

Visualizing Ambiguity: A Grammar of Graphics Approach to Resolving Numerical Ties in Parallel Coordinate Plots

Denise Bradford

1. Introduction

This proposal outlines a systematic approach to visually distinguish tied numerical values in multidimensional datasets by employing parallel coordinate plots (PCPs). Parallel coordinates, first popularized by Alfred Inselberg, are a powerful technique for investigating patterns across multiple attributes simultaneously (Inselberg 2009). However, when datasets contain exact numerical ties, the resulting overlapping lines in PCPs can obscure critical distinctions. To address this, we propose a method for introducing controlled spacing to tied values using both traditional jittering techniques and a novel Vogel-inspired distribution algorithm that leverages principles from natural phyllotaxis patterns. This functionality will be integrated into the `ggpcp` package in R, ensuring a streamlined workflow for users seeking enhanced clarity in their parallel coordinate visualizations.

Importantly, our approach complements recent work on generalized parallel coordinate plots (GPCPs) — an extension of PCPs that supports categorical variables (Vander-Plas et al. 2023). The `ggpcp` package for R implements these GPCPs using a grammar of graphics framework, which seamlessly incorporates both continuous and categorical variables in a single parallel coordinate plot. One of the key contributions of that work is a robust tie-breaking mechanism for categorical variables, implemented through the `pcp_arrange()` function with methods including “from-left” and “from-right” hierarchical sorting. This ensures that individual observations can be traced across multiple dimensions, even when categories induce identical or “tied” values. By adding multiple numerical tie-breaking techniques for continuous data—including our innovative Vogel-based approach—we further refine GPCPs’ capacity to handle the visualization of real-world datasets exhibiting many types of ties.

2. Background and Motivation

2.1 Parallel Coordinate Plots

Parallel coordinate plots assign each dimension of an n -dimensional dataset to a vertical axis arranged in parallel (Wegman 1990). Each observation is drawn as a polyline connecting its values on these axes, providing a visual representation that can illuminate underlying data structures.

2.2 Numerical Ties and Visual Overlap

When multiple observations share the same value in a given dimension, their polylines perfectly overlap, creating “visual collisions.” This masks information about distribution, density, or potential outliers. The treatment of ties is an aspect not generally addressed in the original parallel coordinate plots of Inselberg (1985) and Wegman (1990). However, the ggpcp implementation has demonstrated that careful tie-handling is essential for both continuous and categorical variables.

Introducing a small offset (“jitter”) to these tied values can mitigate overlap without distorting the overall relationships in the data (Peng, Ward, and Rundensteiner 2004). In the context of generalized parallel coordinate plots, the ggpcp package separates data management from visual rendering into three distinct components: variable selection and reshaping, scaling of axes, and treatment of ties in categorical axes (VanderPlas et al. 2023).

2.3 Existing Solutions for Categorical Ties in ggpcp

The ggpcp package currently addresses categorical ties through sophisticated tie-breaking algorithms. The package implements hierarchical sorting through the `pcp_arrange(data, method, space)` function, with two primary methods: “from-left” and “from-right”, meaning that tie breaks are determined hierarchically by variables’ values from the specified direction. The parameter `space` specifies the amount of the y-axis to use for spacing between levels of categorical variables, with a default of 5% of the axis used for spacing.

This hierarchical sorting approach serves as “external cognition” — the additional computational processing reduces the cognitive load required to untangle overlapping lines in the parallel coordinate plot. The categorical tie-breaking creates equispaced tie-breaking that reduces line crossings and allows users to follow individual observations from left to right through the plot even for categorical variables.

2.4 Nature-Inspired Solutions: Vogel’s Sunflower Model

Recent advances in visualization theory suggest that natural phenomena can provide optimal solutions to common visualization challenges. Vogel’s model (1979), Vogel (1979), demonstrates how sunflower seeds achieve near-optimal packing through a simple algorithm using the golden angle ($\theta = 2\pi/\varphi^2 \approx 137.508^\circ$) and radial positioning ($r = c\sqrt{n}$). This produces visually harmonious, near-uniform distributions that avoid clustering artifacts common in random displacement methods. We propose adapting this principle to create a theoretically grounded alternative to traditional jittering that maintains aesthetic appeal while preserving data relationships.

