How to use replication assignments for teaching integrity in empirical archaeology

Ben Marwick
Liying Wang
Ryan Robinson
Hope Loiselle

The value of new archaeological knowledge is strongly determined by how credible it is, and a key measure of scientific credibility is how replicable new results are. However, few archaeologists learn the skills necessary to conduct replication as part of their training. This means there is a gap between the ideals of archaeological science and the skills we teach future researchers. Here, we argue for replications as a core type of class assignment in archaeology courses to close this gap and establish a culture of replication and reproducibility. We review replication assignments in other fields, and describe how to implement a replication assignment suitable for many types of archaeology programs. We describe our experience with replication in an upper-level undergraduate class on stone artefact analysis. Replication assignments can help archaeology programs give students skills that enable transparent and reproducible research.

Introduction

In his influential study of replication, sociologist of science Harry Collins argued that replication is at the core of scientific practice, writing “Replicability […] is the Supreme Court of the scientific system” (Collins 1992, 19). Like other observers of science, Collins claimed that in replication, private observations become communal facts, offering vital protection from error and fraud. In this paper we propose a new type of assignment for the archaeology classroom, the replication report, to better align the practice of teaching archaeology with the transparency and openness ideals of science (Nosek et al. 2015). The replication report assignment involves four steps: 1) students analysing a published report to determine the main claims made by the authors of that report, 2) students obtaining the data used by the authors, 3) analysing that data to help determine if one or more of the authors’ claims are reliable, and 4) submitting a research compendium that documents the students’ work in a reproducible format including the code and data used in the assignment.

We describe how to implement a replication report assignment suitable for upper-level undergraduates and graduate students in archaeology. Our experience is based on an upper-level archaeology class on stone artefact analysis taught during the Spring quarter of 2019 at the University of Washington. The class format includes a weekly cycle of lecture, discussion seminar, and hands-on laboratory activities. The assignments include seminar notes, lecture quizzes, laboratory worksheets, and two longer empirical reports. For the term that we report here, the class had 16 students and one graduate student teaching assistant. This is a typical size for this class, and similar to the usual size of upper-level laboratory classes in the archaeology program at the University of Washington. Our students are mostly in social science and humanities majors with varying levels of statistical competence. Here, we survey the literature on similar types of assignments in other fields to identify common elements that other fields have identified as important principles and skills. We describe our assignment and discuss student feedback on our implementation. Finally, we offer recommendations for how to use replication reports to teach archaeology students.

To enable re-use of our materials and improve reproducibility and transparency according to the principles outlined in Marwick et al. (2017), we include all our assignment materials, as well as the entire R code used for all the analysis and visualizations contained in this paper, in our compendium at http://doi.org/10.17605/OSF.IO/DBSW9. Also in this version-controlled compendium are the raw data for the analyses reported here. The figures and results presented here can be independently reproduced with the code and data in this repository. In our compendium, our code is released under the MIT license, our data as CC-0, and our figures, and assignment instructions and grading rubric as CC-BY, to enable maximum re-use (for more details, see Marwick 2017).

What is replication?

Barba (2018) points out that although there has been prolific discussion of the terms “reproducibility” and “replication” in many disciplines in recent years, confusion and conflicting uses are widespread. In her survey of relevant literature, Barba finds that some fields make no distinction between “reproducibility” and “replication.” Among fields that do recognise a distinction between the two, the meanings are sometimes directly inverted. Here we follow what Barba has identified as the most common, long-established and highly cited definitions of these terms, as also recently recommended by the National Academies of Sciences, Engineering, and Medicine (2019). Reproducibility is the ability to obtain results by using the same data, code, and procedures provided by the original authors (Marwick 2017). This is only possible when the authors make all those materials available, for example in a research compendium (Marwick, Boettiger, and Mullen 2018). Replication is the ability to arrive at the same scientific conclusions in a new study, collecting new data (possibly with different methods) and completing new analyses.

In the following section, we briefly survey replication assignments described in other fields to survey the variety of forms this usually takes. Replication assignments across different fields may not strictly fit into the above definition of replication because they do not always involve a completely new study. Nevertheless, we consider that if a study departs from any of the original materials, e.g. new data with previously published code, or new code with previously published data, then it is broadly within the definition of replication.

How do different disciplines use replication assignments?

Some of the earliest discussions of replication in university curricula appear in economics and psychology (Höffler 2013; Standing et al. 2014). Ball and Medeiros (2012) describe their TIER (Teaching Integrity in Empirical Research) protocol for undergraduate economics students at Haverford College. This is intended to ensure that a student’s work is replicable by the instructors. Their protocol requires that when students submit their final project report, their submission must contain four elements: the raw data files, a metadata file, script files of code used to analyse the data, and public availability (i.e. deposit on an open repository such as Dataverse). Frank and Saxe (2012) describe how they teach undergraduates (at MIT) and graduate students (at Stanford) to do in-class replications of recent, cutting-edge psychology experiments, and note that several projects from their undergraduate course have even been part of successful publications. More recently Hawkins et al. (2018) reported on 11 replication assignments from a psychology graduate seminar at Stanford, finding that the replications typically yielded effects that were smaller than the originally published ones. Similarly, Jern (2018) describes how students in a psychology course completed replication assignments by using statistical methods of the original research articles with new data collected by the students outside of class.

Students in Stanford University's graduate course 'Advanced Topics in Networking' are given a replication assignment in which they are asked to replicate 'classic' computer networking experiments (Yan and McKeown 2017). Students work in pairs and receive modest instructor support. The assignment entails selecting appropriate emulation software, communicating with the original authors, obtaining the authors' materials, replicating the experiment, and publicising their results through both an in-class presentation and a blog post on a program website. We classify this as a replication assignment because many students could not obtain the original code from the authors, and had to write their own for the assignment. Since 2012, over 200 undergraduate and graduate students have participated in this assignment with an 86% success rate (Yan and McKeown 2017). Student feedback suggests high satisfaction with the assignment, citing unique educational value, improved understanding of the original material, and the acquisition of professional skills. In some cases, students personally contributed to the network engineering literature when their replications exposed inaccuracies in original experiments, which were then presented to and publicly amended by the authors (Yan and McKeown 2017).

