




# ggpedigree: Visualizing Pedigrees with ‘ggplot2’ and ‘plotly’

S. Mason Garrison <sup>1</sup>✉

<sup>1</sup> 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Published: unpublished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](#)).

## 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.

## 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.

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 take 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.

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 non-  
46 vectorized recursive layout functions that do not scale well to large families. A partial ggplot2  
47 implementation exists in a modernized kinship2 (called Pedexplorer, Le Nézet, Sinnwell,  
48 Letko, André, & Quignon, 2025), but is non-vectorized and incompatible with other ggplot2  
49 layers. pedtricks, a revival of pedantics (Morrissey & Wilson, 2010), provides a ggplot2-  
50 based implementation for large animal pedigrees but lacks extensibility and interactivity. The  
51 geneHapR (Zhang, Jia, & Diao, 2023) package focuses on haplotype visualization rather than  
52 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 ggpedigree provides. ggpedigree 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 ggpedigree 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 structure analysis 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, S. M., Hunter, Michael D., Lyu, X., Trattner, J. D., & Burt,  
70 S. A., 2024). All functions return standard R objects (ggplot or plotly) that can be further  
71 customized.

## 72 Features

73 I describe the main features of the ggpedigree package below. More detailed descriptions of  
74 features and usage are available in the [package vignettes](#), including examples of how to create  
75 static and interactive pedigree plots, customize aesthetics, and visualize relatedness matrices.  
76 Additional example data include squirrel data from the Kluane Red Squirrel Project (McFarlane  
77 et al., 2014, 2015) and Targaryen family data from the Song of Ice and Fire universe (G. R. R.  
78 Martin, 1997, 2018).

- 79 ■ Data Standardization and Family Structure Analysis: ggPedigree() integrates with  
80 BGmisc functions like ped2fam() to organize individuals by family, recodeSex() to  
81 standardize sex coding, and checkParentIDs() to validate pedigree structures. The  
82 function handles consanguineous relationships and individuals appearing in multiple  
83 pedigree positions.
- 84 ■ Coordinate Calculation: calculateCoordinates() computes optimal positioning for indi-  
85 viduals using algorithms adapted from kinship2::align.pedigree, with enhancements  
86 for large multi-generational pedigrees and complex family structures. These steps are  
87 vectorized as much as possible to ensure computational efficiency and compatibility with  
88 ggplot2.

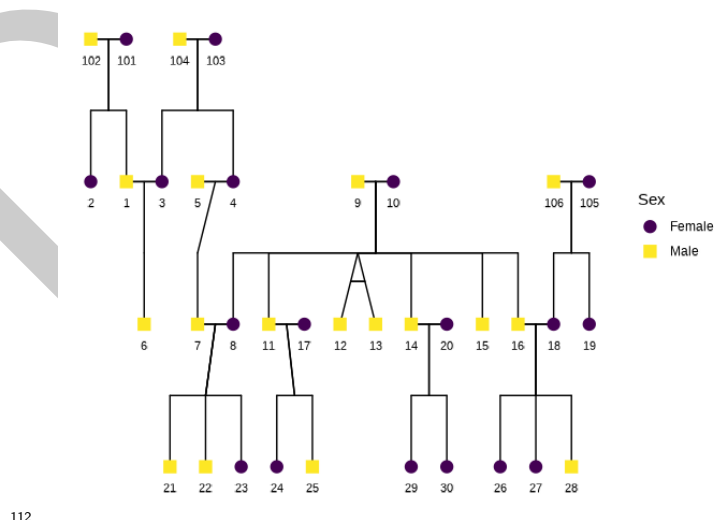
- 89 ■ Relationship Connection Mapping: `calculateConnections()` generates connection paths  
90 between family members, mapping parent-child, sibling, spousal, and twin relationships.  
91 The function determines midpoints for line intersections and handles overlapping con-  
92 nections with specialized curved segments. These calculations are optimized for large  
93 datasets by using vectorized operations rather than the loop-based approaches used in  
94 `kinship2`.
- 95 ■ Layer-based Plot Construction: `ggPedigree()` constructs plots using `ggplot2` geometry  
96 functions, returning standard `ggplot2` objects that integrate with existing R workflows.  
97 `ggPedigreeInteractive()` extends plots into interactive `plotly` widgets. A config system  
98 allows customization of over 100 aesthetic and layout parameters.
- 99 ■ Individual Highlighting: Advanced functionality to highlight specific individuals and their  
100 relatives based on additive genetic, mitochondrial, or other relationship matrices.
- 101 ■ Specific Visualization Functions: `ggPedigree()` creates static pedigree plots using  
102 `ggplot2`. `ggPedigreeInteractive()` generates interactive pedigree plots using `plotly`.  
103 `ggRelatednessMatrix()` creates customizable heatmaps for relatedness matrices with  
104 support for hierarchical clustering, and seamless integration with `BGmisc` relatedness  
105 calculations. `ggPhenotypeByDegree()` visualizes phenotypic correlations as a function  
106 of genetic relatedness, including confidence intervals and statistical summaries for  
107 quantitative genetic analysis.