The mathematical foundation rests on two principles: (1) Each new element bears the same spatial relationship to existing elements and issues at a fixed angle with respect to the preceding element, and (2) Each new element fits into the largest still-existing gap between older elements. This produces visually harmonious, near-uniform distributions that avoid clustering artifacts common in random displacement methods. The 137.5° angle occurs also in the arrangement of leaves or side branches around the main stem of many plants, especially compositae, demonstrating the biological optimality of this approach.

We propose adapting this principle to create a theoretically grounded alternative to traditional jittering that maintains aesthetic appeal while preserving data relationships,

extending the grammar of graphics framework to include nature-inspired transformations.

3. Proposed Method for Tie Spacing in ggpcp

3.1 Overview

Building on the existing ggpcp architecture that separates data wrangling from visual rendering, we propose integrating three R functions into the ggpcp package to handle overlapping polylines caused by numerical ties. These functions introduce minimal, controlled spacing before the parallel coordinates are drawn, following the established pattern where `pcp_select`, `pcp_arrange`, and `pcp_scale` handle variable selection, tie arrangement, and scaling respectively.

Our approach mirrors how ggpcp already handles ties in categorical data through the `pcp_arrange()` function, thereby providing a unified tie-breaking solution for both numeric and factor variables. Users can select from: a straightforward fixed tie-band approach (Solution #1), an adaptive method based on data properties (Solution #2), or our novel Vogel-inspired distribution method (Solution #3).

Below are the core functions that implement these approaches:

Key Features of Each Approach

1. Solution #1 – Static Tie Band

- Relies on a user-defined `tie_band` to offset tied values.
- Offers simplicity and predictable increments in spacing.
- Parallels the existing ggpcp categorical approach with fixed spacing parameters.

2. Solution #2 – Dynamic Tie Band

- Adjusts tie spacing based on data properties (e.g., normality checks, IQR, and outlier detection).
- More adaptive, especially useful for heterogeneous or continuous data with varying scales.
- Extends the ggpcp scaling philosophy to tie-breaking through data-driven parameter selection.

3. Solution #3 – Vogel-Inspired Distribution

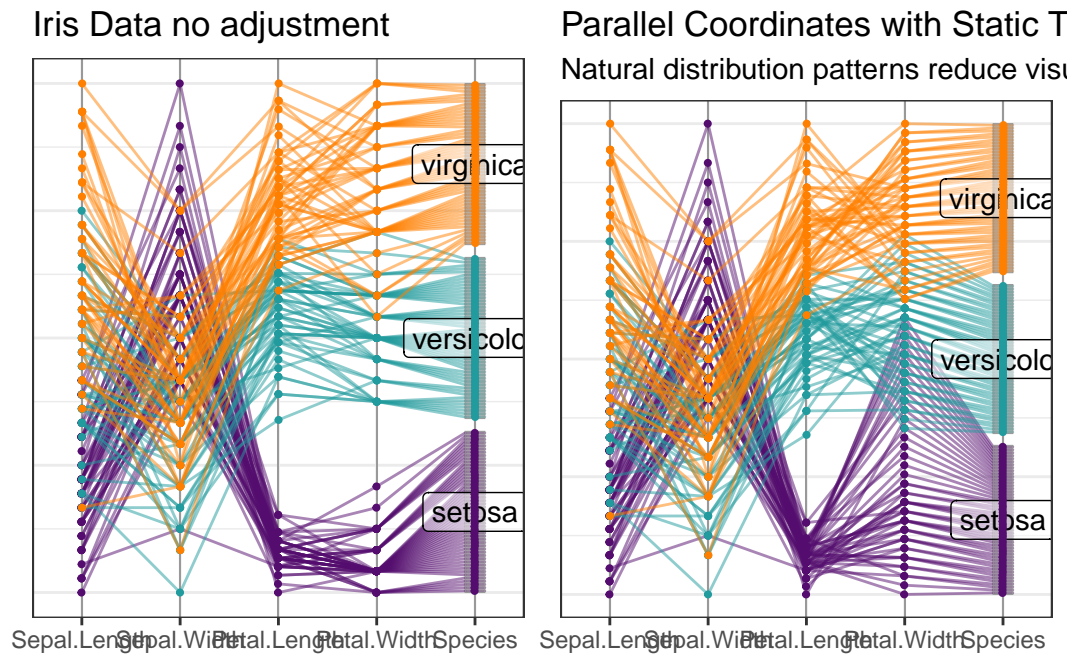
- Applies the golden angle principle ($\theta = 137.508^\circ$) from sunflower seed arrangements to achieve optimal point distribution.
- Implements the cyclotron spiral growth program ($r \approx \sqrt{k}$) that maintains uniform area distribution while following angular uniformity principles.
- Produces aesthetically pleasing, near-uniform spacing that scales naturally with the number of tied observations.
- Deterministic and reproducible, unlike random jittering methods.

