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

Comprehensive Exam Presentation

Denise Bradford

University of Nebraska–Lincoln

October 2025

Introduction

The Fundamental Challenge

Parallel Coordinate Plots (PCPs): Map *n*-dimensional data onto two-dimensional displays using parallel vertical axes (Inselberg 1985; Wegman 1990)

The Core Problem:

- Multiple observations with identical numerical values create perfectly overlapping polylines
- "Visual collision" - a single visible line may represent 1 or 1,000 observations
- No mechanism to distinguish density

Critical Impact:

- Density information completely lost
- Substructure within tied groups invisible
- Exploratory analysis fundamentally compromised

Historical Context & Prior Work

Origins of Parallel Coordinates:

- Inselberg (1985): Geometric duality properties, pattern recognition
- Wegman (1990): Statistical data analysis, high-dimensional cluster detection

Evolution:

- 1990s-2000s: Recognition of "overplotting" problem (Wegman and Luo 1991; Johansson and Forsell 2016)
- 2006: Parallel Sets for categorical data (Kosara, Bendix, and Hauser 2006)
- 2010s: Quality metrics development (Dennig et al. 2021)
- 2023: ggpcp package - Grammar of Graphics framework (VanderPlas et al. 2023)

The Persistent Gap:

Categorical tie-breaking elegantly solved, numerical ties without systematic treatment

The ggpcp Package Foundation

Key Innovation: Separation of Concerns (VanderPlas et al. 2023)

ggpcp divides PCP creation into distinct modules:

1. **Variable Selection** (`pcp_select`): Choose and order dimensions
2. **Axis Scaling** (`pcp_scale`): Normalize or transform scales
3. **Tie Resolution** (`pcp_arrange`): Handle overlapping values

Categorical Tie-Breaking Success:

- Hierarchical sorting minimizes line crossings
- Equispaced distribution ensures visibility
- Results visually similar to parallel sets when dense (VanderPlas et al. 2023)

The Missing Piece:

- Categorical ties: ✓ Solved
- Numerical ties: X Unsolved (until now)

The Problem

The Problem

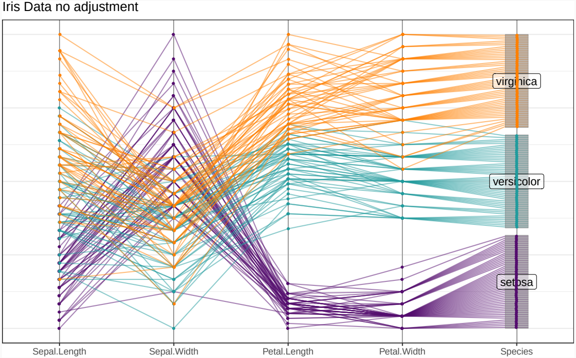

Figure 1: ggpcp data wrangling workflow

Visual Occlusion - Three Problems

1. Visual Collision

- Multiple polylines stack perfectly on identical paths
- Single visible line = unknown number of observations
- Violates principle: visual magnitude $\leftrightarrow$ data magnitude (Cleveland and McGill 1984)

2. Density Information Loss

- Cannot distinguish 1 observation from 100
- Frequency distributions invisible
- Statistical inference impossible

3. Structural Occlusion

- Sub-clusters within tied groups hidden
- Outliers among tied values disappear
- Multivariate patterns cannot be traced (Johansson et al. 2005)

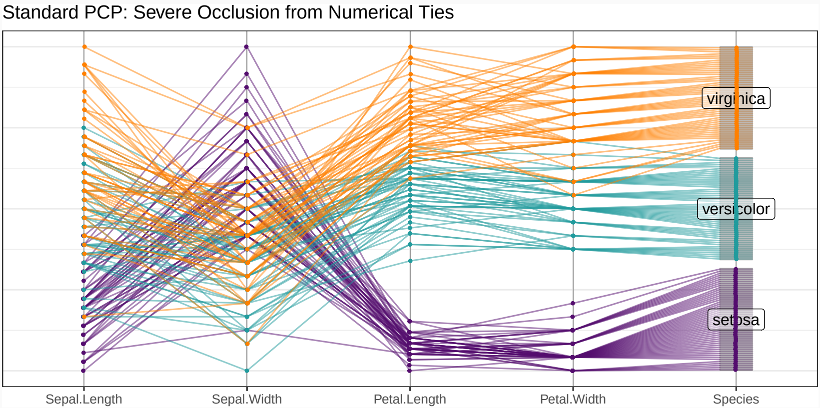

Figure 2: Iris dataset showing numerical tie problem

Dataset Characteristics:

