Paper submission for CAA2023, Amsterdam, S12: "Chronological modelling: formal methods and research software"

Tidying up time: towards an infrastructure for chronological modelling in R

Joe Roe
Institute for Archaeological Sciences, University of Bern
https://joeroe.io

Abstract

The statistical programming language *R* is widely used to conduct a range of analyses in computational archaeology. Time is indeed at the heart of many of these analyses. There are a number of R packages for chronological modelling, including applications developed specifically by or for archaeologists as well as those borrowed from adjacent fields, and many more types of analysis incorporate chronological data in an *ad hoc* way. As the range and complexity of these analytic procedures increases, it can become increasingly difficult to integrate the varied data types, interfaces, and output formats into a single workflow – to get different packages to "talk to each other". This is a common problem, especially in decentralised software ecosystems such as R. In other domains, the solution has often common in the form of computational 'infrastructure': packages that smooth the way by providing consistent formal representations, a common set of basic methods, and a predictable user interface patterns.

In this paper, I introduce three interconnected R packages:

- era provides formal representations of year-based time scales of the type usually used by archaeologists and other palaeoscientists (e.g. Before Present).
- c14 facilitates 'tidy' analysis of radiocarbon data, including further formal classes for uncalibrated measurements and calibrated, calendar probability distributions, as well as functions for converting data types from existing radiocarbon-analysis packages into tidy formats.
- stratigraphr is a framework for working with stratigraphies and relative chronologies as directed graphs, with tools for reading, analysing, and visualing these stratigraphic graphs ('Harris matrices').

Apart from implementing various analytical methods, the aim of these packages is to provide a common infrastructure for the formal representation of chronological data and a set of consistent tools for reasoning with it. All three make extensive use of tooling and design patterns from the 'tidyverse' (Wickham et al. 2019) to present a predictable user interface that fits into a wider ecosystem of R packages. In particular, they use vctrs-based formalisms

to represent different types of chronological vectors, a functional paradigm for composing programs, and assume/encourage the use of a tidy data workflow (Wickham 2014).

A common computational infrastructure for chronological data makes it easier to 'glue together' varied packages for chronological modelling within a single analysis. And by linking it to an existing, widely-used paradigm (the tidyverse) we can take this further, smoothing the way for the integration of chronological data into a wide range of other methodologies. The packages outlined here are not intended to be the definitive version of such an infrastructure but, I suggest, could serve as the rudiments of a more coherent approach of chronological modelling in R.

References

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10): 1–23. https://doi.org/10.18637/jss.v059.i10.

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.