- Maintains visual harmony while preserving data relationships through mathematically principled displacement based on natural optimization principles.

3.2 Integration into ggpcp

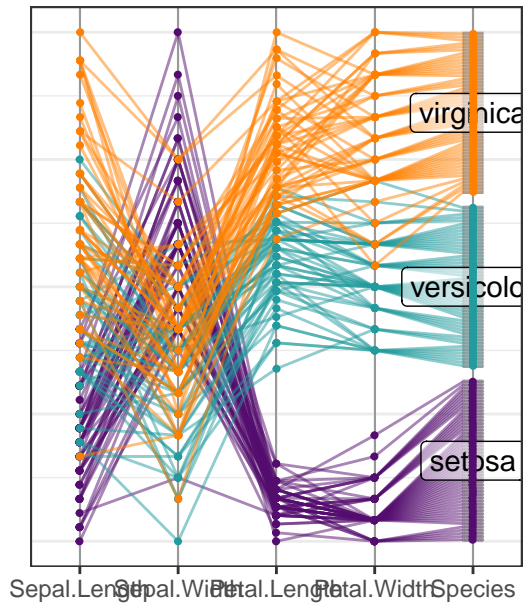
Following the established ggpcp workflow where data management is separated from visual rendering, when integrated into **ggpcp**, users can preprocess their numeric columns with any of the three functions prior to plotting. The parallel coordinate plot (via `geom_pcp()` or similar functionality in **ggpcp**) then operates on the adjusted (tie-spaced) data, preventing line overlap.

This integration follows the existing ggpcp philosophy demonstrated in their standard workflow:

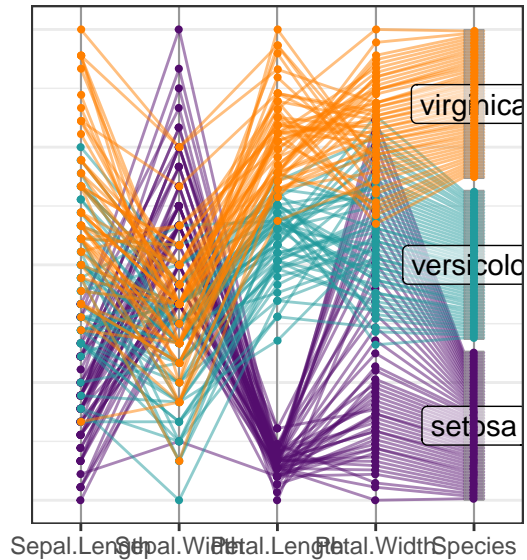


The approach ensures that all forms of ties—categorical or continuous are resolved before rendering the PCP, maintaining the established separation between data preparation and visualization that characterizes the **ggpcp** design philosophy.