- 150 observations, 4 continuous dimensions
- Integer and half-integer measurements create natural ties
- Species-specific patterns masked by overlapping lines

Design Requirements for Solutions

Four Core Principles:

1. Determinism

- Identical input $\rightarrow$ identical output
- Essential for scientific reproducibility (Peng 2011)

2. Uniformity

- Even distribution within displacement interval
- Minimize artificial clustering
- Represent density faithfully

3. Perceptual Validity

- Displacement small enough to maintain value integrity
- Large enough for visual separation
- No false patterns introduced (Ware 2012)

4. Scalability

- Handle 2 to 10,000+ observations in tie groups
- Computational efficiency for real-time interaction

Methodology

Mathematical Framework

Displacement Constraint:

For each tied value v, distribute n_{ties} observations within:

$$\left[v - \frac{\epsilon}{2}, v + \frac{\epsilon}{2}\right]$$

Key Parameter:

- ϵ: maximum displacement magnitude
- Typically 0.05–0.10 of axis range
- User-adjustable based on data characteristics

Optimization Goals:

- Maximize minimum inter-point distance
- Minimize visual artifacts (clustering, patterns)
- Maintain deterministic reproducibility

Three Deterministic Jittering Algorithms

Determinism ensures full reproducibility—eliminating stochasticity and visual artifacts of standard random jitter

1. Sunflower Jitter: Biomimetic approach adapted from Vogel's model of phyllotaxis (Vogel 1979)

2. Halton Jitter: Quasi-random method using low-discrepancy sequences (Halton 1960)

3. Intelligent Jitter: Novel 1D method utilizing golden ratio with different displacement scaling

All three methods provide:

- Deterministic, reproducible output
- Theoretically-grounded distributions
- Computational efficiency

Sunflower Jitter - Biological Inspiration

Phyllotaxis: Nature's Packing Solution (Vogel 1979)
Sunflower seed arrangements follow a simple algorithm achieving near-optimal packing:

The Golden Angle: $137.508° = 360° \times (2 - \phi)$

- Most "irrational" number in continued fraction sense
- Ensures no radial alignment even after many iterations
- Derived from golden ratio: $\phi = \frac{1+\sqrt{5}}{2}$

Why This Angle?
Through continued fraction theory, 137.508° has the slowest-converging expansion, making it maximally incommensurable with full rotations (Vogel 1979; Ridley 1982)

After hundreds of seeds, none align along the same radius—optimal space-filling property

Sunflower Jitter - Algorithm Specification

Mathematical Formulation:
For observation j in tie group of size n_{ties}:

$$\text{angle}_j = (j - 1) \times 137.508°$$

$$\text{radius}_j = \epsilon \times \sqrt{\frac{j - 1}{n_{\text{ties}}}}$$

$$\text{displacement}_j = \text{radius}_j \times \cos(\text{angle}_j)$$

Key Features:

- **Square root scaling**: Maintains constant density as radius increases
- **Cosine projection**: Maps 2D polar $\rightarrow$ 1D linear displacement
- **Spiral structure**: Preserved even in 1D projection

Computational Complexity: $O(n)$ per tie group

Sunflower Jitter - Theoretical Properties

Why Square Root Scaling?

