Introduction to Reproducible Research in Social Sciences

John P. Burkett

January 12, 2018

An interesting empirical research report may inspire its readers to try to reproduce and perhaps extend its data analysis. However, as data sets have grown bigger and analytic methods more complex, reproducing published results has become harder, exacerbating a “reproducibility crisis” [1]. In response to the crisis, scientists in several fields have recommended stronger efforts to ensure reproducibility [24]. Such efforts can benefit not only other investigators but also the original investigators if they resume work on a project after forgetting some of its initial steps.

Reproducibility may have several components, including data availability and rights to use proprietary data and copyrighted materials. However, a central component is usually computational reproducibility. A research project is said to be computationally reproducible “if a second investigator... can re-create the final reported results of the project, including key quantitative findings, tables, and figures, given only a set of files and written instructions” [12, p. xxii].

A good overview of computational reproducibility is provided by The Practice of Reproducible Research[13], which includes an excellent account of how do perform reproducible research, complete with an easily implemented example [11]. While working through this example, I encountered only two minor snags: First, the text suggests creating a data file in which one variable is labeled “Mass” (p. 24) but lists R code in which that variable is called “Weight” (p. 28). Revising the code file to replace “Weight” by “Mass” suffices to make the code work as intended. Second, the text suggests creating a shell script containing the commands r clean_data.R and r analysis.R (p. 29). My computer was unable to execute those commands but responded as intended to slightly revised versions: R CMD BATCH clean_data.R and R CMD BATCH analysis.R.

Interest in reproducible research has been growing in several social sciences, including anthropology [4], archaeology [9, 19, 20], cognitive neuroscience [8, 17, 27], economics [6, 14, 15, 16], epidemiology and public health [3, 25], geography [5], history [23], management [28], political science [2, 7, 18], psychology [10, 22, 24], and sociology [29].

Efforts to train social science students in reproducible research methods have been undertaken by the project on Teaching Integrity in Empirical Research (TIER) [30]. Suggestions for such training with respect to economics are offered by Medeiros and Ball [21]. Further efforts to provide such training could greatly improve the accuracy, credibility, and impact of future social science research.
References

Summary: Baker reports the results of a survey of 1576 researchers. Most respondents believed there is a crisis. Over 80% believed that factors contributing to irreproducibility include unavailable methods and code.

[2] Pablo Barberá. The trade-off between reproducibility and privacy in the use of social media data to study political behavior. In Kitzes et al. [13], pages 103–08.
Summary: Barberá describes the workflow in his study of “political polarization on social media websites” (p. 103).

Abstract: This mini-review briefly discusses the importance of reproducible research in the fields of epidemiology and public health. Advantages of reproducible research as well as pitfalls and shortcomings of failure to conduct reproducible research in these two important fields are discussed.

Summary: Beheim notes that “a statistical analysis in anthropology involves literally thousands of individual steps.... Herein lies the real value of scripting your work in a statistical computing language like R or Stata; the code is self-documenting. With everything in timeless, universal plaintext (think .csv, .txt, .r, .dta, or .dat), an analysis can be re-run automatically, and becomes available for review and cumulative improvement.”

Abstract: Reproducible quantitative research is research that has been documented sufficiently rigorously that a third party can replicate any quantitative results that arise. It is argued here that such a goal is desirable for quantitative human geography, particularly as trends in this area suggest a turn towards the creation of algorithms and codes for simulation and the analysis of Big Data. A number of examples of good practice in this area are considered, spanning a time period from the late 1970s to the present day. Following this, practical aspects such as tools that enable research to be made reproducible are discussed, and some beneficial side effects of adopting the practice are identified. The paper concludes by considering some of the challenges faced by quantitative geographers aspiring to publish reproducible research.

Abstract: We attempted to replicate 67 macroeconomic papers using author-provided data and code files by following a preanalysis plan. Excluding 6 papers that used confidential data, we obtained data and code replication files for 29 of 35 papers (83 percent) that were required to provide such files as a condition of publication, compared to 11 of 26 papers (42 percent) that were not required to provide such files. Also excluding the 2 papers that used software we did not possess, we replicated 29 of 59 papers (49 percent) with assistance from the authors. We conclude with recommendations on improving replication of economics research.


Abstract: In April 2013, a controversy arose when a working paper (Herndon, Ash, and Pollin 2013) claimed to show serious errors in a highly cited and influential economics paper by Carmen Reinhart and Kenneth Rogoff (2010). The Reinhart and Rogoff paper had come to serve as authoritative evidence in elite conversations (Krugman 2013) that high levels of debt, especially above the “90 percent [debt/GDP] threshold” (Reinhart and Rogoff 2010, 577), posed a risk to economic growth. Much of the coverage of this controversy focused on an error that was a “perfect made-for-TV mistake” (Stevenson and Wolfers 2013) involving a simple error in the formula used in their Excel calculations. The real story here, however, is that it took three years for this error and other issues to be discovered because replication files were not publicly available, nor were they provided to scholars when asked. If professional norms or the *American Economic Review* had required that authors publish replication files, this debate would be advanced by three years and discussions about austerity policies would have been based on a more clear-sighted appraisal of the evidence.