In describing her political science classes at the University of Cambridge, Janz (2016) argues that reproducibility and replication should be held as the gold standard for scientific research. She claims that teaching these concepts should be a necessary component of graduate studies, to ensure students can make their own future work reproducible. Janz reports on her class in which about 15 students undertake replication assignments over eight weeks, including providing weekly updates to each other to gain insight and feedback. Janz describes two possible levels of assignment suitable for different lengths of the term and levels of the students: duplication (aiming for the exact same results based on the exact same data set with exactly the same methods) and replication (tests the robustness of previous research results by employing newly collected data, and/or new variables, and/or new model specifications). Duplication, which we would define here as reproduction, may be beneficial for lower-level students, while upper-level students can replicate a study and contribute original data, potentially leading to publication. Janz (2016) describes how replication assignments are a growing trend in political science departments (noting that R and STATA are commonly used), and reviews many of the practical challenges of doing replication assignments in a graduate course. She also responds to six typical criticisms of replication assignments, and points out the need for universities to nurture a culture of reproducibility and replication to ensure that the gold standard of reliable, credible, and valid research is not just an empty phrase.

The Freie Universität Berlin extracurricular graduate seminar course ‘Digital Open Science’ aims to teach open science practice and assigns replication projects, mostly around neuroimaging topics (Toelch and Ostwald 2018). These projects are carried out with a variety of typical open science software tools and services, including Python, R, Git, GitHub and the Open Science Framework. Students first receive extensive lectures and hands-on tutoring, then choose a simple neuroscience experiment to replicate. Finally, they present their results at a symposium. The course’s primary goal is to teach students the value of verifying data upon which their own future research might rely. Students report a high rate of engaging in open science practices after taking the class, and 80% of the participants said that they believe the open science techniques will improve their future research as professionals (Toelch and Ostwald 2018). Millman et al. (2018) describe a similar course at Berkeley that teaches students how to use open science tools to complete a capstone replication assignment on neuroscience topics.

This brief survey demonstrates that replication assignments are widely known in economics, psychology, political science, neuroscience and other fields (e.g. Roettger and Baer-Henney 2018) . Common elements include group work, use of open source software and services to make the replication results openly accessible to anyone, and a scaffolded, stepwise approach to the task to ensure that students receive instructor support at multiple stages in the assignment. To the best of our knowledge, replication assignments are not common in archaeology programs, although the tools and data structures are generally similar to other social sciences. We posted a message to the Society of American Archaeology Teaching Archaeology Interest Group e-community on 27 May 2019 to ask for examples of replication assignments used in teaching archaeology, and received no replies from anyone teaching with replication. More broadly in archaeology, replication and reproducibility has received limited, but growing, attention. Elsewhere, we have documented recent rapid increases in the number of publications that include code and data to enable readers to reproduce the published results (Marwick and Schmidt 2019)

How to conduct a replication assignment in archaeology

In this section we describe our replication assignment and how we assessed its effectiveness. A brief discussion of our replication report assignment was announced at the beginning of the ten week-long term to give students background about the purpose and concepts of replication and our expectations. Our replication report assignment consisted of three small, graded activities to scaffold the preparation of the final report. The first step started from Week Four and each step was separated by one week to give students time to work and submit their final reports due on Week Seven. Students were expected to work in groups of 3-4 people, but submit their assignments for each of the three steps and the final report individually. Submissions for each step were graded as complete/incomplete, with feedback provided individually via the Canvas learning management system, and collectively during class meetings. Our course had no prerequisites, so we assumed no prior knowledge of the free and open source R programming language among the students, and were prepared to teach them as complete novices. We chose R (R Core Team 2019) because it is widely used by archaeologists (Schmidt and Marwick 2019), and also commonly taught in undergraduate classes in social sciences and statistics (Baumer et al. 2014; Cetinkaya-Rundel and Rundel 2018; Dvorak et al. 2019). We were also prepared for students to have no prior experience with replication assignments.

Step 1: In groups, select a study to replicate

For the first step, we supplied students with a list of journal articles that included raw data and R code either in supplemental files or deposited on open data repositories. This list is updated regularly, but is not exhaustive, and is currently online at https://github.com/benmarwick/ctv-archaeology. Working in groups of 3-4 people, students selected a journal article from this list as their target article for this assignment. We encouraged them to choose a target article about a stone artefact analysis that looked interesting to them. We also required students to set up an open communication channel for their group to ensure they had an easy way to discuss their selection of the target article. We used Slack (https://slack.com/), a free cloud-based web application for team communication (Perkel 2017), to help them collaborate with each other efficiently. The instructors were members of all the student group channels to supervise, give guidance and support good communication habits. Students were required to individually submit the full bibliographic reference for their target article to complete step one.

Step 2: Identify the key claims and data in the study

For step two of the assignment, students were required to discuss in their groups to identify 2-3 key claims made by the authors of their target article. They were told to study the data visualisations in the paper to identify which figures seem to give the best support to the authors' claims. Recreating these 1-2 visualisations was a key task for the students in the production of their final report. A second task for step two was for students to identify and obtain the raw data files of their target article. The list of articles that the students chose from only included articles for which data were openly available. This removed the need for students to contact authors to request data, which may have added the risk of a long wait for a favourable reply, refusal to share, or no reply. To complete step two, each student was required to submit a short statement summarising the 2-3 key claims of their target paper, and the raw data file.

