Advances in Archaeological Practice (2024), 1–7
doi:10.1017/aap.2024.10

SAA
Society for American Archaeology

HOW-TO SERIES

The *tesselle* Project: A Collection of R Packages for Research and Teaching in Archaeology

Nicolas Frerebeau

UMR 6034 Archéosciences Bordeaux, Maison de l'Archéologie, Université Bordeaux Montaigne, Pessac, France
Email: nicolas.frerebeau@u-bordeaux-montaigne.fr

Abstract

The use of programming languages in archaeological research has witnessed a notable surge in the last decade, particularly with R, a versatile statistical computing language that fosters the development of specialized packages. This article introduces the *tesselle* project (https://www.tesselle.org/), a comprehensive collection of R packages tailored for archaeological research and education. The *tesselle* packages are centered on quantitative analysis methods specifically crafted for archaeology. They are designed to complement both general-purpose and other specialized statistical packages. These packages serve as a versatile toolbox, facilitating the exploration and analysis of common data types in archaeology—such as count data, compositional data, or chronological data—and enabling the construction of reproducible workflows. Complementary packages for visualization, data preparation, and educational resources augment the *tesselle* ecosystem. This article outlines the project's inception, its objectives, design principles, and key components, along with reflections on future directions.

Resumen

El uso de lenguajes de programación en arqueología ha experimentado un notable aumento en la última década, especialmente con R, un lenguaje de computación estadística versátil que fomenta el desarrollo de paquetes especializados. El proyecto *tesselle* (https://www.tesselle.org/) es una colección completa de paquetes de R adaptados para la investigación arqueológica y la educación. Este artículo describe el inicio del proyecto, sus objetivos, principios de diseño y componentes clave, junto con reflexiones sobre las direcciones futuras. Los paquetes de *tesselle* se centran en métodos de análisis cuantitativos específicamente diseñados para la arqueología. Están diseñados para complementar tanto paquetes estadísticos de propósito general como otros especializados. Estos paquetes sirven como un conjunto de herramientas versátil, facilitando la exploración y análisis de tipos de datos comunes en arqueología, como datos de recuento, datos composicionales o datos cronológicos, y permiten la construcción de flujos de trabajo reproducibles. Paquetes complementarios para visualización, preparación de datos y recursos educativos complementan el ecosistema de *tesselle*.

Keywords: archaeology; archaeometry; statistics; reproducibility; open science; R package

Palabras clave: arqueología; arqueometría; estadística; reproductibilidad; ciencia abierta; paquete R

Over the last decade, R has gradually emerged as the archaeological language for data analysis (Schmidt and Marwick 2020). This trend reflects a renewed interest in formal approaches and the use of statistics in archaeology. Moreover, this usage is situated in a particular context. The 2010s witnessed two major reflections, though not unrelated to each other. The first is the recognition of the reproducibility crisis spanning all disciplines (Baker 2016; Ioannidis 2005), including archaeology (Karoune and Plomp 2022; Marwick 2017). The second is the burgeoning open science movement, supported by national and international initiatives, with varying degrees of institutional commitment. All of this unfolds against a backdrop of substantial growth in the volume of collected and processed archaeological data.

R (R Core Team 2023) is a programming language for statistical computing released under the GNU General Public License. The freedoms offered by the GNU license and the modular structure of R allow the development of packages that provide additional functionality, usually dedicated to a specific task, making R a versatile tool.

The *tesselle* project (https://www.tesselle.org; Figure 1) is a collection of R packages for research and teaching in archaeology that emerged from this evolving landscape of research practices. This article describes the *tesselle* project, its objectives, and the main principles of its design, and it provides some reflections on the encountered challenges and future directions. The aim is to offer an overview of the project, presenting its key components, and encouraging the reader to continue exploring the documentation.

