

¹ ggpedigree: Visualizing Pedigrees with ‘ggplot2’ and ‘plotly’²

³ S. Mason Garrison  

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

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).¹⁶¹⁷¹⁸¹⁹²⁰

⁵ Summary

Pedigree diagrams underpin research and practice across genetics, animal breeding, genealogy, forensics, and counseling. They help medical geneticists trace the inheritance of Mendelian diseases and identify at-risk relatives; enable dairy breeders to plan matings that improve milk yield; support genealogists in reconstructing ancestry; assist forensic scientists in establishing familial connections in criminal investigations; and facilitate family therapists and counselors in understanding their clients' relationships (McGoldrick, Gerson, & Petry, 2020). Early R tools such as kinship2 (Sinnwell, Therneau, & Schaid, 2014) plot simple nuclear families effectively, but they do not scale to today's pedigrees that can exceed 1,000s of individuals. As datasets have grown, researchers now work with increasingly complex family structures, including large-scale plant breeding pedigrees (Shaw, Graham, Kennedy, Milne, & Marshall, 2014), web-based pedigree management systems (Ranaweera, Makalic, Hopper, & Bickerstaffe, 2018), interactive pedigree editors (Carver et al., 2018), and behavior genetic studies of extended family structures (Garrison et al., 2023; Hunter, Garrison, Burt, & Rodgers, 2021). That complexity exposes the limitations of existing tools, which often struggle with large and complex datasets. ggpedigree addresses this need by combining a vectorised layout algorithm, ggplot2 output, and optional plotly interactivity.²¹

²² Statement of need

Pedigree visualization has traditionally relied on proprietary software (e.g., Progeny, GenoPro, Pedigree Viewer) or R packages like kinship2 (Sinnwell et al., 2014), pedtools (Vigeland, 2021a), or pedtricks (J. Martin, Wolak, Johnston, & Morrissey, 2025). While these tools are functional for many use cases, their limitations become pronounced when working with complex, modern pedigree datasets or when more detailed customization is required. Most R packages focus on base graphics or simple ggplot2 implementations.²³²⁴²⁵²⁶²⁷²⁸

Existing R solutions face three main challenges. First, current solutions are often poorly integrated with tidyverse workflows and do not expose the full theming and layering capabilities familiar to ggplot2 users (Wickham, 2016). In animal-focused workflows, rapid rendering seems to takes precedence over aesthetic flexibility. I suspect that this is because users tend to work with more uniform data and fewer phenotypes. By contrast, human-focused workflows—particularly in behavior genetics and genetic epidemiology (Garrison et al., 2023; Lyu et al., 2025; McArdle & McDonald, 1984)—require closer integration with tidyverse pipelines and more flexible plotting systems to accommodate complex pedigree structures and harmonization of phenotypes across data sources. In other words, the needs are different.²⁹³⁰³¹³²³³³⁴³⁵³⁶³⁷

Second, most R-based tools offer no interactivity. Static graphics are often sufficient for publication, but interactivity improves exploration and communication during model development or data cleaning. A notable exception is pedtools (Vigeland, 2021b), which offers a sister shiny

41 app, QuickPed ([Vigeland, 2022](#)). While the R ecosystem includes libraries, like `plotly`, that
42 support interactive plotting, these features have yet to be integrated into pedigree functions.

43 Third, scalability and extensibility remain limited across existing tools. Several R packages
44 attempt to address these challenges with built-in pedigree plotting functions. `kinship2`
45 ([Sinnwell et al., 2014](#)) remains widely used but produces static base graphics and relies on
46 non-vectorized recursive layout functions that do not scale well to large families. A partial
47 `ggplot2` implementation exists in a modernized `kinship2` (called `Pedixplorer`, [Le Nézet,](#)
48 [Sinnwell, Letko, André, & Quignon, 2025](#)), but is non-vectorized and incompatible with other
49 `ggplot2` layers. `pedtricks`, a revival of `pedantics` ([Morrissey & Wilson, 2010](#)), provides a
50 `ggplot2`-based implementation for large animal pedigrees but lacks extensibility and interactivity.
51 The `geneHapR` ([Zhang, Jia, & Diao, 2023](#)) package focuses on haplotype visualization rather
52 than general pedigree structure. The `pedgene` package ([Schaid & Sinnwell, 2024](#)) offers some
53 plotting functions but is primarily designed for association testing. The `pedigreejs` package
54 ([Carver et al., 2018](#)) provides an interactive pedigree editor but does not integrate with R or
55 `ggplot2`, limiting its utility for R users.

