The Journal of Open Source Software

ggpedigree: Visualizing Pedigrees with 'ggplot2' and 'plotly'

S. Mason Garrison [1¶]

1 Department of Psychology, Wake Forest University, North Carolina, USA ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Pedigree visualization is essential across multiple research disciplines, including human genetics, animal breeding, genealogical research, forensic science, and counseling. Human geneticists use pedigree diagrams to trace disease inheritance patterns and identify at-risk individuals in families affected by genetic disorders. Animal breeders rely on pedigree visualization to track lineages, plan breeding programs, and optimize genetic traits across generations of livestock and crops. Genealogical researchers use family trees to document ancestral relationships and explore heritage, while forensic scientists analyze pedigree structures to establish familial connections in criminal investigations. Family therapists and counselors use genograms to visualize family dynamics, relationship patterns, and psychological factors across generations to inform therapeutic interventions.

Traditional pedigree plotting tools such as kinship2 (**?**) have served these communities well for basic family structures.

Recently, these fields have expanded to analyze increasingly complex family structures, including large-scale plant breeding pedigrees (**?**), web-based pedigree management systems (**?**), interactive pedigree editors (**?**), and behavior genetic studies of extended family structures (Garrison et al., 2023; Hunter, Garrison, Burt, & Rodgers, 2021). These developments have highlighted the limitations of existing pedigree visualization tools, which often struggle to handle large datasets and complex relationships.

`ggpedigree` simplifies the visualization of these complex family structures by offering a comprehensive suite of functions for creating publication-quality pedigree plots of any size and complexity.

Statement of need

Pedigree visualization has traditionally relied on specialized software or R packages like kinship2 (Sinnwell, Therneau, & Schaid, 2014) or pedtricks that use base R graphics. While these tools are functional, they have several limitations that become apparent when working with complex, modern pedigree datasets. First, base R graphics lack the modularity and extensibility that researchers need for publication-quality figures. Second, existing tools offer limited interactivity, making it difficult to explore large pedigrees dynamically. Third, most current solutions do not integrate well with modern data science workflows that rely heavily on the tidyverse and ggplot2 ecosystem (**?**).

The shift toward analyzing extended family designs in behavior genetics (Garrison et al., 2023; **?**) has created new visualization challenges. Modern pedigree datasets can include hundreds or thousands of individuals across multiple generations, with complex relationship patterns that are difficult to represent clearly. Additionally, researchers increasingly need to overlay

phenotypic information, genetic relatedness values, and other analytical results onto pedigree diagrams for interpretive purposes.

Several existing R packages provide pedigree visualization capabilities. The kinship2 package (Sinnwell et al., 2014) is widely used but limited to static base R plots. (They have an alpha version that can work with ggplot, but it is not vectorized). The HaplotypePlot package focuses on haplotype visualization rather than general pedigree structure. The pedgene package provides some plotting functions but is primarily designed for association testing. None of these packages offers the combination of modern ggplot2 integration, interactive capabilities, and extensibility that ggpedigree provides. ggpedigree addresses these limitations by providing a comprehensive visualization framework built on modern R graphics infrastructure. It leverages the extensive customization capabilities of ggplot2 while adding specialized functionality for pedigree-specific visualization challenges.

Software Architecture

ggpedigree is built on a modular architecture that separates data processing, layout calculation, and visualization layers. The core workflow involves: (1) data standardization and family structure analysis using BGmisc functions, (2) coordinate calculation using algorithms adapted from kinship2, (3) relationship connection mapping, and (4) layer-based plot construction using ggplot2 geometry functions. This design allows users to customize any aspect of the visualization while maintaining computational efficiency for large pedigrees. The package integrates tightly with the broader R ecosystem, particularly the tidyverse (**?**) and BGmisc (**?**). All functions return standard R objects (ggplot or plotly) that can be further customized using familiar syntax, ensuring accessibility for users already comfortable with modern R data science workflows.