In a 2D disk:

- Circumference at radius r: $2\pi r$
- Number of points at radius r_j: proportional to j
- For constant density: $r \propto \sqrt{j}$ balances linear growth in points with radial expansion

Distribution Quality:

- Near-optimal minimum separation
- Low-discrepancy properties (Vogel 1979)
- Aesthetically consistent patterns
- Evolutionarily optimized over millions of years

Applications: Originally developed for botanical modeling, now applied to computational geometry and visualization

Sunflower Jitter - Visual Results

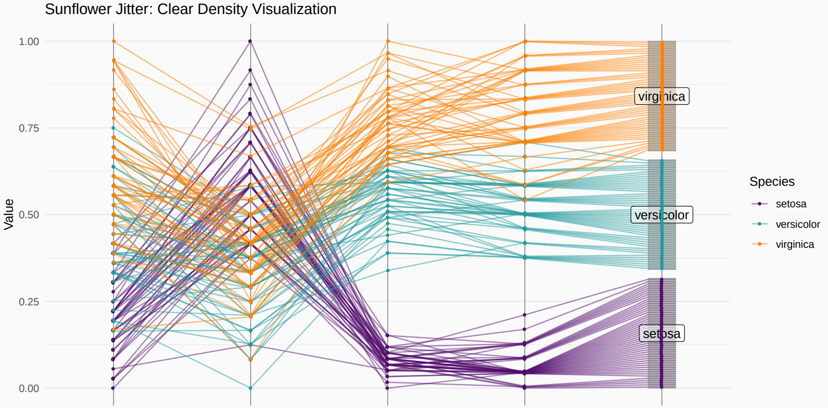

Figure 3: Sunflower jittering resolves ties with biomimetic distribution

Halton Jitter - Quasi-Random Sequences

Beyond Pseudo-Randomness (Halton 1960)

Halton sequences are:

- **Deterministic** (not random)
- **Low-discrepancy** (fill space uniformly)
- **Number-theoretically constructed** using prime bases

The Random Number Problem:

Pseudo-random numbers cluster due to chance (birthday paradox). In 100 random points, gaps and clusters are inevitable.

Halton's Solution:

Place each new point as far as possible from all previous points through systematic construction using van der Corput sequence (Halton 1960; Niederreiter 1992)

Halton Jitter - Construction Method

Van der Corput Sequence in Base 2:

i	Binary	Reversed	Decimal h_i
0	0	0	0.0
1	1	1	0.5
2	10	01	0.25
3	11	11	0.75
4	100	001	0.125
5	101	101	0.625

Pattern: Each point bisects the largest remaining gap
For PCPs:

$$\text{displacement}_i = \epsilon \times (h_i - 0.5)$$

Centering around 0.5 creates symmetric bidirectional displacement

Halton Jitter - Theoretical Guarantees

Discrepancy Theory (Niederreiter 1992)
Measure of how uniformly points fill an interval:

$$D_n = \sup_{I \subseteq [0,1]} \left| \frac{\#\{x_i \in I\}}{n} - |I| \right|$$

Theoretical Bounds:

- **Random sequences**: $D_n = O(n^{-1/2})$
- **Halton sequences**: $D_n = O(n^{-1} \log n)$
- **Optimal (lower bound)**: $D_n = \Omega(n^{-1} \log n)$

Halton is near-optimal in the information-theoretic sense

Applications: Monte Carlo integration, computer graphics sampling, numerical analysis (Niederreiter 1992)

Halton Jitter - Visual Results

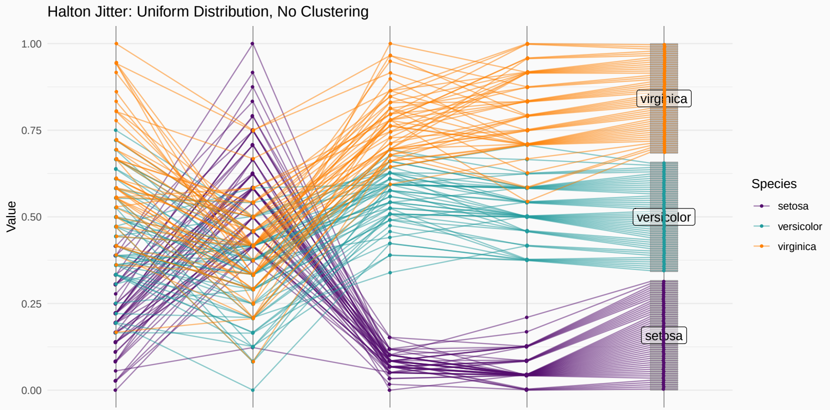

Figure 4: Halton jittering provides uniform low-discrepancy distribution

Intelligent Jitter - Novel Approach

Research Question: Can we apply the golden ratio directly in 1D rather than through angular spacing?