Motivation

The increasing use of R poses a number of new challenges for the archaeological community. These challenges can be grouped into two main categories, allowing us to distinguish between intrinsic and extrinsic difficulties in the use of programming languages—including R—in archaeology. Intrinsic difficulties pertain to the very use of programming languages and, in general, the use of

Figure 1. Logos of the *tesselle* packages (CC-BY 4.0).

https://doi.org/10.1017/aap.2024.10 Published online by Cambridge University Press

any research software, especially in light of the challenges of research reproducibility. Despite all the care taken during development, no software is entirely free from bugs. Similarly, unintended uses by users can lead to unexpected, if not erroneous, results. Finally, each language—and by extension, each software—has its life cycle: more or less significant changes may occur over time (whether or not visible to the user), and maintenance may also cease.

This latter point echoes what can be termed as extrinsic difficulties: those that do not directly relate to the use of programming languages but to the organization and functioning of the archaeological discipline as a community. Baptiste and Roe (2021) have highlighted the fragility of open-source archaeology: most projects have a short lifespan and rely on precarious work, that which often lacks professional and institutional recognition. Additionally, there is the issue of training for archaeologists; as emphasized by Schmidt and Marwick (2020), it is unlikely that established professionals would be motivated to program if they had not been trained to work with code early in their career. Reflection should be undertaken at the institutional level on how digital tools are becoming prominent in the professional context (Tufféry 2019) and on the additional workload that open science may represent (Hostler 2023).

The *tesselle* project was conceived as an attempt to respond to some of the intrinsic challenges associated with using R in archaeology. In doing so, the project encounters the same extrinsic challenges as the rest of open-source archaeology. This project is driven by two primary objectives: to move away from proprietary environments and advance toward more transparent and open methodologies in archaeological research. At the time of writing this article, there are over 20,000 packages available on the Comprehensive R Archive Network (CRAN; https://cran.r-project.org), providing a vast array of tools to meet most analytical needs. Furthermore, owing to the collective efforts of the community, a wealth of high-quality packages tailored to archaeology have been developed (for a comprehensive list, see Marwick et al. 2022).

The *tesselle* packages are centered on quantitative analysis methods specifically crafted for archaeology. They are designed to complement both general-purpose and other specialized statistical packages. These packages serve as a versatile toolbox, facilitating the exploration and analysis of common data types in archaeology—such as count data, compositional data, or chronological data—and enabling the construction of reproducible workflows.

Additionally, the project was designed with a focus on university-level teaching. Although this last point requires an in-depth discussion beyond the scope of this article, it is worth noting that improved statistical and scientific programming training contributes to addressing research reproducibility issues (Munafò et al. 2017). Numerous teaching resources are available (e.g., Carlson 2017), but the importance of these courses appears to vary widely across archaeology programs.[1] The *tesselle* project also aims to help novice programmers start analyzing their data in R by offering a consistent toolbox.

Design Principles

The design of the *tesselle* project and its packages drew inspiration from certain aspects of the *tidyverse* (https://www.tidyverse.org)—particularly its emphasis on prioritizing end users, given that R is primarily used by nonprogrammers (Wickham et al. 2019). This is manifested through the attention given to package documentation. Each package is accompanied by a website consolidating all the documentation, which is accessible from the portal https://packages.tesselle.org. The enhancement of documentation represents one of the most significant ongoing endeavors: providing novice users with sufficient resources to facilitate their initial use of the tools.

The *tesselle* project also aims to adhere to the recommendations of the *tinyverse* (https://www.tinyverse.org) by trying to minimize external hard dependencies to the bare essentials. This simplifies maintenance by avoiding external changes that might impact or break the project. Keeping the project as lightweight as possible also serves to minimize the impact on the end user, ensuring that the installation of one package does not entail installing dozens of others. Although not all packages in the *tesselle* project are entirely dependency-free (Figure 2), the dependencies, with a few exceptions, are internal to the project (the arkhe package, for example, was initially designed for internal use by other packages within the project).

https://doi.org/10.1017/aap.2024.10 Published online by Cambridge University Press