Features

The `ggpedigree` package offers comprehensive visualization capabilities organized into several main categories:

Pedigree Visualization and Customization

- Data Standardization and Family Structure Analysis: `ggPedigree()` integrates with BGmisc functions like `ped2fam()` to organize individuals by family, `recodeSex()` to standardize sex coding, and `checkParentIDs()` to validate pedigree structures. The function handles complex scenarios including consanguineous relationships and individuals appearing in multiple pedigree positions.
- Coordinate Calculation: `calculateCoordinates()` computes optimal positioning for individuals using algorithms adapted from `kinship2::align.pedigree`, with enhancements for large multi-generational pedigrees and complex family structures. The function returns coordinate grids that optimize spacing and minimize visual overlap. These steps are vectorized as much as possible to ensure computational efficiency, especially for large pedigrees.
- Relationship Connection Mapping: `calculateConnections()` generates connection paths between family members, mapping parent-child, sibling, spousal, and twin relationships. The function determines midpoints for line intersections and handles overlapping connections with specialized curved segments. These calculations are optimized for large datasets by using vectorized operations rather than the loop-based approaches used in kinship2.
- Layer-based Plot Construction: `ggPedigree()` constructs plots using ggplot2 geometry functions, returning standard ggplot2 objects that integrate with existing R workflows. ggPedigreeInteractive() extends plots into interactive plotly widgets with hover tooltips

and dynamic exploration capabilities. A comprehensive config system allows customization of over 100 aesthetic and layout parameters.

- Focal Individual Highlighting: Advanced functionality to highlight specific individuals and their relatives based on additive genetic, mitochondrial, or other relationship matrices.

Specific Visualizations

- Pedigree Plotting: ggPedigree() creates static pedigree plots using ggplot2, supporting complex family structures, multiple generations, and customizable aesthetics. It can handle large pedigrees with thousands of individuals while maintaining clarity and readability.
- Interactive Pedigree Visualization: `ggPedigreeInteractive()` generates interactive pedigree plots using plotly, allowing users to explore large pedigrees dynamically. Users can configure tooltip content to display individual IDs, names, phenotypic information, and other relevant data. It supports tooltips, zooming, and panning for detailed exploration of family structures.
- Relatedness Matrix Heatmaps: `ggRelatednessMatrix()` creates customizable heatmap visualizations for genetic relatedness matrices with support for hierarchical clustering, interactive exploration, and seamless integration with BGmisc relatedness calculations.
- Phenotype-Relatedness Analysis: `ggPhenotypeByDegree()` produces specialized plots for visualizing phenotypic correlations as a function of genetic relatedness, including confidence intervals and statistical summaries for quantitative genetic analysis.

Collectively, these tools provide a valuable resource for behavior geneticists and others who work with extended family data. They were developed as part of a grant and have been used in several ongoing projects and forthcoming papers (Burt, 2023; Garrison et al., 2023; Hunter et al., 2023; Lyu et al., 2023) and theses (Lyu, 2023).

Availability

The ggpedigree package is open-source and available on both GitHub at https://github.com/R-Computing-Lab/ggpedigree and the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/package=ggpedigree. It is licensed under the GNU General Public License.

Acknowledgments

The current research is supported by the National Institute on Aging (NIA), RF1-AG073189. The authors would like to thank Michael Hunter for their support and enthusiasm to the development of this package.

References

Burt, S. A. (2023). Mom genes: Leveraging maternal lineage to estimate the contributions of mitochondrial DNA. *Behavior Genetics*. doi:10.1007/s10519-023-10156-9

Garrison, S. M., Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., & Burt, S. A. (2023). Analyzing extended cousin similarity to unravel the mystery of mtDNA and longevity. *Behavior Genetics*. doi:10.1007/s10519-023-10156-9

Hunter, M. D., Garrison, S. M., Burt, S. A., & Rodgers, J. L. (2021). The Analytic Identification of Variance Component Models Common to Behavior Genetics. *Behavior Genetics*, *51*(4), 425–437. doi:10.1007/s10519-021-10055-x

Hunter, M. D., Lyu, X., Garrison, S. M., Rodgers, J. L., Smith, K., Coon, H., & Burt, S. A. (2023). Modeling mtDNA effects from extended pedigrees in the utah population database. *Behavior Genetics*. doi:10.1007/s10519-023-10156-9

Lyu, X. (2023, May). *Statistical power analysis on mtDNA effects estimation* (Master's thesis). Wake Forest University.

Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., Burt, S. A., & Garrison, S. M. (2023). Statistical power analysis on mtDNA effects estimation. *Behavior Genetics*. doi:10.1007/s10519-023-10156-9

Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 r package for pedigree data. *Human Heredity*, *78*, 91–93. doi:10.1159/000363105