

# Digital Ecosystems in Archaeological Science: A History and Taxonomy of R packages in Archaeology

Ben Marwick, University of Washington, bmarwick@uw.edu

## Introduction

The digital ecosystem of a research community is important to understand because it reveals what is possible, what is common, and what unrealised opportunities exist for future work. The recent growth in the use of R in archaeological science is of especial interest because it is a free, open source, and highly extensible programming language that any researcher can contribute to by writing a package and sharing it for others to use. In this presentation we present the results of a study of how archaeologists use R packages, and how archaeologists write them for other researchers to use. We ask: what are the patterns in R package use and production among archaeologists, and what packages are yet to be developed that might satisfy the specific needs of archaeological scientists?

## Methods and materials

We have two data sources for this study: (1) a list of packages written by archaeologists, and (2) a list of 150+ scholarly publications that include R code. Both datasets are openly curated at <https://github.com/benmarwick/ctv-archaeology>. To answer our question about how packages are used, we will extract all the packages referenced in the 150+ papers that include R code. We will classify these packages according to keywords in their title to identify the types of packages that are often used by archaeologists, and how they associate with journals and archaeological topics. We will study co-citation patterns to understand what packages tend to be used together, and for what types of data. To answer our question about archaeologists produce packages, we will use our dataset of packages authored by archaeologists and analyse data on dates of first publication, usage data from CRAN downloads and citations, and co-citation patterns. We will also examine software engineering attributes, such as the presence/absence of vignettes, tests, continuous integration, dependencies, etc. to understand how archaeologists adhere to modern package development conventions ,

## Results

The analysis is underway and results are not available yet.

## Discussion

Our results will be the first empirical study to examine the frequency of R package use for the reported analyses of archaeological data in publications. We will also report on patterns of R package production by archaeologists. We

expect that our results will be useful for archaeologists looking to begin using R but feel overwhelmed by the 10,000+ packages currently on CRAN. Our results will show the most commonly used packages, which will guide archaeologists towards packages that will probably be a good choice for their data analysis needs, assuming a strong frequency-based bias in software choice. In looking at how archaeologists produce R packages, we will reveal where current effort has been concentrated, and recommend where future effort be directed. We will also report on the best practices currently in use by archaeologists writing R packages, to inspire and motivate future package authors to produce high quality software that will be used and cited by the research community

## References

- Joo, Rocio, Matthew E. Boone, Thomas A. Clay, Samantha C. Patrick, Susana Clusella-Trullas, and Mathieu Basille. "Navigating through the r packages for movement." *Journal of Animal Ecology* 89, no. 1 (2020): 248-267.
- Lai, Jiangshan, Christopher J. Lortie, Robert A. Muenchen, Jian Yang, and Keping Ma. "Evaluating the popularity of R in ecology." *Ecosphere* 10, no. 1 (2019): e02567.
- Li, Kai, and Erjia Yan. "Co-mention network of R packages: Scientific impact and clustering structure." *Journal of Informetrics* 12, no. 1 (2018): 87-100.
- Lortie, Christopher J., Jenna Braun, Alessandro Filazzola, and Florencia Miguel. "A checklist for choosing between R packages in ecology and evolution." *Ecology and evolution* 10, no. 3 (2020): 1098-1105.