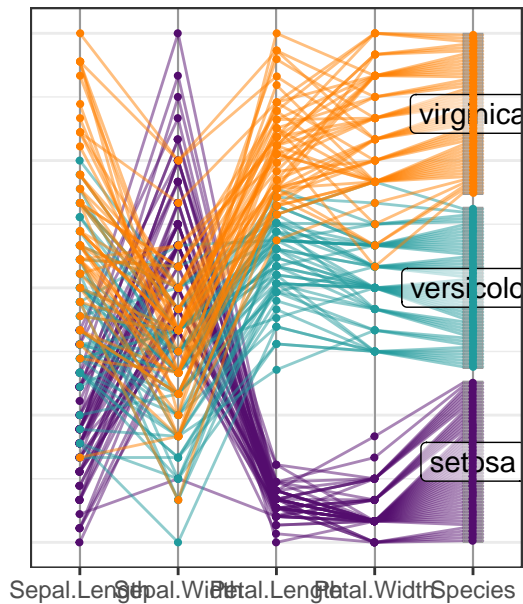
Iris Data no adjustment



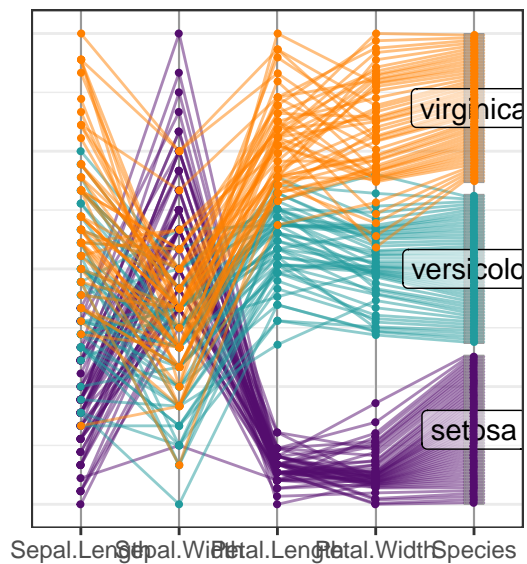
Parallel Coordinates with Dynamic Natural distribution patterns reduce visual clutter



Iris Data no adjustment



Parallel Coordinates with Vogel-I Natural distribution patterns reduce visual clutter



3.3 Theoretical Foundation and Grammar of Graphics Integration

The Vogel approach represents a significant advancement in the theoretical foundation of position adjustments within the Grammar of Graphics framework. Unlike ad hoc random jittering, the Vogel transformation provides a mathematically principled solution that follows natural optimization principles derived from phyllotaxis patterns.

Building on Vogel's original biological observations, the method:

- Preserves topological relationships between distinct data values through consistent angular relationships.
- Reflects data density naturally through spiral tightness, following the cyclotron growth program.

- Maintains aesthetic consistency across all scales through the golden angle property (137.508°).
- Ensures reproducibility through deterministic calculations based on the Fibonacci sequence relationships inherent in the golden angle.

Within Wilkinson’s Grammar of Graphics, this becomes a new category of biomimetic transformations that leverage evolutionary optimization principles to solve visualization challenges. This theoretical contribution extends beyond the immediate application to parallel coordinates, establishing a framework for incorporating natural pattern principles into data visualization grammar.

4. Proposed Research

4.1 Primary Research Question

How can the formal structure of the Grammar of Graphics be extended to systematically incorporate and evaluate methods for resolving numerical ties in parallel coordinate plots, and what is the quantifiable impact of these methods on the accuracy and efficiency of visual data analysis?

4.2 Secondary Question 1: Theory

How can the management of numerical ties be most effectively and coherently formalized within the layered grammar of graphics, building on the established ggpcp framework for categorical tie management?

Methodology: Theoretical analysis, literature synthesis, and specification of a new or extended grammatical element that harmonizes with existing ggpcp architecture.

4.3 Secondary Question 2: Methodology

What are the optimal algorithmic criteria for ordering and spacing tied data points to maximize visual clarity while preserving underlying data properties, particularly in comparison to the hierarchical sorting methods already established in ggpcp for categorical variables?

Methodology: Algorithm design and implementation in R. Develop a novel method that balances separation and density preservation while maintaining compatibility with existing ggpcp workflows.

4.4 Secondary Question 3: Perception

How do different visualization strategies for numerical ties affect an analyst’s ability to perform key visual tasks (cluster identification, outlier detection) compared to the categorical tie-breaking approaches already validated in ggpcp?

Methodology: Controlled user studies with human subjects. Measure task accuracy and completion time across different visualization conditions, including comparisons with existing ggpcp categorical tie-breaking performance.

4.5 Secondary Question 4: Practice

Can a set of evidence-based heuristics be developed to guide practitioners in selecting the most appropriate numerical tie-breaking method for their specific data context, integrating with existing `ggpcp` parameter selection guidelines?

Methodology: Synthesize findings from RQ1-3 into a practical framework. Validate with case studies on real-world datasets from different domains, ensuring compatibility with existing `ggpcp` use cases.

5. Implementation Roadmap

1. Data Preparation

- Ensure compatibility with existing `ggpcp` tidy format and `pcp_select` output structure.
- Normalize or standardize dimensions if necessary, following existing `pcp_scale` methodology, so the offset remains meaningful and does not distort the scale.

2. Integration into `ggpcp` Architecture

- Incorporate all three tie-breaker functions as user-facing options within the established `ggpcp` workflow.
- Provide grammar-based syntax following `ggpcp` conventions: `pcp_numerical_ties(method = "vogel", displacement_factor = 0.02)`
- Expose parameters (e.g., `tie_band`, `base_fraction`, `displacement_factor`) through well-documented function arguments and vignettes, following existing `ggpcp` documentation patterns.
- Ensure compatibility with existing `aes_pcp()` and `geom_pcp()` infrastructure.