#### 108 Code example

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

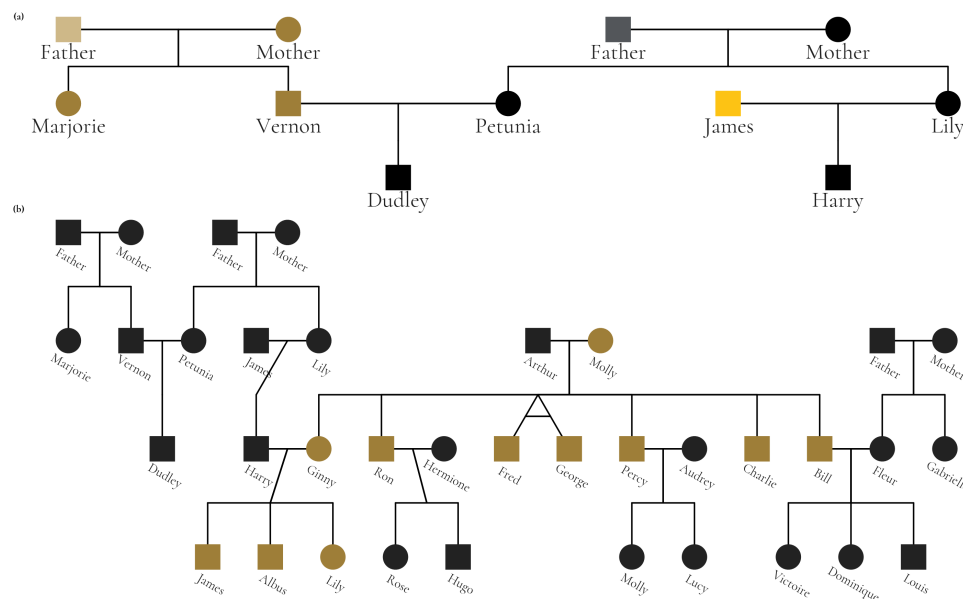
```
ggPedigree(potter,
  famID = "famID",
  personID = "personID"
)
```

111 This code produces the following pedigree plot:



112  
113 The package supports extensive customization of visual aesthetics. The following example  
114 applies the family lineage tracing algorithms from Hunter et al. (Hunter et al., 2025) and  
115 `BGmisc` to the Potter pedigree data. The figure has been restyled according to Wake Forest  
116 University brand identity guidelines to demonstrate `ggpedigree`'s customization capabilities.

117 The figure combines two panels: panel (a) highlights unique mitochondrial lines in the Dursley  
118 and Evans families, while panel (b) shows the full pedigree with Molly Weasley's mitochondrial  
119 descendants in gold.