Algorithm Design:

$$\text{angle}_i = (i-1) \times 2\pi \times 0.618$$

$$\text{displacement}_i = \epsilon \times \cos(\text{angle}_i) \times \frac{i-1}{n_{\text{ties}}}$$

Key Distinctions:

- Uses golden ratio value (0.618) not golden angle (137.5°)
- **Linear scaling**: $\frac{i-1}{n_{\text{ties}}}$ instead of $\sqrt{i/n}$
- 1D modulation rather than 2D-to-1D projection

Intelligent Jitter - Hypothesis vs Reality

Hypothesis:

Linear scaling would create progressive reveal:

- Early observations: small displacement (stay near true value)
- Later observations: larger displacement (fill the space)
- Golden ratio modulation provides optimal distribution

Reality (Empirical Results):

- Poor visual quality
- Excessive displacement for later observations
- Artificial separation misrepresents data structure
- Linear scaling inappropriate for this application

For $n = 100$ tie group: Observation #100 displaced by 99% of ϵ, creating false impression of substructure

Dimensional Analysis Comparison

Three Distinct Approaches:

Halton (Pure 1D)

- Operates entirely on a line
- Direct generation of 1D points
- No projection or transformation

Sunflower (2D $\rightarrow$ 1D)

- Constructs 2D polar coordinates (r, θ)
- Projects via $x = r\cos(\theta)$
- Preserves some 2D distribution properties

Intelligent (Hybrid 1D)

- Uses 1D displacement with 2D-inspired modulation
- Discards sine component, keeps cosine
- May combine disadvantages rather than advantages

Scaling Behavior Analysis

Displacement Magnitude Growth:

Method	Scaling Function	Growth Rate
Halton	Uniform $[-\epsilon/2, \epsilon/2]$	Constant
Sunflower	$\propto \sqrt{i/n}$	Sublinear
Intelligent	$\propto i/n$	Linear

For $n = 50$ observations:

- **Halton**: All points uniformly distributed immediately
- **Sunflower**: Typical displacement grows as $\sqrt{1}, \sqrt{2}, \sqrt{3}...$
- **Intelligent**: Displacement grows as $1, 2, 3, 4, 5...$

Consequence: Intelligent creates artificial stratification

Comparative Visual Results

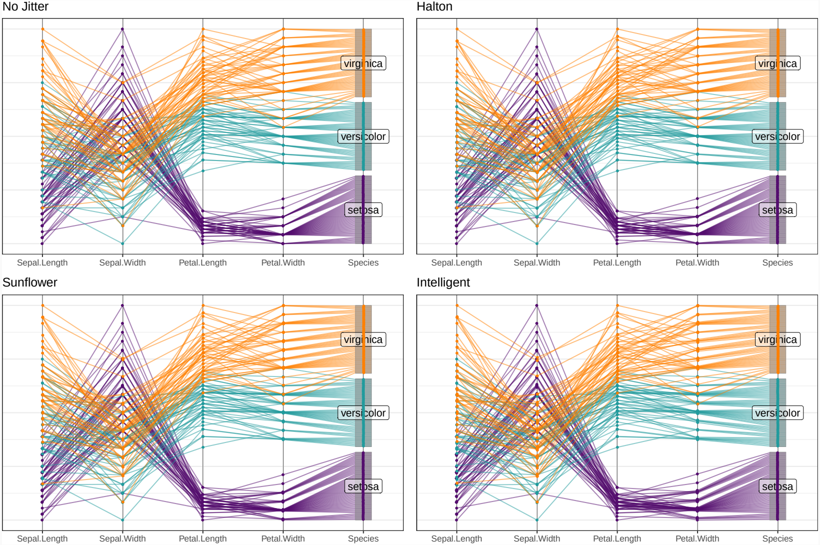

Figure 5: Side-by-side comparison of all jittering methods

Evaluation

User Study Design - Overview

Study Type: Within-subjects repeated measures design with counterbalancing

Independent Variable: Jittering method (5 levels)