Step 3: Begin the replication analysis and get instructor feedback

Step three of the assignment required students to create a file structure on their computer to organise their assignment files, following basic guidance in Marwick, Boettiger, and Mullen (2018). They also had to download an R Markdown file and write a small amount of R code to read in the raw data and explore it with one basic visualisation, using data in the target article. R Markdown is a file format for making reproducible documents with R. An R Markdown document is written in markdown (a simple plain text format) and contains chunks of embedded R code (Xie, Allaire, and Grolemund 2018). The document can be easily converted into many standard formats, such as Microsoft Word, PDF and HTML, we give more detail about this in Marwick (2017). We prepared an R Markdown template file with some basic headings (following the IMRaD or Introduction-Method-Results-and-

Discussion format) and empty code chunks to provide guidance to the students on how many code chunks were expected, and where in the document they should appear. As students wrote their R code and encountered errors, they were encouraged to share screenshots on Slack for the instructors to assist with troubleshooting. After completing this step, the instructor met with each group to review the main claims identified by the students, review the visualisation they had chosen to replicate, and provide guidance on writing the R code to produce the key visualisations.

Step 4: Complete the replication analysis and submit the compendium of report, code and data

The final task was for the students to write their report, and submit a reproducible research compendium. This included three files: (1) their R Markdown document, (2) the raw data file, and (3) the output document (e.g. Microsoft Word document that is produced when they knit the R Markdown file). The students submitted these materials to Canvas for grading. Two complete student submissions are available for inspection in our compendium at http://doi.org/10.17605/OSF.IO/DBSW9. We did not make all the student work public, unlike some of examples described above that deposit student work in public on the Open Science Framework. Our expectation was that we could reproduce any student's results by running their submitted R Markdown document with the raw data file to produce the Word document they submitted. The final report was graded with a rubric which was presented to the students at the first step to help set expectations about what the final product should look like.

In the time between students submitting their final report and the grades being released we administered an online survey on Canvas to obtain feedback anonymously from the students. The purpose of the feedback survey was to collect information about how to improve the assignment for future classes, to understand the students' experience of the assignment, and what value they perceive in replication skills for archaeology in general, and for them individually. Two questions were designed to learn about students' prior experiences of replication assignment and using the R language. We asked about students' opinions and attitudes toward replication assignments in archaeology and collected responses on a likert scale. Two open-ended questions sought to know more about the students' thoughts on replication in the classroom in general. They had one week to respond to the survey, which was not a requirement.

Observations on the assignment process

The first step, choosing the target article, revealed the need for some intervention from the instructor to guide students to articles that used relatively simple statistical methods. For example, one group initially chose Breslawski et al. (2018) as their target article, but the key claims in this paper depend on multiple comparisons of multilevel regression models. We explained to the students that if they attempted to replicate a key claim of this paper then they would likely be doing substantially more work than other groups in the class. We invited this group to choose a different target article to ensure a more comparable experience, which they accepted. The statistical backgrounds of our students was highly diverse, so we could not expect students to be very discerning about the statistical complexity of the methods in the articles on the list of potential target articles. As a consequence, we were prepared to intervene to guide their selection of a paper that we could be sure they could successfully replicate, given the time available. The target articles used by this class were Marwick et al. (2016), Bicho and Cascalheira (2020), and Marwick (2013).

The second step was mostly straightforward, with students engaging in discussion in class and on Slack to identify the 2-3 key claims of their target paper, and identify the data visualisation that provided the most relevant support to one or more of those claims. Given that not all students have statistical background, the instructor covered some statistical methods they might encounter while reading during lecture, such as principal components analysis, to give them the mathematical concepts behind it. Identifying the data files was less straightforward, with about one third of students failing to correctly identify the data files accompanying their target article. We attribute this to the relatively low level of familiarity of the students in working with raw data, so they are not sure when they are looking at it, and the high degree of variability in how the target article authors make their data open. Some authors include their data as a file in the supplementary information attached to the article, while others deposit their files on an open data repository such as osf.io or figshare.com, and then cite the DOI to the files in their article. When the data files were nested in several layers of folders, some students struggled to find them.

The ability to easily share screenshots on Slack was important to the success of the third step. Our intention was that two lab classes earlier in the term that introduced students to some methods for data visualisation using R would provide the foundation for succeeding in this step. We have expected that two lab reports completed earlier in the term that required to be written with R markdown would help them practice crucial code they might need later. For the lab reports, students used R markdown templates we provided to complete tasks of reading data into R, basic data tidying, and data visualisation by modifying sample code. However, we found that for some students this was not sufficient practice, and substantial instructor guidance was required to help them complete this step. At the completion of this step, the instructor met one on one with each group to check how successful they were producing a basic visualisation using data from the target article, and to discuss the group’s strategy to complete the report. This was the most time-consuming aspect of the assignment for the instructor, holding a one hour meeting with each of the five groups.

Analysis of the students’ anonymous feedback

Thirteen out of sixteen students completed an anonymous feedback survey (Figure 1). Only one student had done replication before and two had used R previously for an archaeology assignment. Most students strongly agreed with the statements about having sufficient support and clear instructions. Most students strongly disagreed with the statement, “I am likely to attempt to replicate published research in my future studies and work”. This contrasts with the high proportion of students that agree with the statement “The ability to replicate published research is an important skill for professional archaeologists”. Taken together, these two responses show that while students see the value of replication to archaeology in general, they do not see any specific benefits to doing it themselves. This may reflect a failure of the instructor to communicate the individual benefits of developing skills for replication. It may also reflect uncertainty among the students about their plans for a career in archaeology. Most, but not all, students agreed that the replication assignment helped them to learn more effectively than reading a paper to write a traditional paper.

Figure 2 shows the correlations between the five feedback questions that have likert scale responses. The statements about instructions and instructor support are highly positively correlated, showing the positive effects of the assignment design, a detailed rubric, and the instructor meeting with each group to discuss their work, and answering students' questions promptly on Slack. The strongest negative correlation is between the statements about instructor support and doing replication in future work. This might suggest that the students received so much support that they do not feel capable to do a replication like this by themselves. We see confirmation of this in the free-form comments, such as 'it would not have been possible for us to do this correctly on our own'. These correlations indicate a need to equip students with skills to work more independently of the instructor, and strengthen students' self-efficacy with replication skills.