3. Rendering Parallel Coordinates

- Use existing `ggpcp` parallel coordinate functions (`geom_pcp`, `geom_pcp_axes`, `geom_pcp_labels`) on the tie-spaced data.
- Provide relevant axis annotations, legends, and tooltips consistent with existing `ggpcp` interactive capabilities.
- Add visual indicators to show which axes have been numerically adjusted, following the pattern of existing `ggpcp` categorical indicators.

4. Testing and Validation

- Assess the impact of numerical tie spacing by comparing plots with and without offsets, using existing `ggpcp` categorical tie-breaking as a performance baseline.
- Conduct user studies comparing visual clarity and pattern recognition between random jittering, Vogel distribution, and existing `ggpcp` categorical approaches.
- Gather feedback from existing `ggpcp` user community to fine-tune default parameters and ensure consistency with established user experience patterns.

6. Historical Context of Multidimensional Data Visualization

6.1 Early Developments

The concept of parallel coordinates dates back to Maurice d’Ocagne’s work on “coordinate parallelism” (d’Ocagne 1885). While his contributions did not see immediate mainstream adoption, they formed the conceptual foundation for later multidimensional plotting.

6.2 Alfred Inselberg’s Contributions

Alfred Inselberg developed the mathematical underpinnings of parallel coordinates for analyzing high-dimensional data, demonstrating the approach’s strength in various domains Inselberg (2009).

6.3 Wegman’s Statistical Insights

Wegman (1990) further popularized parallel coordinates by integrating statistical perspectives, showing how PCPs can be leveraged in exploratory data analysis. Subsequent advances in interactive visualization have enabled real-time manipulation of axes, brushing, and filtering, allowing users to distill meaningful insights from large datasets.

6.4 Generalized Parallel Coordinates

More recently, VanderPlas et al. (2023) extended PCPs to handle both categorical and continuous variables within a single framework, implemented via the R package `ggpcp`. Their work, *Penguins Go Parallel*, details a grammar-of-graphics approach that accommodates a mix of variable types by applying sophisticated tie-breaking and ordering strategies on categorical data to preserve the continuity of observation trajectories.

The `ggpcp` implementation represents a significant advance in parallel coordinate visualization by separating data management from visual rendering. The package’s modular approach allows users to customize variable selection (`pcp_select`), scaling (`pcp_scale`), and categorical tie arrangement (`pcp_arrange`) independently, providing unprecedented flexibility in parallel coordinate plot construction.

A key innovation of `ggpcp` is its ability to follow a single observation through multiple categorical and continuous axes, providing a perceptual advantage over other alternatives. This continuity is achieved through hierarchical sorting algorithms that minimize line crossings while maintaining the ability to trace individual data points across dimensions.

6.5 Nature-Inspired Visualization Methods

Our work builds upon emerging research in biomimetic visualization techniques that apply natural optimization principles to data presentation challenges. Vogel’s sunflower model represents one of the earliest mathematical formalizations of natural packing efficiency, demonstrating how the golden angle (137.508°) produces optimal point distributions through simple growth rules (Vogel 1979).

Vogel’s mathematical formulation provides both angular uniformity (through the golden angle) and area uniformity (through the cyclotron spiral growth program $r \approx \sqrt{k}$), creating patterns that are both mathematically optimal and aesthetically pleasing. His work demonstrates that two simple principles—most uniform angular distribution and most uniform area distribution—explain natural patterns down to the most minute irregularities.

By adapting this principle to parallel coordinate visualization, we contribute to a growing body of work that recognizes nature as a source of elegant solutions to complex visualization problems, while building on the solid foundation established by the **ggpcp** framework.

7. Expected Outcomes

1. Improved Visual Clarity

- The introduction of numerical tie spacing in **ggpcp** will ensure that overlaid lines become distinguishable, unveiling patterns and anomalies otherwise hidden by overlap, complementing the existing categorical tie-breaking capabilities.
- The Vogel method will provide superior aesthetic appeal and consistent spacing compared to random jittering, following natural optimization principles proven effective in biological systems.

2. Enhanced Data Interpretation

- Researchers and data analysts will be able to more accurately interpret subtle differences in high-dimensional datasets, especially in domains like bioinformatics, finance, and engineering, building on the proven success of **ggpcp** in handling mixed variable types.
- The natural distribution patterns will improve pattern recognition and reduce cognitive load, extending the “external cognition” benefits already demonstrated in **ggpcp**’s categorical tie-breaking.

