller. 2008. “Illustrative Parallel Coordinates.” In *Computer Graphics Forum*, 27:1031–38. 3. Wiley Online Library.
- Moustafa, Rida E. 2011. “Parallel Coordinate and Parallel Coordinate Density Plots.” *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (2): 134–48.
- Moustafa, Rida EA. 2009. “QGPCP: Quantized Generalized Parallel Coordinate Plots for Large Multivariate Data Visualization.” *Journal of Computational and Graphical Statistics* 18 (1): 32–51.
- Munzner, Tamara. 2014. *Visualization Analysis and Design*. CRC press.
- Novotny, Matej, and Helwig Hauser. 2006. “Outlier-Preserving Focus+ Context Visualization in Parallel Coordinates.” *IEEE Transactions on Visualization and Computer Graphics* 12 (5): 893–900.
- Palmas, Gregorio, Myroslav Bachynskyi, Antti Oulasvirta, Hans Peter Seidel, and Tino Weinkauff. 2014. “An Edge-Bundling Layout for Interactive Parallel Coordinates.” In *2014 IEEE Pacific Visualization Symposium*, 57–64. IEEE.
- “Parallel Coordinates Plot in Ggplot2.” n.d. <https://plotly.com/ggplot2/parallel-coordinates-plot/>.
- Pilhöfer, Alexander, Alexander Gribov, and Antony Unwin. 2012a. “Comparing Clusterings Using Bertin’s Idea.” *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2506–15.
- . 2012b. “Comparing Clusterings Using Bertin’s Idea.” *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2510.
- Poco, Jorge, Ronak Etemadpour, Fernando Vieira Paulovich, TV Long, Paul Rosenthal, Maria Cristina Ferreira de Oliveira, Lars Linsen, and Rosane Minghim. 2011. “A Framework for Exploring Multidimensional Data with 3d Projections.” In *Computer Graphics Forum*, 30:1111–20. 3. Wiley Online Library.
- Qu, Huamin, Wing-Yi Chan, Anbang Xu, Kai-Lun Chung, Kai-Hon Lau, and Ping Guo. 2007. “Visual Analysis of the Air Pollution Problem in Hong Kong.” *IEEE Transactions on Visualization and Computer Graphics* 13 (6): 1408–15.
- Raidou, Renata Georgia, Martin Eisemann, Marcel Breeuwer, Elmar Eisemann, and Anna Vilanova. 2015. “Orientation-Enhanced Parallel Coordinate Plots.” *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 589–98.
- Schonlau, Matthias. 2003. “Visualizing Categorical Data Arising in the Health Sciences Using Hammock Plots.” In *Proceedings of the Joint Statistical Meetings, Section on Statistical Graphics*. American Statistical Association. https://schonlau.net/publication/03jsm_hammockplot.pdf.
- . 2024. “Hammock Plots: Visualizing Categorical and Numerical Variables.” *Journal of Computational and Graphical Statistics*, 1–16.
- Siirtola, Harri, and Kari-Jouko Räihä. 2006. “Interacting with Parallel Coordinates.” *Interacting with Computers* 18 (6): 1278–1309.
- Silverman, Bernard W. 2018. *Density Estimation for Statistics and Data Analysis*. Routledge.
- Swayne, Deborah F, Duncan Temple Lang, Andreas Buja, and Dianne Cook. 2003.

- “GGobi: Evolving from XGobi into an Extensible Framework for Interactive Data Visualization.” *Computational Statistics & Data Analysis* 43 (4): 423–44.
- Symanzik, Jürgen, Michael Friendly, and Ortac Onder. 2018. “The Unsinkable Titanic Data.”
- Theus, Martin, and Simon Urbanek. 2008. *Interactive Graphics for Data Analysis: Principles and Examples*. CRC Press.
- Tufte, Edward R. 2001. *The Visual Display of Quantitative Information (2nd Edition)*. USA: Graphics Press.
- van der Maaten, Laurens, and Geoffrey Hinton. 2008. “Visualizing Data Using t-SNE.” *Journal of Machine Learning Research* 9 (86): 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html>.
- VanderPlas, Susan, Yawei Ge, Antony Unwin, and Heike Hofmann. 2023. “Penguins Go Parallel: A Grammar of Graphics Framework for Generalized Parallel Coordinate Plots.” *Journal of Computational and Graphical Statistics* 32 (4): 1572–87.
- Velagala, Vijay, Weitao Chen, Mark Alber, and Jeremiah J Zartman. 2020. “Multiscale Models Coupling Chemical Signaling and Mechanical Properties for Studying Tissue Growth.” In *Mechanobiology*, 173–95. Elsevier.
- Wang, Shisong, Debajyoti Mondal, Sara Sadri, Chanchal K Roy, James S Famiglietti, and Kevin A Schneider. 2022. “Set-Stat-Map: Extending Parallel Sets for Visualizing Mixed Data.” In *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*, 151–60. IEEE.
- Wegman, Edward J. 1990. “Hyperdimensional data analysis using parallel coordinates.” *Journal of the American Statistical Association* 85: 664–75.
- Wegman, Edward J, and Qiang Luo. 1997b. “High Dimensional Clustering Using Parallel Coordinates and the Grand Tour.” In *Classification and Knowledge Organization: Proceedings of the 20th Annual Conference of the Gesellschaft für Klassifikation eV, University of Freiburg, March 6–8, 1996*, 93–101. Springer.
- . 1997a. “High Dimensional Clustering Using Parallel Coordinates and the Grand Tour.” In *Classification and Knowledge Organization: Proceedings of the 20th Annual Conference of the Gesellschaft für Klassifikation eV, University of Freiburg, March 6–8, 1996*, 94. Springer.
- Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2011. “tourr: An R Package for Exploring Multivariate Data with Projections.” *Journal of Statistical Software, Articles* 40 (2): 1–18. <https://doi.org/10.18637/jss.v040.i02>.
- Wilcox, Rand. 2017. *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction*. Chapman; Hall/CRC.
- Yang, Vincent, Harrison Nguyen, Norman Matloff, and Yingkang Xie. 2017. “Top-Frequency Parallel Coordinates Plots.” *arXiv Preprint arXiv:1709.00665*.
- Yuan, Xiaoru, Peihong Guo, He Xiao, Hong Zhou, and Huamin Qu. 2009. “Scattering Points in Parallel Coordinates.” *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1001–8.
- Zhao, Xin, and Arie Kaufman. 2012. “Structure Revealing Techniques Based on Parallel Coordinates Plot.” *The Visual Computer* 28: 541–51.
- Zhou, Hong, Xiaoru Yuan, Huamin Qu, Weiwei Cui, and Baoquan Chen. 2008. “Visual Clustering in Parallel Coordinates.” In *Computer Graphics Forum*, 27:1047–54. 3. Wiley Online Library.