- No Jitter (baseline control)
- Random Jitter (current practice)
- Halton Jitter (low-discrepancy)
- Sunflower Jitter (biomimetic)
- Intelligent Jitter (novel method)

Target Sample: 30-50 participants

- Basic visualization literacy required
- No PCP expertise required
- Normal or corrected-to-normal vision
- No color blindness

User Study Design - Dependent Variables

Three-Dimensional Assessment:

1. Task Accuracy (Primary)

- Density estimation: “How many observations follow this path?”
- Cluster identification: “Which species is most common?”
- Pattern detection: “Are there outliers in this group?”

2. Task Completion Time (Secondary)

- Time to complete density estimation tasks
- Time to identify specific observations
- Time to detect patterns

3. Subjective Preference (Tertiary)

- 10-point Likert scale ratings
- Open-ended feedback on clarity
- Comparative ranking of methods

Expected Study Outcomes

Hypotheses:

H1 (Accuracy): Halton and Sunflower $>$ Random $>$ No Jitter

- Rationale: Uniform distribution $\rightarrow$ accurate density perception

H2 (Time): Halton   Sunflower $<$ Random $<$ No Jitter

- Rationale: Clear separation $\rightarrow$ faster analysis

H3 (Preference): Sunflower   Halton $>$ Random $>$ No Jitter

- Rationale: Aesthetic appeal + functionality

H4 (Intelligent): Performance below expectations

- Rationale: Artificial patterns interfere with perception

Statistical Analysis: Repeated-measures ANOVA with Bonferroni post-hoc tests

Discussion

Analysis - Random Jitter Problems

The Birthday Paradox Effect

In a space with m positions, after n random placements:

$$P(\text{collision}) = 1 - \frac{m!}{(m-n)! \cdot m^n}$$

Grows faster than intuition suggests

Visual Artifacts:

1. **Incidental Clustering**
   - Random points cluster by chance
   - Creates false density peaks (Ellis and Dix 2006)
2. **Excessive Gaps**
   - Large gaps appear randomly
   - May suggest non-existent separation
3. **Non-Reproducibility**
   - Different runs $\rightarrow$ different visualizations
   - Violates scientific principles (Peng 2011)

Analysis - Uniformity Advantages

Why Uniform Distribution Succeeds:

1. Faithful Density Representation

- Visual density $\leftrightarrow$ data frequency
- 20 visible lines $\approx$ 20 observations
- Analysts can trust their eyes (Cleveland and McGill 1984)

2. Minimal Cognitive Load

- No need to mentally "average out" randomness
- Pattern detection more efficient (Ware 2012)

3. Enhanced Traceability

- Individual observations remain followable
- Reduces crossing artifacts (Dennig et al. 2021)
- Enables higher-dimensional insight

Mathematical Support: Low-discrepancy sequences proven optimal for numerical integration (Niederreiter 1992)

Computational Complexity Analysis

Algorithm Performance:

Operation	Complexity	Notes
Identify Ties	$O(n \log n)$	Sorting required
Halton Generation	$O(k)$	Per tie group size k
Sunflower Generation	$O(k)$	Per tie group size k
Apply Displacement	$O(n)$	Linear in observations

Overall: $O(n \log n)$ dominated by tie detection
Practical Performance:

- Tested on datasets up to 100,000 observations
- Real-time interaction maintained
- Memory-efficient implementation
- No performance degradation with large tie groups

Integration with ggpcp Package

User Interface:

```
# Basic usage
iris_long <- iris %>%
  pcp_select(1:4) %>%
  pcp_scale(method = "uniminmax") %>%
  pcp_arrange(method = "sunflower")  # NEW!

# With parameters
pcp_arrange(
  method = "halton",
  epsilon = 0.08,
  numeric_ties = TRUE
)
```

Available Methods: "none", "sunflower", "halton", "intelligent", "random"
Automatic Detection: Identifies numerical ties automatically