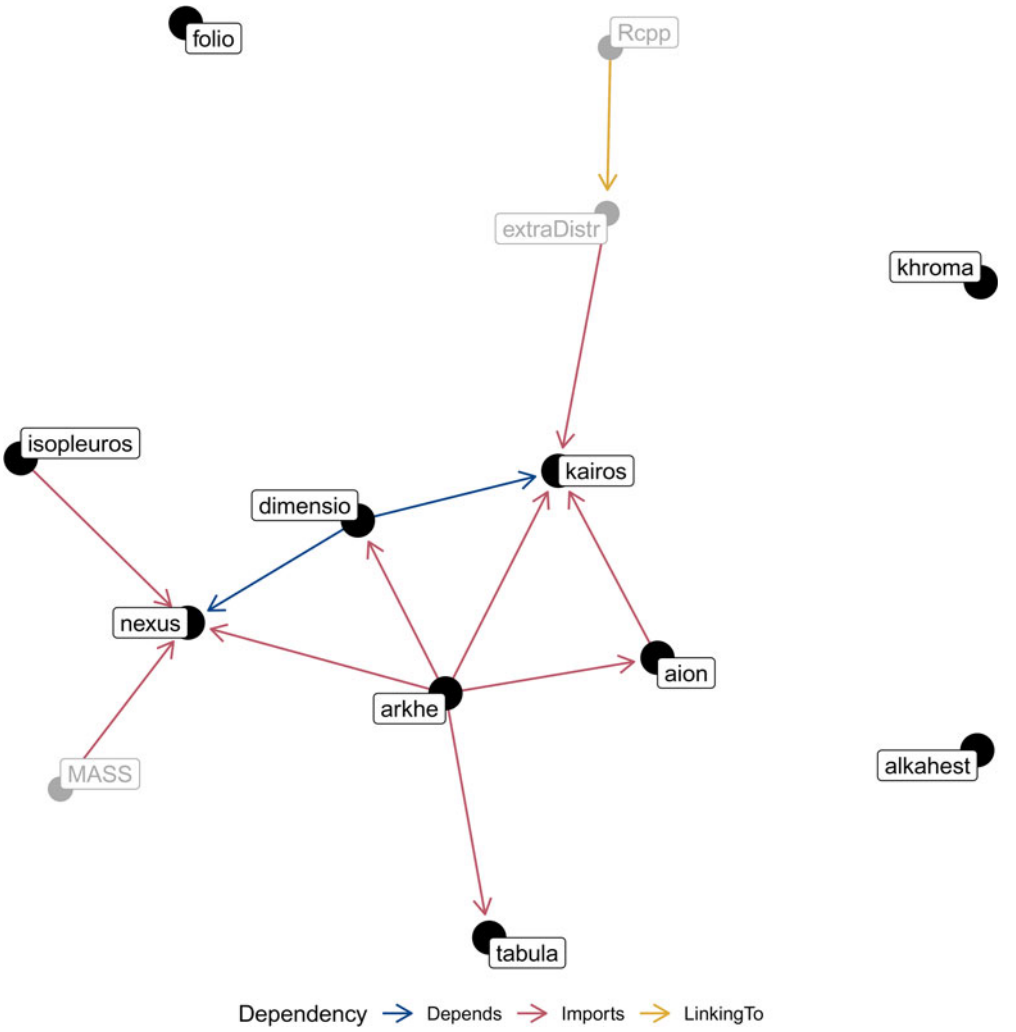

Figure 2. Dependency network of the *tesselle* packages (black dots) as of February 2024. For easier reading, the *tesselle* meta-package is not shown. Data collected with miniCRAN (de Vries 2022) and processed with tidygraph (Pedersen 2023), ggraph (Pedersen 2024), and ggplot2 (Wickham 2016).

The project is developed with transparency and reliability in mind, as indicated by the following:

- All packages are distributed under GNU General Public License (https://www.gnu.org/licenses/gpl-3.0.html): this makes it possible to freely run, copy, distribute, study, change, and improve them.
- All packages are publicly maintained, with source code accessible and versioned on GitHub (https://github.com/tesselle/).
- All packages undergo rigorous testing and code coverage. Most of them are distributed on CRAN, which implies adherence to stringent standards (Chambers 2020).