Analysis of the students' grades for the replication assignment

We graded the students' final submissions using a rubric with criteria that covered content, the introduction, methods, result, and conclusion sections, and style. In Figure 3 we show the distribution and means of student scores for each criterion. The two criteria showing the highest mean score are "Style: use commas and apostrophes correctly, and spell consistently", and "Intro: has clear statement of the purpose of the report". High scores for the grammar criterion are expected because these reflect basic writing skills required for many undergraduate-level courses. Students are expected to have learned these in lower-level classes before taking this class. The high scores relating to the introduction section may reflect the effectiveness of the scaffolding steps that focused students on the specific purpose of the assignment. The lowest mean score is for "Content: minimum of 4 scholarly items in the reference list" that shows some people did not include four items. This might result from insufficient prior training in searching for scholarly publications, suggesting that although this is also a skill that should have been acquired before taking this class, many students remain weak at this task. A low mean is also evident for "Intro: has names, locations, and basic chronology of sites," because some students neglected to supply these archaeological details. Future use of this assignment will incorporate these low-scoring criteria into the scaffolding steps to emphasize their importance to students and give an opportunity for early feedback to students.

The criteria most relevant to the replication component of the assignment, in order from highest mean score to lowest, are "Content: submission includes Rmd file, Data file, and Word file", "Conclusion: state whether the author's claims appear to be robust, unreliable, etc", "Results: includes 1-2 original plots & description of these", and "Methods: identify the specific results you will replicate". This suggests that we could help students to develop better skills in narrating their process (writing about methods), and describing and interpreting their data visualizations. In the future we may include more fundamental exercises focusing on these tasks in the scaffolding steps. Overall, we find that comparison of the scores for the replication criteria and other criteria shows there is no clear evidence that the replication component of this assignment lowers students' grades. The two lowest scoring criteria are more generic research and writing skills rather than skills specific to the replication aspect of the assignment.

Discussion

Replication of results is widely claimed as a gold standard in science. When a result can be independently validated, we can build on it to advance knowledge in our field. Teaching students about replication and giving them skills to conduct replication is thus a vital part of preparing them for professional work in scientific archaeology. The immediate practical benefits of students doing replication assignments include realistic experience with analysing and visualising real-word data (rather than toy datasets often used for class activities), and getting them working at the research frontier by taking an in-depth look at recently published work, going beyond the usual reading and discussion.

The longer term practical benefits include cultivating a reproducibility routine for students to develop a natural habit of organising their code and data for future work so another person can use it to reproduce their results. Benefits also include developing professionalism among students: by working through the steps of an analysis students gain an understanding of what kinds of decisions in all steps of an analysis are acceptable (Janz 2016). Although the small scale of our assignment did not offer the potential for students to publish new findings from replication, we anticipate this may be a benefit for archaeology students participating in more extensive replication assignments.

The challenges of requiring students to do replication assignments are similar to those faced in many types of archaeology classes with quantitative skills at their core. In our case, the absence of a prerequisite and the high variability of statistical and programming skills meant that some students needed a lot more support than others. This may make replication assignments impractical where instructor-student contact hours are limited. In addition to time, the instructor needs to have a high skill level in quantitative methods to guide students in their engagement with the literature. The instructor will also benefit from a high tolerance for helping students to solve coding problems, and having a teaching assistent with a suitable background and similar qualities. To mitigate this in the future we will add a step of student peer review (Wessa 2009) to distribute the feedback task beyond the instructor and give students an opportunity to get assistance from each other more formally.

Conclusion

Our main finding is that replication assignments are valued and in common use throughout the social sciences, and that they can be effective in teaching archaeology. Specifically, we found it possible to conduct a small-scale replication assignment as part of an upper-level undergraduate archaeology class. Student feedback indicated that it was a valuable new experience for them, and for the discipline, even if they could not see themselves doing it again. We found that although this was a new and unconventional assignment, these elements did not have a negative effect on students' grades. Although our study is limited by its small size, when considered with the numerous other reports of replication assignments in other fields, we believe this approach will work in many types of archaeology classes. Replication assignments have an important role in closing the gap between the ideals of archaeological science and preparing students to tackle the practical challenges of doing archaeological science transparently and reproducibly.

To make it easier to conduct replication assignments with archaeology students in the future, we recommend instructors share their syllabi and assignment instructions in trustworthy repositories such as the Open Science Framework (https://osf.io/) or Dataverse (https://dataverse.harvard.edu/). Currently, it is hard to find examples and a more systematic and open way of sharing might reduce preparation time for instructors (cf. Höffler 2017). A second future direction in teaching replication is for archaeologists to share information about the software tools they use to make reproducible research easier. This information can be useful to guide instructors on what to teach students as part of their methods and software training (Janz 2016). In our review above we noted that teaching the use of tools like R, Markdown, Git and GitHub has already been embraced in many fields as core elements of graduate programs. Archaeology programs must place a greater emphasis on giving students skills to use tools that enable transparent and reproducible research.

Acknowledgements

Thanks to the students of ARCHY 483 in Spring 2019 for participating in the assignment. Thanks to the peer reviewers for their detailed feedback and suggestions for improvement.

References

Ball, Richard, and Norm Medeiros. 2012. "Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis." *The Journal of Economic Education* 43 (2): 182–89. https://doi.org/10.1080/00220485.2012.659647.

Barba, Lorena A. 2018. "Terminologies for Reproducible Research." *arXiv Preprint arXiv:1802.03311*.

Baumer, Ben, Mine Cetinkaya-Rundel, Andrew Bray, Linda Loi, and Nicholas J Horton. 2014. "R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics." *Technology Innovations in Statistics Education* 8 (1).

Bicho, N., and J. Cascalheira. 2020. "The Use of Lithic Assemblages for the Definition of Short-Term Occupations in Hunter-Gatherer Prehistory." In *Short-Term Occupations in Paleolithic Archaeology. Interdisciplinary Contributions to Archaeology*, edited by A. Picin and J. Cascalheira. Switzerland: Interdisciplinary Contributions to Archaeology. Springer. https://doi.org/https://doi.org/10.31235/osf.io/3wgsa.