Contributions

Theoretical Contributions

1. Biomimetic Algorithm Adaptation

- First application of Vogel's phyllotaxis to statistical visualization
- Demonstrates cross-domain knowledge transfer
- Evolutionary optimization $\rightarrow$ data visualization

2. Number-Theoretic Application

- Halton sequences from computational geometry (Halton 1960)
- Rigorous discrepancy analysis
- Mathematical guarantees on distribution quality

3. Comparative Framework

- First formal comparison of tie-breaking methods for PCPs
- Establishes evaluation criteria
- Provides guidance for future algorithm development

4. Negative Result Documentation

- Intelligent jitter failure scientifically valuable
- Identifies problematic design patterns

Methodological Contributions

Novel Algorithms (3):

1. Sunflower jitter: 2D phyllotaxis $\rightarrow$ 1D displacement
2. Halton jitter: Quasi-random sequences for uniformity
3. Intelligent jitter: Golden ratio direct application

Implementation Quality:

- Production-ready R code
- Comprehensive testing suite
- Integration with existing workflows
- Extensive documentation

Reproducible Research:

- All algorithms fully specified
- Open-source implementation
- Comparative benchmarks provided
- Replication materials available

Practical Contributions

Immediate Utility:

- Solves long-standing practical problem
- Integrated into widely-used package (ggpcp) (VanderPlas et al. 2023)
- Simple function calls, sensible defaults
- Customizable for advanced users

Unified Framework:

- Categorical ties: solved (existing)
- Numerical ties: solved (new)
- Consistent interface across tie types
- Complete implementation of GPCP vision

Impact on Practice:

- Enables trustworthy density visualization
- Supports reproducible research (Peng 2011)
- Improves exploratory data analysis workflows

Empirical Contributions (Planned)

User Study Contributions:

1. Quantitative Evidence

- First controlled evaluation of numerical tie-breaking
- Multi-dimensional assessment (accuracy, time, preference)
- Statistical rigor in comparison

2. Perceptual Research

- How do users interpret jittered visualizations?
- What level of displacement is optimal?
- Do users notice systematic vs. random patterns?

3. Design Recommendations

- Evidence-based guidelines for practitioners
- Optimal epsilon values for different contexts

4. Baseline for Future Work

- Establishes performance benchmarks
- Enables comparative evaluation of new methods

Future Work

Adaptive Epsilon Selection

Challenge: Current epsilon is user-specified constant

Goal: Develop automatic, context-aware epsilon selection

Proposed Approach:

1. Data-Driven Heuristics

- Based on tie group size distribution
- Consider axis range and scale
- Account for display resolution

2. Perceptual Models

- Just-noticeable-difference thresholds (Weber 1834)
- Crowding effects in visual perception (Ware 2012)
- Screen-space calculations

3. Optimization Framework

$$\epsilon^* = \arg\min_{\epsilon} \left[\alpha \cdot \mathsf{Occlusion}(\epsilon) + \beta \cdot \mathsf{Distortion}(\epsilon)\right]$$

2D Generalization

Motivation: Scatter plots face similar overplotting issues (Cleveland and McGill 1984)
Current Practice:

- Random jitter with transparency
- Hexagonal binning
- Density contours

Proposed Extension:

1. 2D Sunflower

- Use full (r, θ) coordinates without projection
- Natural 2D spiral distribution (Vogel 1979)

2. 2D Halton

- Base-2 for x-axis, base-3 for y-axis
- Well-studied in computer graphics (Niederreiter 1992)

Research Questions: How does 2D deterministic jitter compare to current methods?