However, some reservations must be addressed regarding the implementation of these guiding principles. Like many open-source software, the *tesselle* packages come without any warranty. As highlighted by Kreutzer et alia (2017), software quality assurance is a shared responsibility between developers and users. Even with adherence to rigorous development practices (testing, cross-validation, code coverage, etc.), incorrect or unexpected results may arise (flaws in design, corner cases, etc.), or breaking changes may be introduced.

https://doi.org/10.1017/aap.2024.10 Published online by Cambridge University Press

End users must accurately report and cite any software used, along with its version number, to ensure transparency and reproducibility of published results. By doing so, the published results are associated with a specific state of the software, ensuring traceability in case a software error is discovered later. Within the *tesselle* project, semantic versioning (https://semver.org) is employed to assign version numbers. Semantic versioning is a versioning scheme used to convey meaningful information: it supports compatibility and stability, because it distinguishes between major changes that may require adjustments in existing code and minor changes that can be safely integrated without major disruptions. Furthermore, every version of each package is archived on Zenodo (https://zenodo.org) and receives a DOI to be easily citable.

Components

A meta-package, called *tesselle*, lets one download and install the project's core packages with a single R command:

```
install.packages("tesselle")
```

Using the library() function, one can then attach the core *tesselle* packages:

```
library("tesselle")
```

The following core packages are designed to work seamlessly together and can be used to explore and analyze common data types in archaeology:

- tabula (Frerebeau 2023a; https://packages.tesselle.org/tabula/) allows for the examination of archaeological **count data**. It provides several tests and measures of diversity: heterogeneity and evenness, richness and rarefaction, turnover, and similarity. This package makes it easy to visualize count data and statistical thresholds—rank versus abundance plots, heatmaps, and Ford and Bertin diagrams.
- kairos (Frerebeau 2024a; https://packages.tesselle.org/kairos/) provides a tool kit for absolute dating and the analysis of **chronological patterns**. This package includes functions for chronological modeling and dating of archaeological assemblages from count data. It provides methods for matrix seriation and allows for the computation of time point estimates and density estimates of the occupation and duration of an archaeological site. This package relies on aion (Frerebeau and Roe 2023; https://packages.tesselle.org/aion/), which makes it easier to work with **time series** in archaeology.
- nexus (Frerebeau and Philippe 2024; https://packages.tesselle.org/nexus/) allows for the exploration and analysis of **compositional data**. It provides tools for chemical fingerprinting and source tracking of ancient materials by chemical composition.
- dimensio (Frerebeau 2024b; https://packages.tesselle.org/dimensio/) offers methods to compute, extract, summarize, and visualize results of simple **multivariate data** analysis (Principal Components Analysis [PCA] and Correspondence Analysis [CA]).
- isopleuros (Frerebeau 2024c; https://packages.tesselle.org/isopleuros/) enables the creation of **ternary plots** and includes common ternary diagrams useful for archaeologists (e.g., soil texture charts, ceramic phase diagrams).

Additionally, companion packages complement these core packages for specific tasks, such as data visualization or preparation, and can be installed separately. khroma (Frerebeau 2024d; https://packages.tesselle.org/khroma/) provides accessible color schemes tailored for each type of data (qualitative, diverging, or sequential). alkahest (Frerebeau 2023b; https://packages.tesselle.org/alkahest/) is a toolbox for preprocessing XY data from experimental methods (i.e., any signal that can be measured along a continuous variable): it provides methods for baseline estimation and correction, smoothing, normalization, and more. For **teaching** purposes, folio (Frerebeau 2024e; https://packages.tesselle.org/folio/) offers several datasets related to broad topics in archaeology and paleontology, which can be used to illustrate statistical methods in the classroom.

Concluding Words

The *tesselle* project has reached a stable state and is actively being developed. This collection of R packages aims to contribute to the development of open-source computational archaeology. It provides a

https://doi.org/10.1017/aap.2024.10 Published online by Cambridge University Press

consistent and reproducible tool kit that can be easily extended. Users are invited to contribute, share feedback, request new features, or report bugs on GitHub: https://github.com/tesselle/.