56 None of these packages offers the combination of modern `ggplot2` integration, interactive
57 capabilities, and extensibility that `ggpPedigree` provides. `ggpPedigree` addresses these limitations
58 by providing a comprehensive visualization framework built on modern R graphics infrastructure.
59 It leverages the extensive customization capabilities of `ggplot2` while adding specialized
60 functionality for pedigree-specific visualization challenges.

61 Software Architecture

62 `ggpPedigree` is built on a modular architecture that separates data processing, layout calculation,
63 and visualization layers. The core workflow involves: (1) data standardization and family
64 restructuring using `BGmisc` functions, (2) coordinate calculation using algorithms adapted
65 from `kinship2`, (3) relationship connection mapping, and (4) layer-based plot construction
66 using `ggplot2` geometry functions. This design allows users to customize any aspect of the
67 visualization while maintaining computational efficiency for large pedigrees. The package
68 integrates tightly with the broader R ecosystem, particularly the `tidyverse` ([Wickham et al.,](#)
69 [2019](#)) and `BGmisc` ([Garrison, Hunter, Lyu, Trattner, & Burt, 2024](#)). All functions return
70 standard R objects (`ggplot` or `plotly`) that can be further customized.

71 `BGmisc`, as described in Garrison et al. (2024), is a dependency for its relatedness-heavy
72 workflows, supplying network-based validation utilities (`checkParentIDs()`) and relatedness
73 components, like `ped2fam()`, `ped2paternal()`, and `ped2maternal()`. These components
74 allow `ggpPedigree` to visualize how related any two individuals are based on additive genetic,
75 mitochondrial, or other relationship matrices. Burt et al. (2025) uses these features to create
76 mitochondrial lineages in human pedigrees ($n > 176$ million), finding that mitochondrial DNA
77 explains significant variance in longevity.

78 Features

79 I describe the main features of the `ggpPedigree` package below. Detailed descriptions of features
80 and usage are available in the [package vignettes](#), including how to create static and interactive
81 pedigree plots, customize aesthetics, and visualize relatedness matrices. Additional example
82 data include squirrels from the Kluane Red Squirrel Project ([McFarlane et al., 2014, 2015](#))
83 and Targaryens from the Song of Ice and Fire universe ([G. R. R. Martin, 1997, 2018](#)).

- 84
 - 85 ▪ Data Standardization and Family Structure Analysis: `ggPedigree()` integrates with
86 network-based functions from `BGmisc` like `ped2fam()` to organize individuals by family and
87 `checkParentIDs()` to validate pedigree structures. The function handles consanguineous
88 relationships and individuals appearing in multiple pedigree positions. More details are
89 in the [complex pedigree data vignette](#), as well as in these works ([Garrison et al., 2024;](#)
 [Hunter et al., 2025, 2021](#)).

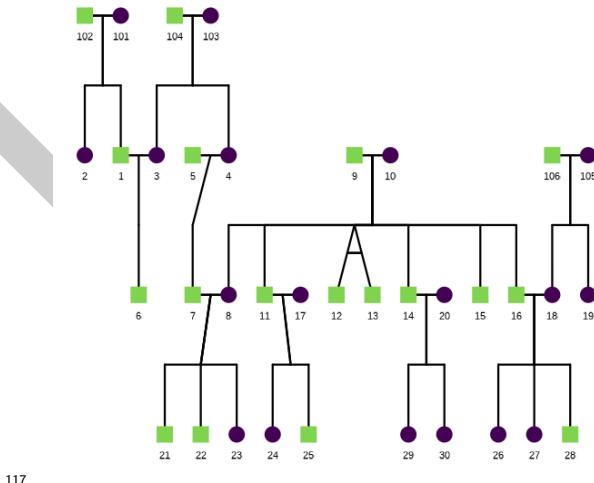
- 90 ▪ Coordinate Calculation: `calculateCoordinates()` computes optimal positioning for individuals using algorithms adapted from `kinship2::align.pedigree`, with enhancements for complex multi-generational pedigrees. These steps are vectorized as much as possible to ensure efficient computation and compatibility with `ggplot2`.
- 91
- 92
- 93
- 94 ▪ 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 curved segments. These calculations are optimized for large datasets by using vectorized operations rather than the loop-based approaches used in `kinship2`.
- 95
- 96
- 97
- 98
- 99 ▪ Layer-based Plot Construction: `ggPedigree()` constructs plots using `ggplot2` geometry functions, returning standard `ggplot2` objects that integrate with existing R workflows.
- 100 `ggPedigreeInteractive()` extends plots into interactive `plotly` widgets. A config system allows customization of over 150 aesthetic and layout parameters. More details are in the [configuration vignette](#).
- 101
- 102
- 103
- 104 ▪ Individual Highlighting: Advanced functionality to highlight specific individuals and their relatives based on additive genetic, mitochondrial, or other relationship matrices.
- 105
- 106 ▪ Specific Visualization Functions: `ggPedigree()` creates static pedigree plots using `ggplot2`. `ggPedigreeInteractive()` generates interactive pedigree plots using `plotly`. `ggRelatednessMatrix()` creates customizable heatmaps for relatedness matrices with support for hierarchical clustering, and seamless integration with `BGmisc` relatedness calculations. `ggPhenotypeByDegree()` visualizes phenotypic correlations as a function of genetic relatedness, including confidence intervals and statistical summaries for quantitative genetic analysis.
- 107
- 108
- 109
- 110
- 111
- 112