Breslawski, Ryan P., Bonnie L. Etter, Ian Jorgeson, and Matthew T. Boulanger. 2018. "The Atlatl to Bow Transition: What Can We Learn from Modern Recreational Competitions?" *Lithic Technology* 43 (1): 26–37. https://doi.org/10.1080/01977261.2017.1416918.

Cetinkaya-Rundel, Mine, and Colin Rundel. 2018. "Infrastructure and Tools for Teaching Computing Throughout the Statistical Curriculum." *The American Statistician* 72 (1). Taylor & Francis: 58–65.

Collins, Harry. 1992. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.

Dvorak, Tomas, Simon D Halliday, Michael O'Hara, and Aaron Swoboda. 2019. "Efficient Empiricism: Streamlining Teaching, Research, and Learning in Empirical Courses." *The Journal of Economic Education*. Taylor & Francis, 1–16.

Frank, Michael C, and Rebecca Saxe. 2012. "Teaching Replication." *Perspectives on Psychological Science* 7 (6). Sage Publications Sage CA: Los Angeles, CA: 600–604.

Hawkins, Robert XD, Eric N Smith, Carolyn Au, Juan Miguel Arias, Rhia Catapano, Eric Hermann, Martin Keil, et al. 2018. "Improving the Replicability of Psychological Science Through Pedagogy." *Advances in Methods and Practices in Psychological Science* 1 (1). SAGE Publications Sage CA: Los Angeles, CA: 7–18.

Höffler, Jan H. 2013. "Teaching Replication in Quantitative Empirical Economics." In *World Economics Association (Wea), Conference on the Economics Curriculum: Towards a Radical Reformation: May*, 3–31.

———. 2017. "ReplicationWiki: Improving Transparency in Social Sciences Research." *D-Lib Magazine* 23 (3). Corporation for National Research Initiatives: 1.

Janz, Nicole. 2016. "Bringing the Gold Standard into the Classroom: Replication in University Teaching." *International Studies Perspectives* 17 (4). Oxford University Press: 392–407.

Jern, Alan. 2018. "A Preliminary Study of the Educational Benefits of Conducting Replications in the Classroom." *Scholarship of Teaching and Learning in Psychology* 4 (1). Educational Publishing Foundation: 64.

Marwick, Ben. 2013. "Multiple Optima in Hoabinhian Flaked Stone Artefact Palaeoeconomics and Palaeoecology at Two Archaeological Sites in Northwest Thailand." *Journal of Anthropological Archaeology* 32 (4): 553–64. https://doi.org/http://dx.doi.org/10.1016/j.jaa.2013.08.004.

———. 2017. "Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation." *Journal of Archaeological Method and Theory* 24 (2): 424–50. https://doi.org/10.1007/s10816-015-9272-9.

Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. "Packaging Data Analytical Work Reproducibly Using R (and Friends)." *The American Statistician* 72 (1). Taylor & Francis: 80–88.

Marwick, Ben, Chris Clarkson, Sue O'Connor, and Sophie Collins. 2016. "Early Modern Human Lithic Technology from Jerimalai, East Timor." *Journal of Human Evolution* 101: 45–64.

Marwick, Ben, Jade d'Alpoim Guedes, C Michael Barton, Lynsey A Bates, Michael Baxter, Andrew Bevan, Elizabeth A Bollwerk, et al. 2017. "Open Science in Archaeology." *SAA Archaeological Record* 17 (4). Society for American Archaeology: 8–14.

Marwick, Ben, and Sophie C Schmidt. 2019. "Tool-Driven Revolutions in Archaeological Science." OSF. https://doi.org/10.17605/OSF.IO/RHVN5.

Millman, K. Jarrod, Matthew Brett, Ross Barnowski, and Jean-Baptiste Poline. 2018. "Teaching Computational Reproducibility for Neuroimaging." *Frontiers in Neuroscience* 12. https://doi.org/10.3389/fnins.2018.00727.

National Academies of Sciences, Engineering, Medicine, and others. 2019. "Reproducibility and Replicability in Science." *Washington, DC: The National Academies Press), Doi* 10: 25303.

Nosek, Brian A, George Alter, George C Banks, Denny Borsboom, Sara D Bowman, Steven J Breckler, Stuart Buck, et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242). American Association for the Advancement of Science: 1422–5.

Perkel, Jeffrey M. 2017. "How Scientists Use Slack." *Nature News* 541 (7635): 123.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roettger, Timo B, and Dinah Baer-Henney. 2018. "Towards a Replication Culture in Phonetic Research: Speech Production Research in the Classroom." *PsyArXiv Preprint Https://Psyarxiv.com/Q9t7c*.

Schmidt, S, and Ben Marwick. 2019. "Tool-Driven Revolutions in Archaeological Science." *SocArXiv Preprint Doi:10.31235/Osf.io/4nkxv*.

Spearman, Charles. 1904. "The Proof and Measurement of Association Between Two Things." *American Journal of Psychology* 15 (1): 72–101.

Standing, Lionel G, Manuel Grenier, Erica A Lane, Meigan S Roberts, and Sarah J Sykes. 2014. "Using Replication Projects in Teaching Research Methods." *Psychology Teaching Review* 20 (1). ERIC: 96–104.

Toelch, Ulf, and Dirk Ostwald. 2018. "Digital Open Science—Teaching Digital Tools for Reproducible and Transparent Research." *PLoS Biology* 16 (7). Public Library of Science: e2006022.

Wessa, Patrick. 2009. "How Reproducible Research Leads to Non-Rote Learning Within Socially Constructivist Statistics Education." *Electronic Journal of E-Learning* 7 (2). ERIC: 173–82.

Xie, Yihui, Joseph J Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. CRC Press.