Further reading—including examples, tutorials, and detailed documentation—can be found at http://www.tesselle.org.

Acknowledgments. The following contributors have made it possible to develop this project by providing helpful discussion and bringing in new ideas: Jean-Baptiste Fourvel, Ben Marwick, Anne Philippe, and Joe Roe. The author would like to thank Brice Lebrun for creating the project and package logos. The development and maintenance of packages within the *tesselle* project are greatly facilitated by the following packages: usethis (Wickham et al. 2024), devtools (Wickham, Hester, et al. 2022), pkgdown (Wickham, Hesselberth, and Salmon 2022), tinytest (van der Loo 2020), tinysnapshot (Arel-Bundock 2024), codemetar (Boettiger and Salmon 2022), and cffr (Hernangómez 2021). The project also benefits from these infrastructures built and maintained by the R community for package distribution: the Comprehensive R Archive Network (https://cran.r-project.org) and R-universe (Ooms 2021; https://r-universe.dev).

Funding Statement. This research received no specific grant funding from any funding agency or from commercial or not-for-profit sectors.

Data Availability Statement. No original data have been presented in this article. The source code for all R packages is available on GitHub (https://github.com/tesselle/) and archived on Zenodo (see references cited).

Competing Interests. The author declares none.

Note

1. For instance, in France in 2023, only one-third of undergraduate programs in archaeology offered instruction in applied statistics, according to institutional websites.

References Cited

Arel-Bundock, Vincent. 2024. tinysnapshot: Snapshots for Unit Tests Using the "Tinytest" Framework. R package version 0.0.4. https://CRAN.R-project.org/package=tinysnapshot.

Baker, Monya. 2016. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* 533(7604):452–454. https://doi.org/10.1038/533452a.

Baptiste, Zack, and Joe Roe. 2021. Open Archaeology: A Survey of Collaborative Software Engineering in Archaeological Research. Paper presented at the 2021 Computer Applications & Quantitative Methods in Archaeology, virtual meeting.

Boettiger, Carl, and Maëlle Salmon. 2022. Codemetar: Generate "CodeMeta" Metadata for R Packages. R package version 0.3.5. https://CRAN.R-project.org/package=codemetar.

Carlson, David L. 2017. *Quantitative Methods in Archaeology Using R*. Cambridge Manuals in Archaeology. Cambridge University Press, Cambridge.

Chambers, John, M. 2020. S, R, and Data Science. *R Journal* 12(1):462. https://doi.org/10.32614/RJ-2020-028.

de Vries, Andrie. 2022. miniCRAN: Create a Mini Version of CRAN Containing Only Selected Packages. R package version 0.2.16. https://CRAN.R-project.org/package=miniCRAN.

Frerebeau, Nicolas. 2023a. tabula: Analysis and Visualization of Archaeological Count Data. R package version 3.0.1. https://zenodo.org/doi/10.5281/zenodo.1489944.

Frerebeau, Nicolas. 2023b. alkahest: Pre-Processing XY Data from Experimental Methods. R package version 1.1.1. https://doi.org/10.5281/zenodo.7081524.

Frerebeau, Nicolas. 2024a. kairos: Analysis of Chronological Patterns from Archaeological Count Data. R package version 2.0.2. https://doi.org/10.5281/zenodo.5653896.

Frerebeau, Nicolas. 2024b. dimensio: Multivariate Data Analysis. R package version 0.5.0. https://doi.org/10.5281/zenodo.4478530.

Frerebeau, Nicolas. 2024c. isopleuros: Ternary Plots. R package version 1.0.0. https://doi.org/10.5281/zenodo.7940389.

Frerebeau, Nicolas. 2024d. khroma: Colour Schemes for Scientific Data Visualization. R package version 1.11.0. https://doi.org/10.5281/zenodo.1472077.

Frerebeau, Nicolas. 2024e. folio: Datasets for Teaching Archaeology and Paleontology. R package version 1.3.0. https://doi.org/10.5281/zenodo.4476182.