3. Comprehensive Tie-Handling

- By integrating numerical tie-breaking (including the novel Vogel approach) with existing categorical tie-breaking in **ggpcp**, users will enjoy a complete solution for parallel coordinate visualization that handles all forms of data ties within a unified grammar of graphics framework.

4. Easy Integration with R Ecosystem

- Users of **ggpcp** will have ready-to-use, integrated solutions for all tie handling without resorting to external data preprocessing or manual jittering methods, maintaining the package’s established philosophy of seamless workflow integration.

5. Theoretical Contribution

- The Vogel approach establishes a new paradigm for biomimetic transformations within the Grammar of Graphics, opening avenues for other nature-inspired visualization techniques while building on the theoretical foundations already established by the **ggpcp** framework.

8. Conclusion

Parallel coordinates have evolved from early geometric solutions (d’Ocagne?) to the sophisticated, grammar-of-graphics-based implementations exemplified by `ggpcp` (VanderPlas et al. 2023). While the `ggpcp` package successfully addresses categorical ties through hierarchical sorting and equispaced arrangements, numerical ties remain a persistent challenge, causing overlapping lines that diminish interpretability.

By integrating minimal offset mechanisms—including both traditional methods and our novel Vogel-inspired distribution algorithm—directly into the existing `ggpcp` architecture, this proposal ensures a streamlined approach that balances the preservation of data fidelity with the need for visual clarity. The approach maintains compatibility with `ggpcp`’s established separation of data management from visual rendering while extending the package’s comprehensive tie-handling capabilities.

The incorporation of Vogel’s sunflower model represents a significant theoretical advancement, introducing nature-inspired optimization principles to the Grammar of Graphics framework already established in `ggpcp`. This biomimetic approach not only solves the practical problem of numerical visual overlap but also establishes a foundation for future research into natural pattern applications in data visualization, building on the solid theoretical and practical foundations of the `ggpcp` ecosystem.

In tandem with the existing developments in generalized parallel coordinate plots, the resulting enhancement will aid analysts in effectively discerning underlying structures and relationships across multiple dimensions, while providing theoretically grounded, aesthetically pleasing, and reproducible visualizations within the familiar `ggpcp` workflow that users have already adopted.

References

- d’Ocagne, Maurice. 1885. “Coordonnées parallèles et axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles.” Gauthier-Villars, 112. <https://archive.org/details/coordonnesparal00ocaggoog/page/n10>.
- Inselberg, Alfred. 1985. “The plane with parallel coordinates.” *The Visual Computer* 1 (2): 69–91. <https://doi.org/10.1007/BF01898350>.
- Inselberg, Alfred. 2009. “Parallel Coordinates: Interactive Visualisation for High Dimensions.” *Trends in Interactive Visualization: State-of-the-Art Survey*, 49–78.
- Peng, Wei, Matthew O Ward, and Elke A Rundensteiner. 2004. “Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering.” In *IEEE Symposium on Information Visualization*, 89–96. IEEE.
- Vogel, H. 1979. “A better way to construct the sunflower head.” *Mathematical Biosciences* 44(1): 179–189.
- 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.
- Wegman, Edward J. 1990. “Hyperdimensional data analysis using parallel coordinates.” *Journal of the American Statistical Association* 85: 664–75.
- Wilkinson, Leland. 2005. *The Grammar of Graphics*. 2nd ed. Springer-Verlag.

- d'Ocagne, Maurice. 1885. "Coordonnées parallèles et axiales : Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles." *Gauthier-Villars*, 112. <https://archive.org/details/coordonnesparal00ocaggoog/page/n10>.
- Inselberg, Alfred. 1985. "The plane with parallel coordinates." *The Visual Computer* 1 (2): 69–91. <https://doi.org/10.1007/BF01898350>.
- . 2009. "Parallel Coordinates: Interactive Visualisation for High Dimensions." *Trends in Interactive Visualization: State-of-the-Art Survey*, 49–78.
- Peng, Wei, Matthew O Ward, and Elke A Rundensteiner. 2004. "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering." In *IEEE Symposium on Information Visualization*, 89–96. IEEE.
- 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.
- Vogel, Helmut. 1979. "A Better Way to Construct the Sunflower Head." *Mathematical Biosciences* 44 (3-4): 179–89.
- Wegman, Edward J. 1990. "Hyperdimensional data analysis using parallel coordinates." *Journal of the American Statistical Association* 85: 664–75.