Yan, Lisa, and Nick McKeown. 2017. “Learning Networking by Reproducing Research Results.” *ACM SIGCOMM Computer Communication Review* 47 (2). ACM: 19–26.

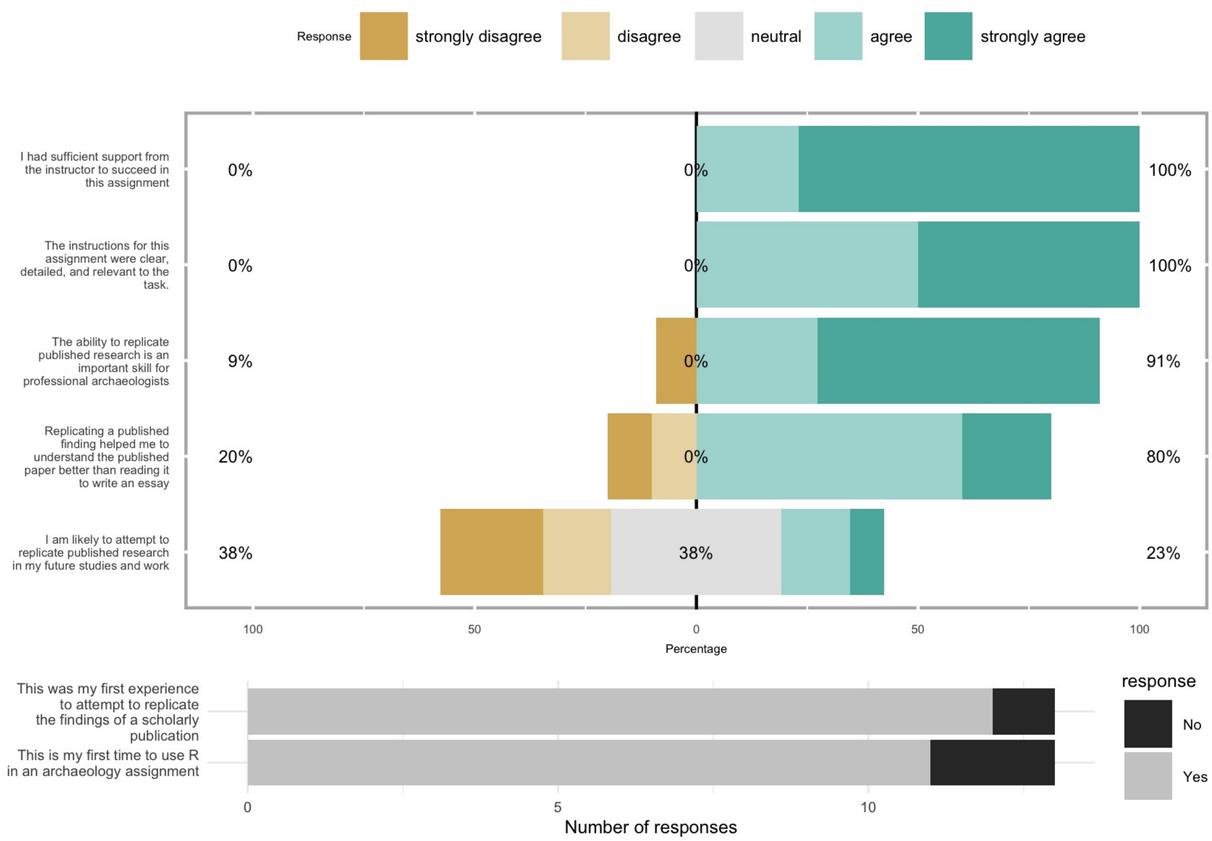

Figure 1: Results of the anonymous feedback survey on the replication assignment

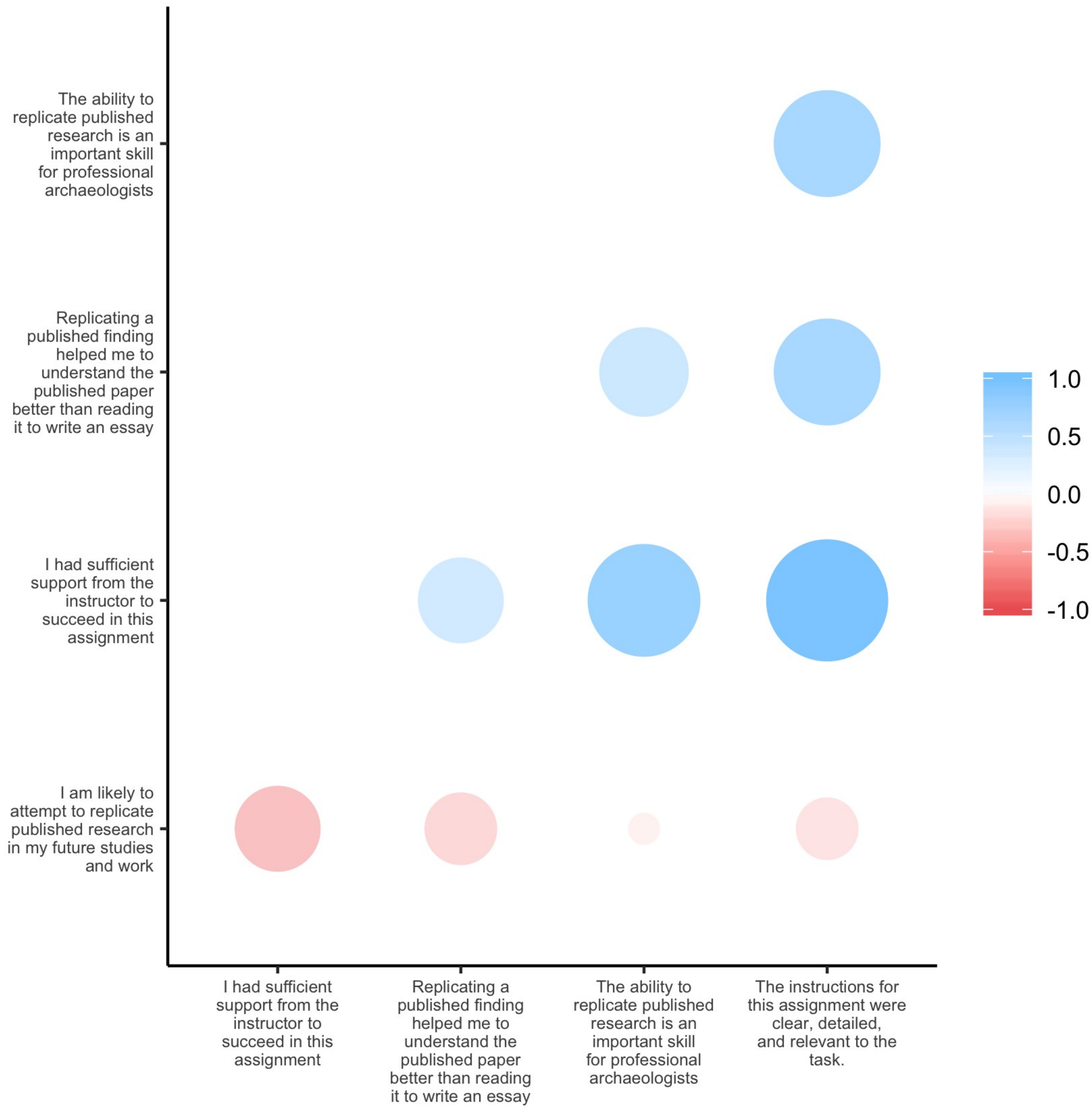

Figure 2: Correlations among feedback items with Likert scale responses. The size