113 Code example

114 This example shows how to use `ggpedigree` to visualize a pedigree. The `potter` dataset
 115 includes several wizarding families from the Harry Potter series.

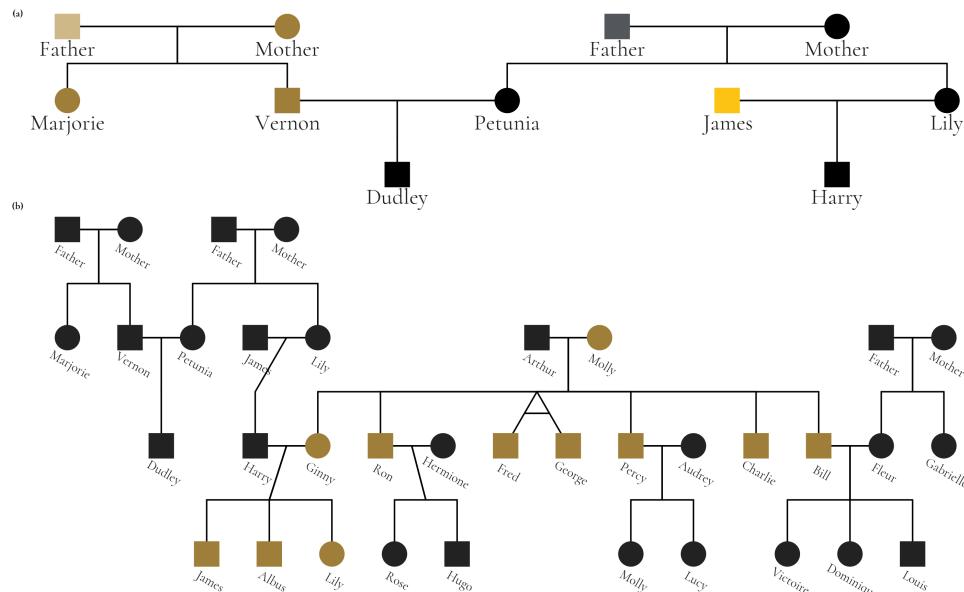
```
116     ggPedigree(potter,
117         famID = "famID",
118         personID = "personID"
119     )
```

116 This code produces the following pedigree plot:



118 The package supports extensive customization of visual aesthetics. The following example
 119 is a figure from Hunter et al. (2025) that used the Potter pedigree data. The figure has

120 been restyled according to Wake Forest University brand identity guidelines to demonstrate
 121 ggpedigree's customization capabilities. The figure combines two panels: panel (a) highlights
 122 unique mitochondrial lines in the Dursley and Evans families, while panel (b) shows the full
 123 pedigree with Molly Weasley's mitochondrial descendants in gold.



124
 125 The complete source code for this example is available in the package documentation [website](#).
 126 Collectively, these tools provide a valuable resource for those work with extended family data.
 127 They were developed as part of a grant and have been used in several ongoing projects,
 128 presentations ([Garrison, 2024](#); [Hunter, Garrison, Lyu, Good, & Burt, 2024](#)), and forthcoming
 129 papers ([Burt et al., 2025](#); [Hunter et al., 2025](#); [Lyu et al., 2025](#)).