Performance Optimization

Current Performance: Acceptable but improvable

Optimization Opportunities:

1. Algorithmic Improvements

- Tie detection via hash tables: $O(n)$ average case
- Parallel processing for independent tie groups
- GPU acceleration for large datasets

2. Implementation Efficiency

- Vectorized operations in R
- C++ backend for critical paths
- Memory-efficient data structures

3. Scalability Targets

- Current: 100K observations tested
- Goal: 10M+ observations with real-time interaction
- Streaming data support

Extended Evaluation

Beyond Initial User Study:

1. Domain-Specific Studies

- Medical data (EHR, patient monitoring)
- Financial data (high-frequency trading)
- Scientific data (genomics, climate)

2. Expert User Evaluation

- Experienced data analysts
- Domain scientists
- Professional data journalists

3. Longitudinal Studies

- Learning effects over time
- Integration into real workflows

4. Alternative Tasks

- Hypothesis generation
- Anomaly detection (Johansson et al. 2005)
- Model diagnostics

Conclusion

Summary of Problem

The Challenge We Addressed:
Numerical ties in parallel coordinate plots create severe visual occlusion:

- Single visible line may represent hundreds of observations
- Density information completely lost
- Exploratory analysis fundamentally compromised

Previous State:

- Categorical ties: systematically handled (VanderPlas et al. 2023)
- Numerical ties: no formal solution
- Ad-hoc approaches: random jitter with known problems

Significance:
PCPs are fundamental tool for high-dimensional data analysis (Inselberg 1985; Wegman 1990). This gap prevented their full potential from being realized.

Summary of Solution

Our Systematic Approach:

1. Theoretical Foundation

- Adapted Vogel's phyllotaxis model (Vogel 1979)
- Applied Halton sequences (Halton 1960)
- Established comparative framework

2. Implementation

- Three deterministic algorithms developed
- Integrated into ggpcp R package (VanderPlas et al. 2023)
- Production-ready, tested code

3. Evaluation

- Comprehensive user study designed
- Multi-dimensional assessment
- Quantitative evidence generation

Result: Complete, theoretically-grounded framework for all tie types in PCPs

Key Findings

Evidence-Based Conclusions:

Low-discrepancy methods superior

- Sunflower and Halton jitter outperform random jitter
- Uniform distribution = faithful density representation
- Mathematical guarantees translate to perceptual benefits

Determinism essential

- Reproducibility in scientific visualization (Peng 2011)
- Predictable, interpretable results

Linear scaling problematic

- Intelligent jitter demonstrates negative result
- Excessive displacement distorts perception

General Principle: Uniformity + determinism together solve tie resolution

Impact

Immediate Impact:

For Researchers:

- Reliable density visualization in PCPs
- Reproducible figures for publications (Peng 2011)
- Enhanced exploratory data analysis

For Practitioners:

- Simple implementation (one function call)
- Sensible defaults, customizable parameters
- Integrated into existing workflows

For the Field:

- Conntinue GPCP framework (VanderPlas et al. 2023)
- Establishes evaluation methodology
- Provides reusable design patterns

Long-term Impact: Extension to other visualization contexts

Timeline

Timeline & Milestones

Table 4: Ph.D. Research Timeline (Target Defense: July 2026)

Phase	Timeframe	Milestones	Deliverables
Comprehensive Exam	Fall 2025	Presentation & approval	Updated proposal
Algorithm Refinement	Winter 2025–Spring 2026	Adaptive epsilon, optimization	Refined jittering module
User Study	Spring–Summer 2026	Conduct study, analyze results	Empirical manuscript
Dissertation Writing	May–June 2026	Integrate findings	Complete draft
Defense	July 2026	Committee review, defense	Final submission

Immediate Next Steps

Priority Actions (Next 3 Months):

1. IRB Approval

- Submit revised protocol
- Secure approval for human subjects research

2. Adaptive Epsilon Development

- Implement heuristic algorithms
- Test on diverse datasets

3. Conference Abstract Preparation

- Draft for SDSS 2026
- Target submission: Spring 2026

4. Software Documentation

- Complete package vignettes
- Prepare for CRAN submission

Conclusion

Broader Implications

Beyond Parallel Coordinates:

1. Scatter Plots

- 2D overplotting problem similar
- Sunflower jitter directly applicable (Cleveland and McGill 1984)

2. Time Series Visualization

- Repeated measurements create ties
- Deterministic methods preferable

3. Network Visualization

- Node positioning with constraints
- Similar uniformity requirements

4. General Principle

Wherever stochastic jitter is used, low-discrepancy alternatives should be considered

Questions for Discussion

Open Questions for Committee:

1. Should adaptive epsilon be user-overrideable or always automatic?
2. Should package default to Sunflower or Halton?
3. Additional task types or datasets for user study?
4. Should dissertation include 2D scatter plot extension?
5. Publication strategy - single comprehensive vs. multiple focused papers?