Frerebeau, Nicolas, and Anne Philippe. 2024. nexus: Sourcing Archaeological Materials by Chemical Composition. R package version 0.2.0. https://doi.org/10.5281/zenodo.10225630.

Frerebeau, Nicolas, and Joe Roe. 2023. aion: Archaeological Time Series. R package version 1.0.2. https://doi.org/10.5281/zenodo.8032278.

Hernangómez, Diego. 2021. cffr: Generate Citation File Format Metadata for R Packages. *Journal of Open Source Software* 6(67):3900. https://doi.org/10.21105/joss.03900.

Hostler, Thomas J. 2023. The Invisible Workload of Open Research. *Journal of Trial and Error* 4(1). https://doi.org/10.36850/mr5.

Ioannidis, John P. A. 2005. Why Most Published Research Findings Are False. *PLoS Medicine* 2(8):e124. https://doi.org/10.1371/journal.pmed.0020124.

Karoune, Emma, and Esther Plomp. 2022. Removing Barriers to Reproducible Research in Archaeology. Zenodo, version 5, peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.7320029.

Kreutzer, Sebastian, Christoph Burow, Michael Dietze, Margret C. Fuchs, Manfred Fischer, and Christoph Schmidt. 2017. Software in the Context of Luminescence Dating: Status, Concepts and Suggestions Exemplified by the R Package "Luminescence." *Ancient TL* 35(2):1–11.

Marwick, Ben. 2017. Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 24(2):424–450. https://doi.org/10.1007/s10816-015-9272-9.

Marwick, Ben, Li-Ying Wang, Domenico Giusti, Enrico R. Crema, Tal Galili, Bjørn Peare Bartholdy, Laure Spake, et al. 2022. CRAN Task View: Archaeological Science. GitHub. https://github.com/benmarwick/ctv-archaeology, accessed October 19, 2023.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. A Manifesto for Reproducible Science. *Nature Human Behaviour* 1(1):0021. https://doi.org/10.1038/s41562-016-0021.

Ooms, Jeroen. 2021. A First Look at the R-Universe Build Infrastructure. Electronic document, https://ropensci.org/blog/2021/03/04/r-universe-buildsystem, accessed October 19, 2023.

Pedersen, Thomas Lin. 2023. tidygraph: A Tidy API for Graph Manipulation. R package version 1.3.0. https://CRAN.R-project.org/package=tidygraph.

Pedersen, Thomas Lin. 2024. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 2.2.0. https://CRAN.R-project.org/package=ggraph.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Electronic document, https://www.R-project.org, accessed October 23, 2024.

Schmidt, Sophie C., and Ben Marwick. 2020. Tool-Driven Revolutions in Archaeological Science. *Journal of Computer Applications in Archaeology* 3(1):18–32. https://doi.org/10.5334/jcaa.29.

Tufféry, Christophe. 2019. Les compétences numériques en archéologie: Un défi majeur et des risques de déni. *¿Interrogations?* 28. https://www.revue-interrogations.org/Les-competences-numeriques-en, accessed October 19, 2023.

van der Loo, Mark P. J. 2020. A Method for Deriving Information from Running R Code. *R Journal* 13(1):42–52. https://doi.org/10.32614/RJ-2021-056.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wickham, Hadley, Jay Hesselberth, and Maëlle Salmon. 2022. pkgdown: Make Static HTML Documentation for a Package. R package version 2.0.7. https://CRAN.R-project.org/package=pkgdown.

Wickham, Hadley, Jim Hester, Winston Chang, and Jennifer Bryan. 2022. devtools: Tools to Make Developing R Packages Easier. R package version 2.4.5. https://CRAN.R-project.org/package=devtools.

Wickham, Hadley, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2024. usethis: Automate Package and Project Setup. R package version 2.2.3. https://CRAN.R-project.org/package=usethis.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43):1686. https://doi.org/10.21105/joss.01686.

Cite this article: Frerebeau, Nicolas. 2024. The *tesselle* Project: A Collection of R Packages for Research and Teaching in Archaeology. *Advances in Archaeological Practice*. https://doi.org/10.1017/aap.2024.10.