130 Availability

131 The ggpedigree package is open-source and available on both [GitHub](#) and the [Comprehensive](#)
 132 [R Archive Network \(CRAN\)](#). It is licensed under the GNU General Public License.

133 Acknowledgments

134 The current research is supported by the National Institute on Aging, [R01-AG073189](#). The
 135 author thanks Michael Hunter for his enthusiasm for the development of this package.

136 References

- 137 Burt, S. A., Garrison, S. M., Lyu, X., Rodgers, J. L., Carroll, S. L., Smith, K. R., &
 138 Hunter, M. D. (2025). Contributions of inherited mtDNA to longevity: Evidence
 139 from extended pedigrees with 176 million kinship pairs. *eBioMedicine*, *119*, 105911.
 140 doi:[10.1016/j.ebiom.2025.105911](https://doi.org/10.1016/j.ebiom.2025.105911)
- 141 Carver, T., Cunningham, A. P., Babb de Villiers, C., Lee, A., Hartley, S., Tischkowitz,
 142 M., Walter, F. M., et al. (2018). Pedigreejs: A web-based graphical pedigree editor.
 143 *Bioinformatics*, *34*(6), 1069–1071. doi:[10.1093/bioinformatics/btx705](https://doi.org/10.1093/bioinformatics/btx705)
- 144 Garrison, S. M. (2024). Charting new paths in behavior genetics: Developing a comprehensive

- 145 R visualization atlas. *Behavior Genetics* (Vol. 54, p. 488488). doi:[10.1007/s10519-024-10197-8](https://doi.org/10.1007/s10519-024-10197-8)
- 146
- 147 Garrison, S. M., Hunter, M. D., Lyu, X., Trattner, J. D., & Burt, S. A. (2024). BGmisc: An R
148 package for extended behavior genetics analysis. *Journal of Open Source Software*, 9(94).
149 doi:[10.21105/joss.06203](https://doi.org/10.21105/joss.06203)
- 150 Garrison, S. M., Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., & Burt, S.
151 A. (2023). Analyzing extended cousin similarity to unravel the mystery of mtDNA and
152 longevity. *Behavior Genetics*. doi:[10.1007/s10519-023-10156-9](https://doi.org/10.1007/s10519-023-10156-9)
- 153 Hunter, M. D., Garrison, S. M., Burt, S. A., & Rodgers, J. L. (2021). The Analytic Identification
154 of Variance Component Models Common to Behavior Genetics. *Behavior Genetics*, 51(4),
155 425–437. doi:[10.1007/s10519-021-10055-x](https://doi.org/10.1007/s10519-021-10055-x)
- 156 Hunter, M. D., Garrison, S. M., Lyu, X., Good, R., & Burt, S. A. (2024). Tools for
157 biometric modeling in large population databases. *Behavior Genetics* (Vol. 54, p. 531531).
158 doi:[10.1007/s10519-024-10197-8](https://doi.org/10.1007/s10519-024-10197-8)
- 159 Hunter, M. D., Garrison, S. M., Lyu, X., Good, R., Carroll, S. L., & Burt, S. A. (2025). *Tracing
160 the right path: Determination of large pedigree segmentation and relatedness. Revise and
161 Resubmit at Behavior Genetics.*
- 162 Le Nézet, L., Sinnwell, J., Letko, A., André, C., & Quignon, P. (2025). Pedixplorer: A
163 Bioconductor package to streamline pedigree design and visualization. *Bioinformatics*,
164 41(6), btaf329. doi:[10.1093/bioinformatics/btaf329](https://doi.org/10.1093/bioinformatics/btaf329)
- 165 Lyu, X., Hunter, M. D., Burt, S. A., Good, R., Carroll, S. L., & Garrison, S. M. (2025). Detecting mtDNA effects with an extended pedigree model: An analysis of statistical power
166 and estimation bias. *Behavior Genetics*, 55, 320–337. doi:[10.1007/s10519-025-10225-1](https://doi.org/10.1007/s10519-025-10225-1)
- 168 Martin, G. R. R. (1997). *A game of thrones*. Bantam.
- 169 Martin, G. R. R. (2018). *Fire & blood: 300 years before a game of thrones*. New York:
170 Random House Worlds.
- 171 Martin, J., Wolak, M., Johnston, S., & Morrissey, M. (2025). Pedtricks: Visualize, summarize
172 and simulate data from pedigrees. doi:[10.32614/CRAN.package.pedtricks](https://doi.org/10.32614/CRAN.package.pedtricks)
- 173 McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action
174 model for moment structures. *British Journal of Mathematical and Statistical Psychology*,
175 37, 234–251. doi:[10.1111/j.2044-8317.1984.tb00802.x](https://doi.org/10.1111/j.2044-8317.1984.tb00802.x)
- 176 McFarlane, S. E., Boutin, S., Humphries, M. M., McAdam, A. G., Gorrell, J. C., & Coltmann,
177 D. W. (2015, January 21). Data from: Very low levels of direct additive genetic
178 variance in fitness and fitness components in a red squirrel population. Dryad.
179 doi:[10.5061/DRYAD.N5Q05](https://doi.org/10.5061/DRYAD.N5Q05)
- 180 McFarlane, S. E., Gorrell, J. C., Coltmann, D. W., Humphries, M. M., Boutin, S., & McAdam,
181 A. G. (2014). Very low levels of direct additive genetic variance in fitness and fitness
182 components in a red squirrel population. *Ecology and Evolution*, 4(10), 1729–1738.
183 doi:[10.1002/ece3.982](https://doi.org/10.1002/ece3.982)
- 184 McGoldrick, M., Gerson, R., & Petry, S. (2020). *Genograms: Assessment and Treatment*.
185 Erscheinungsort nicht ermittelbar: W. W. Norton & Company.
- 186 Morrissey, M. B., & Wilson, A. J. (2010). pedantics: an R package for pedigree-based genetic
187 simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology
188 Resources*, 10(4), 711–719. doi:[10.1111/j.1755-0998.2009.02817.x](https://doi.org/10.1111/j.1755-0998.2009.02817.x)
- 189 Ranaweera, T., Makalic, E., Hopper, J. L., & Bickerstaffe, A. (2018). An open-source,
190 integrated pedigree data management and visualization tool for genetic epidemiology.
191 *International Journal of Epidemiology*, 47(4), 1034–1039. doi:[10.1093/ije/dyy049](https://doi.org/10.1093/ije/dyy049)