of the dot indicates the magnitude of the correlation, and the colour indicates the direction (red is negative, blue is positive). Correlations were computed using Spearman's (1904) method.

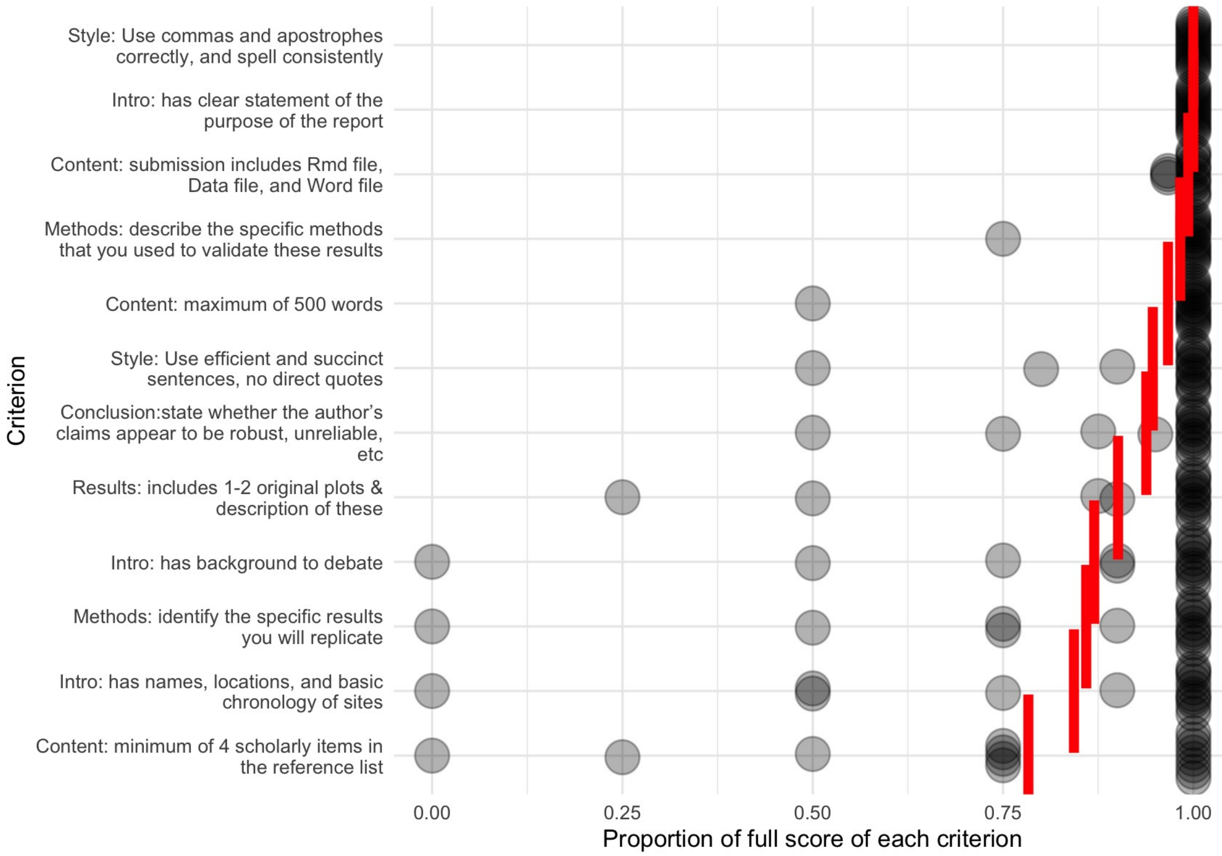

Figure 3: Distribution of students' scores across the grading rubric criteria. Each point is one student. Red lines indicate the mean score for all students per criterion.

