The Journal of Open Source Software

tabula: An R Package for Analysis, Seriation, and Visualization of Archaeological Count Data

Nicolas Frerebeau[1, 2]

1 IRAMAT-CRP2A (UMR 5060 - CNRS, Université Bordeaux Montaigne) **2** Laboratoire de Mathématiques Jean Leray (UMR 6629 - CNRS, Université de Nantes)

DOI: 10.21105/joss.01821

Software

- Review
- Repository
- Archive

Editor: Olivia Guest

Reviewers:

- @izaromanowska
- @soodoku

Submitted: 20 September 2019
Published: 25 November 2019

Background

Detecting and quantifying material and cultural variations in time and space are important methodological issues in archaeology. To solve these issues, we need to construct reliable chronologies and quantitative descriptions of archaeological assemblages, i. e. archaeological sites or intrasite units, each described as a set of p different objects.

Building chronologies involves distinguishing between relative (providing only a chronological sequence) and absolute dating methods (that yield calendric indicators) (O'Brien & Lyman, 2002). Within relative dating, matrix seriation is a long-established method—it was first formulated by Petrie (1899)—and has allowed the construction of reference chronologies (Ihm, 2005). For a set X of n archaeological assemblages, the seriation problem comes down to discovering in X an order inferred as chronological. This approach relies on a set of well-defined statistical and archaeological assumptions (Dunnell, 1970). It may use *a priori* information, e.g., absolute dates or stratigraphical constraints: Poblome & Groenen (2003), and allows the analysis of chronological patterns in a socio-economic or cultural perspective, e.g., Bellanger & Husi (2012), Lipo, Madsen, & Dunnell (2015).

The quantitative analysis of archaeological assemblages can thus be carried out in a synchronic, e.g., diversity measurements, or diachronic, e.g., evolutionary studies: selection process, patterns of cultural transmission, etc., way. These approaches cover a wide range of applications and have led to the development of a multitude of statistical models, but none have been systematically implemented to enable the deployment of reproducible workflows.

Summary

`tabula` provides a convenient and reproducible toolkit for analyzing, seriating, and visualizing archaeological count data, e.g., artifacts, faunal remains, etc.

Several R packages, e.g. `ade4` (Dray & Dufour, 2007), `SpadeR` (Chao, Ma, Hsieh, & Chiu, 2016) or `vegan` (Oksanen et al., 2019), allow the estimation of diversity indices and implements seriation/ordination methods, but these packages are mainly oriented towards ecological issues. `tabula` provides archaeologically-orientated implementations that allow the integration of specific data (dates, stratigraphy, etc.) and offers a consistent framework. The latter is of particular value since `tabula` is designed to be used both by archaeologists and by students with little background in courses on dating methods and applied statistics in archaeology.

The package uses a set of S4 classes for archaeological data matrices that extend the `matrix` data type. These new classes represent different specialized matrices: incidence, abundance, co-occurrence, and (dis)similarity. Methods for a variety of functions applied to objects from these classes provide tools for relative and absolute dating and analysis of (chronological) patterns.

`tabula` includes functions for matrix seriation (`seriate_*`), as well as chronological modeling and dating (`date_*`) of archaeological assemblages and objects. Resulting models can be checked for stability and refined with resampling methods (`refine_*`). Estimated dates can then be displayed as tempo or activity plot (Dye, 2016) to assess rhythms over long periods. Beyond these, `tabula` provides several tests (`test_*`) and measures of diversity within and between archaeological assemblages (`index_*`): heterogeneity and evenness (Brillouin, Shannon, Simpson, etc.), richness and rarefaction (Chao1, Chao2, ACE, ICE, etc.), turnover and similarity (Brainerd-Robinson, etc.). Finally, the package makes it easy to visualize count data and statistical thresholds (`plot_*`): rank vs. abundance plots, heatmaps, Ford (1962), and Bertin (1977) diagrams.

Acknowledgements

All the contributors have made it possible to develop this project through their helpful discussion and by bringing in new ideas: Jean-Baptiste Fourvel, Brice Lebrun, Ben Marwick, Matthew Peeples, and Anne Philippe.

References

Bellanger, L., & Husi, P. (2012). Statistical tool for dating and interpreting archaeological contexts using pottery. *Journal of Archaeological Science*, *39*(4), 777–790. doi:10.1016/j.jas.2011.06.031

Bertin, J. (1977). *La graphique et le traitement graphique de l'information*. Nouvelle Bibliothèque Scientifique. Paris: Flammarion.

Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C.-H. (2016). *SpadeR: Species-richness prediction and diversity estimation with r*. Retrieved from https://CRAN.R-project.org/package=SpadeR

Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. doi:10.18637/jss.v022.i04

Dunnell, R. C. (1970). Seriation Method and Its Evaluation. *American Antiquity*, *35*(03), 305–319. doi:10.2307/278341

Dye, T. S. (2016). Long-term rhythms in the development of Hawaiian social stratification. *Journal of Archaeological Science*, *71*, 1–9. doi:10.1016/j.jas.2016.05.006

Ford, J. A. (1962). *A quantitative method for deriving cultural chronology*. Technical manual. Washington, DC: Pan American Union.

Ihm, P. (2005). A Contribution to the History of Seriation in Archaeology. In C. Weihs & W. Gaul (Eds.), *Classification the Ubiquitous Challenge* (pp. 307–316). Presented at the 28th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Dortmund, march 9-11, 2004, Berlin Heidelberg: Springer. doi:10.1007/3-540-28084-7_34

Lipo, C. P., Madsen, M. E., & Dunnell, R. C. (2015). A Theoretically-Sufficient and Computationally-Practical Technique for Deterministic Frequency Seriation. (N. Bicho, Ed.)*PLOS ONE*, *10*(4), e0124942. doi:10.1371/journal.pone.0124942

O'Brien, M. J., & Lyman, R. L. (2002). *Seriation, Stratigraphy, and Index Fossils: The Backbone of Archaeological Dating*. Dordrecht: Springer.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., et al. (2019). *Vegan: Community ecology package*. Retrieved from https://CRAN.R-project.org/package=vegan

Petrie, W. M. F. (1899). Sequences in Prehistoric Remains. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *29*(3/4), 295–301. doi:10.2307/2843012

Poblome, J., & Groenen, P. J. F. (2003). Constrained Correspondence Analysis for Seriation of Sagalassos Tablewares. In M. Doerr & A. Sarris (Eds.), *The Digital Heritage of Archaeology*. Presented at the CAA2002 (Heraklion, Crete; April 2002), Athens: Hellenic Ministry of Culture.