- 192 Schaid, D., & Sinnwell, J. (2024). Pedgene: Gene-level variant association tests for pedigree
193 data. doi:[10.32614/CRAN.package.pedgene](https://doi.org/10.32614/CRAN.package.pedgene)
- 194 Shaw, P. D., Graham, M., Kennedy, J., Milne, I., & Marshall, D. F. (2014). Helium: Visualization
195 of large scale plant pedigrees. *BMC Bioinformatics*, 15(1), 259. doi:[10.1186/1471-2105-15-259](https://doi.org/10.1186/1471-2105-15-259)
- 196 Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 R package for pedigree
197 data. *Human Heredity*, 78, 91–93. doi:[10.1159/000363105](https://doi.org/10.1159/000363105)
- 198 Vigeland, M. D. (2021b). *Pedigree analysis in R*. London: Academic Press, an imprint of
199 Elsevier. Retrieved from <https://shop.elsevier.com/books/pedigree-analysis-in-r/vigeland/978-0-12-824430-2>
- 200 Vigeland, M. D. (2021a). *Pedigree analysis in R*. doi:[10.32614/CRAN.package.pedtools](https://doi.org/10.32614/CRAN.package.pedtools)
- 201 Vigeland, M. D. (2022). QuickPed: An online tool for drawing pedigrees and analysing
202 relatedness. *BMC Bioinformatics*, 23(1), 220. doi:[10.1186/s12859-022-04759-y](https://doi.org/10.1186/s12859-022-04759-y)
- 203 Wickham, H. (2016). *ggplot2: Use R!* Cham: Springer International Publishing.
204 doi:[10.1007/978-3-319-24277-4](https://doi.org/10.1007/978-3-319-24277-4)
- 205 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G.,
206 et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
207 doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686)
- 208 Zhang, R., Jia, G., & Diao, X. (2023). geneHapR: An r package for gene haplotypic statistics
209 and visualization. *BMC Bioinformatics*, 24(1), 199. doi:[10.1186/s12859-023-05318-9](https://doi.org/10.1186/s12859-023-05318-9)
- 210
- 211