Colophon

This report was generated on 2019-08-17 10:26:04 using the following computational environment and dependencies:

```
#> ─ Session info
──────────────────────────────────────────────────────────
#>  setting  value
#>  version  R version 3.6.0 (2019-04-26)
#>  os       macOS Mojave 10.14.6
#>  system   x86_64, darwin15.6.0
#>  ui       X11
#>  language (EN)
#>  collate  en_US.UTF-8
#>  ctype    en_US.UTF-8
#>  tz       America/Los_Angeles
#>  date     2019-08-17
#>
#> ─ Packages
──────────────────────────────────────────────────────────────
#>  package     * version    date       lib
```

```
#>  assertthat    0.2.1      2019-03-21 [1]
#>  backports     1.1.4      2019-04-10 [1]
#>  beeswarm      0.2.3      2016-04-25 [1]
#>  bookdown      0.12       2019-07-11 [1]
#>  broom         0.5.2      2019-04-07 [1]
#>  callr         3.3.1      2019-07-18 [1]
#>  cellranger    1.1.0.9000 2019-05-28 [1]
#>  cli           1.1.0      2019-03-19 [1]
#>  colorspace    1.4-1      2019-03-18 [1]
#>  corrr       * 0.4.0      2019-07-12 [1]
#>  cowplot     * 1.0.0      2019-07-11 [1]
#>  crayon        1.3.4      2019-05-28 [1]
#>  desc          1.2.0      2018-05-01 [1]
#>  devtools      2.1.0      2019-07-06 [1]
#>  digest        0.6.20     2019-07-04 [1]
#>  dplyr       * 0.8.3      2019-07-04 [1]
#>  evaluate      0.14       2019-05-28 [1]
#>  forcats     * 0.4.0      2019-02-17 [1]
#>  foreign       0.8-71     2018-07-20 [1]
#>  fs            1.3.1      2019-05-06 [1]
#>  generics      0.0.2      2018-11-29 [1]
#>  ggbeeswarm  * 0.6.0      2017-08-07 [1]
#>  ggplot2     * 3.2.1      2019-08-10 [1]
#>  glue          1.3.1      2019-03-12 [1]
#>  gridExtra     2.3        2017-09-09 [1]
#>  gtable        0.3.0      2019-03-25 [1]
#>  haven         2.1.1      2019-07-04 [1]
#>  here          0.1        2017-05-28 [1]
#>  highr         0.8        2019-03-20 [1]
#>  hms           0.5.0      2019-07-09 [1]
#>  htmltools     0.3.6      2017-04-28 [1]
#>  httr          1.4.1      2019-08-05 [1]
#>  jsonlite      1.6        2018-12-07 [1]
#>  knitr       * 1.24       2019-08-08 [1]
#>  labeling      0.3        2014-08-23 [1]
#>  lattice       0.20-38    2018-11-04 [1]
#>  lazyeval      0.2.2      2019-03-15 [1]
#>  likert      * 1.3.5      2016-12-31 [1]
#>  lubridate     1.7.4      2018-04-11 [1]
#>  magrittr      1.5        2014-11-22 [1]
#>  memoise       1.1.0      2017-04-21 [1]
#>  mnormt        1.5-5      2016-10-15 [1]
#>  modelr        0.1.5      2019-08-08 [1]
#>  munsell       0.5.0      2018-06-12 [1]
#>  nlme          3.1-140    2019-05-12 [1]
#>  pillar        1.4.2      2019-06-29 [1]
#>  pkgbuild      1.0.4      2019-08-05 [1]
#>  pkgconfig     2.0.2      2018-08-16 [1]
#>  pkgload       1.0.2      2018-10-29 [1]
#>  plyr          1.8.4      2016-06-08 [1]
```

```
#>  png            0.1-7      2013-12-03 [1]
#>  prettyunits    1.0.2      2015-07-13 [1]
#>  processx       3.4.1      2019-07-18 [1]
#>  ps             1.3.0      2018-12-21 [1]
#>  psych          1.8.12     2019-01-12 [1]
#>  purrr        * 0.3.2      2019-03-15 [1]
#>  R6             2.4.0      2019-02-14 [1]
#>  Rcpp           1.0.2      2019-07-25 [1]
#>  readr        * 1.3.1      2018-12-21 [1]
#>  readxl         1.3.1      2019-03-13 [1]
#>  remotes        2.1.0      2019-06-24 [1]
#>  reshape2       1.4.3      2017-12-11 [1]
#>  rlang          0.4.0      2019-06-25 [1]
#>  rmarkdown      1.14       2019-07-12 [1]
#>  rprojroot      1.3-2      2018-01-03 [1]
#>  rstudioapi     0.10       2019-03-19 [1]
#>  rvest          0.3.4      2019-05-15 [1]
#>  scales         1.0.0      2018-08-09 [1]
#>  sessioninfo    1.1.1      2018-11-05 [1]
#>  stringi        1.4.3      2019-03-12 [1]
#>  stringr      * 1.4.0      2019-02-10 [1]
#>  testthat       2.2.1      2019-07-25 [1]
#>  tibble       * 2.1.3      2019-06-06 [1]
#>  tidyr        * 0.8.3      2019-03-01 [1]
#>  tidyselect     0.2.5      2018-10-11 [1]
#>  tidyverse    * 1.2.1      2017-11-14 [1]
#>  usethis        1.5.1      2019-07-04 [1]
#>  vctrs          0.2.0      2019-07-05 [1]
#>  vipor          0.4.5      2017-03-22 [1]
#>  withr          2.1.2      2018-03-15 [1]
#>  xfun           0.8        2019-06-25 [1]
#>  xml2           1.2.2.9000 2019-08-15 [1]
#>  xtable       * 1.8-4      2019-04-21 [1]
#>  yaml           2.2.0      2018-07-25 [1]
#>  zeallot        0.1.0      2018-01-28 [1]
#>  source
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  Github (rsheets/cellranger@7ecde54)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  Github (gaborcsardi/crayon@84be620)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
```

```
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
```

```
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  Github (hadley/xml2@5bf8de1)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>  CRAN (R 3.6.0)
#>
#> [1] /Library/Frameworks/R.framework/Versions/3.6/Resources/library
```

The current Git commit details are:

```
#> Local:    master /Users/bmarwick/Desktop/teaching-replication-in-
archaeology
#> Remote:   master @ origin (https://github.com/benmarwick/teaching-
replication-in-archaeology)
#> Head:     [e42f3ec] 2019-08-16: Add Hope's commas 👓
```