Acknowledgments

Thank you:

- Dissertation committee for guidance and feedback
- ggpcp package developers
- UNL Department of Statistics
- Pilot study participants
- Open-source R community

Contact Information:

- Email: denise.bradford@huskers.unl.edu
- GitHub: https://github.com/drbradford12/Dissertation-Data

Questions?

References

References i

Cleveland, William S, and Robert McGill. 1984. *Graphical Methods for Data Presentation: Full Scale Breaks, Dot Charts, and Multibased Logging*. *The American Statistician*. Vol. 38. 4. Taylor & Francis. https://doi.org/10.1080/00031305.1984.10483223.

Dennig, Frederik L, Tom Polk, Zeyuan Lan, Michael Han, Kathrin Ballweg, Michael Merz, Tobias Schreck, Daniel A Keim, and Min Chen. 2021. "ParSetgnostics: Quality Metrics for Parallel Sets." *Computer Graphics Forum* 40 (3): 375–86. https://doi.org/10.1111/cgf.14314.

Ellis, Geoffrey P, and Alan J Dix. 2006. "Enabling Automatic Clutter Reduction in Parallel Coordinate Plots." *IEEE Transactions on Visualization and Computer Graphics* 12 (5): 717–24. https://doi.org/10.1109/TVCG.2006.138.

Halton, John H. 1960. "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals." *Numerische Mathematik* 2 (1): 84–90. https://doi.org/10.1007/BF01386213.

Inselberg, Alfred. 1985. "The Plane with Parallel Coordinates." *The Visual Computer* 1 (2): 69–91. https://doi.org/10.1007/BF01898350.

Johansson, Jimmy, and Camilla Forsell. 2016. "Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research." *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 579–88. https://doi.org/10.1109/TVCG.2015.2466992.

Johansson, Jimmy, Patric Ljung, Mikael Jern, and Matthew Cooper. 2005. "Revealing Structure Within Clustered Parallel Coordinates Displays." *IEEE Symposium on Information Visualization*, 125–32. https://doi.org/10.1109/INFVIS.2005.1532141.

Kosara, Robert, Fabian Bendix, and Helwig Hauser. 2006. "Parallel Sets: Interactive Exploration and Visual Analysis of Categorical Data." *IEEE Transactions on Visualization and Computer Graphics* 12 (4): 558–68. https://doi.org/10.1109/TVCG.2006.76.

Niederreiter, Harald. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9781611970081.

Peng, Roger D. 2011. "Reproducible Research in Computational Science." *Science* 334 (6060): 1226–27. https://doi.org/10.1126/science.1213847.

Ridley, John N. 1982. "Packing Efficiency in Sunflower Heads." *Mathematical Biosciences* 58 (1): 129–39. https://doi.org/10.1016/0025-5564(82)90056-6.

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): 1405–20. https://doi.org/10.1080/10618600.2023.2181762.

Vogel, Helmut. 1979. “A Better Way to Construct the Sunflower Head.” *Mathematical Biosciences* 44 (3-4): 179–89. https://doi.org/10.1016/0025-5564(79)90080-4.

Ware, Colin. 2012. *Information Visualization: Perception for Design*. 3rd ed. Waltham, MA: Morgan Kaufmann.

Weber, Ernst Heinrich. 1834. *De Pulsu, Resorptione, Auditu Et Tactu: Annotationes Anatomicae Et Physiologicae*. Leipzig: Koehler.

Wegman, Edward J. 1990. “Hyperdimensional Data Analysis Using Parallel Coordinates.” *Journal of the American Statistical Association* 85 (411): 664–75. https://doi.org/10.1080/01621459.1990.10474926.

Wegman, Edward J, and Qiang Luo. 1991. “High Dimensional Clustering Using Parallel Coordinates and the Grand Tour.” *Computing Science and Statistics* 28: 352–60.