120

121 The complete source code for this example is available in the package documentation at  
122 <https://r-computing-lab.github.io/ggpedigree/articles/>.

123 Collectively, these tools provide a valuable resource for behavior geneticists and others who  
124 work with extended family data. They were developed as part of a grant and have been used  
125 in several ongoing projects, presentations (Garrison, 2024; Hunter, Garrison, Lyu, Good, &  
126 Burt, 2024), and forthcoming papers (Burt et al., 2025; Hunter et al., 2025; Lyu et al., 2025).

## 127 Availability

128 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  
129 <https://cran.r-project.org/package=ggpedigree>. It is licensed under the GNU General Public  
130 License.  
131

## 132 Acknowledgments

133 The current research is supported by the National Institute on Aging (NIA), RF1-AG073189.  
134 The author would like to thank Michael Hunter for his enthusiasm for the development of this  
135 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 from  
139 extended pedigrees with 176 million kinship pairs. *eBioMedicine*, 119, 105911. doi:<https://doi.org/10.1016/j.ebiom.2025.105911>  
140  
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-](https://doi.org/10.1007/s10519-024-10197-8)  
146 [10197-8](https://doi.org/10.1007/s10519-024-10197-8)
- 147 Garrison, S. M., Hunter, Michael D., Lyu, X., Trattner, J. D., & Burt, S. A. (2024). BGmisc:  
148 An r package for extended behavior genetics analysis. *Journal of Open Source Software*,  
149 9(94). 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). Pedexplorer: 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).  
166 Detecting mtDNA effects with an extended pedigree model: An analysis of statistical power  
167 and estimation bias. *Behavior Genetics*. 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* (Vol. 1). 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., & Colt-  
177 man, D. W. (2015, January 21). Data from: Very low levels of direct additive ge-  
178 netic 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., Coltman, 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: Visualiza-  
195 tion of large scale plant pedigrees. *BMC Bioinformatics*, 15(1), 259. doi:[10.1186/1471-](https://doi.org/10.1186/1471-2105-15-259)  
196 [2105-15-259](https://doi.org/10.1186/1471-2105-15-259)
- 197 Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 r package for pedigree  
198 data. *Human Heredity*, 78, 91–93. doi:[10.1159/000363105](https://doi.org/10.1159/000363105)
- 199 Vigeland, M. D. (2021b). *Pedigree analysis in r*. London: Academic Press, an imprint of  
200 Elsevier. Retrieved from [https://shop.elsevier.com/books/pedigree-analysis-in-r/vigeland/](https://shop.elsevier.com/books/pedigree-analysis-in-r/vigeland/978-0-12-824430-2)  
201 [978-0-12-824430-2](https://shop.elsevier.com/books/pedigree-analysis-in-r/vigeland/978-0-12-824430-2)
- 202 Vigeland, M. D. (2021a). *Pedigree analysis in {r}*. doi:[10.32614/CRAN.package.pedtools](https://doi.org/10.32614/CRAN.package.pedtools)
- 203 Vigeland, M. D. (2022). QuickPed: An online tool for drawing pedigrees and analysing  
204 relatedness. *BMC Bioinformatics*, 23(1), 220. doi:[10.1186/s12859-022-04759-y](https://doi.org/10.1186/s12859-022-04759-y)
- 205 Wickham, H. (2016). *ggplot2*. Use R! Cham: Springer International Publishing.  
206 doi:[10.1007/978-3-319-24277-4](https://doi.org/10.1007/978-3-319-24277-4)
- 207 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G.,  
208 et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.  
209 doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686)
- 210 Zhang, R., Jia, G., & Diao, X. (2023). geneHapR: An r package for gene haplotypic statistics  
211 and visualization. *BMC Bioinformatics*, 24(1), 199. doi:[10.1186/s12859-023-05318-9](https://doi.org/10.1186/s12859-023-05318-9)