Deniz describes how she used FlickrAPI, Git, GitHub, and Python in her cognitive neuroscience research.


Summary: Ducke discusses “how digital science and archaeological computing interact with each other at the interface of software technology” and notes conflicts between “good scientific practice” and “commercially licensed, proprietary software.”

Abstract: Authors who wish to publish their work with us have the option of a registered report. With this format, acceptance in principle happens before the research outcomes are known. As a result, publication bias is neutralized, as are incentives for practices that undermine the validity of scientific research.


Summary: Kitzes describes a “reproducible workflow that any scientist should master” and provides a simple example of how to document and archive raw data, cleaned data, computer code used in cleaning and analyzing data, and results.


Summary: Kitzes states that the book’s “goal is to provide concrete advice and examples that will demonstrate how you can make your computational and data-intensive research more clear, transparent, and organized” (p. xxii).


Summary: Kleiber discusses reproducibility in economics, considers case studies in “forensic econometrics,” and proposes steps to increase reproducibility.


Abstract: The practice of reproducible research, a central component of the burgeoning “open science” movement, has been thrust into the public spotlight over the past few years. In this paper, I offer an overview of reproducibility in science, review specific concerns for the real estate field, and survey the current policy regarding reproducibility among top real estate journals. Performing research reproducibly requires a change from the status quo and represents an educational issue. Toward that end, I demonstrate reproducible research via a fully documented and freely-available example of a reproducible hedonic price analysis complete with all data, code, and results hosted online.

Summary: The authors describe how they conducted and documented their research using GitHub, knitr, \LaTeX, R, and RStudio.


Summary: The authors describe how they conducted and documented their cognitive neuroscience research using Advanced Normalization Tools, ANFI, FSL, FreeSurfer, Make, and R Markdown.


Summary: Magallanes describes how he conducted and documented his research using GitHub, \LaTeX, Python, R, RStudio, and Zotero.


Abstract: The use of computers and complex software is pervasive in archaeology, yet their role in the analytical pipeline is rarely exposed for other researchers to inspect or reuse. This limits the progress of archaeology because researchers cannot easily reproduce each other’s work to verify or extend it. Four general principles of reproducible research that have emerged in other fields are presented. An archaeological case study is described that shows how each principle can be implemented using freely available software. The costs and benefits of implementing reproducible research are assessed. The primary benefit, of sharing data in particular, is increased impact via an increased number of citations. The primary cost is the additional time required to enhance reproducibility, although the exact amount is difficult to quantify.


Summary: Marwick describes how he conducted and documented his research with the help of boot2docker, CircleCI, Docker, Git, GitHub, R, RStudio, and R packages such as dplyr, ggplot2, and knitr.


Summary: The chapter describes Medeiros and Ball’s approach to helping students get into the habit of “writing and iteratively revising editable command files, rather than using drop-down menus or interactively executing one command at a time.”
The chapter also describes guidelines for students about how to assemble “a set of electronic documents—including the original data used in the study, the command files they write, and supplementary information that serve as comprehensive replication documentation” (p. 262).


Summary: The authors describe their statistical analysis of psychological data collected during clinical interaction “with children on the autistic spectrum” (p. 201). Parts of their analysis were conducted and documented using Git, GitHub, \LaTeX, NumPy, Python (especially its nose and permute packages), reStructuredText, and Sphinx.


Summary: The author explains to his fellow historians why and how to ensure that their computational work is documented, version controlled, “automated with a build script,” and “written as a literate program...with all dependencies managed.”


Abstract: Improving the reliability and efficiency of scientific research will increase the credibility of the published scientific literature and accelerate discovery. Here we argue for the adoption of measures to optimize key elements of the scientific process: methods, reporting and dissemination, reproducibility, evaluation and incentives. There is some evidence from both simulations and empirical studies supporting the likely effectiveness of these measures, but their broad adoption by researchers, institutions, funders and journals will require iterative evaluation and improvement. We discuss the goals of these measures, and how they can be implemented, in the hope that this will facilitate action toward improving the transparency, reproducibility and efficiency of scientific research.


Summary: Ottoboni describes how she used knitr, \LaTeX, R, and RStudio to conduct and document her research regarding the effects of salt consumption “on a nation’s life expectancy at 30” (p. 221).

Abstract: Powerful new social science data resources are emerging. One particularly important source is administrative data, which were originally collected for organisational purposes but often contain information that is suitable for social science research. In this paper we outline the concept of reproducible research in relation to micro-level administrative social science data. Our central claim is that a planned and organised workflow is essential for high quality research using micro-level administrative social science data. We argue that it is essential for researchers to share research code, because code sharing enables the elements of reproducible research. First, it enables results to be duplicated and therefore allows the accuracy and validity of analyses to be evaluated. Second, it facilitates further tests of the robustness of the original piece of research. Drawing on insights from computer science and other disciplines that have been engaged in e-Research we discuss and advocate the use of Git repositories to provide a usable and effective solution to research code sharing and rendering social science research using micro-level administrative data reproducible.

Summary: Poldrack describes how he used GitHub, LATEX, Matlab, Python, R, R Markdown, and Vagrant to conduct and document his research on “brain systems involved in decision-making and executive control” (p. 311).

Objectives: The authors aim to provide guidelines for reporting “simulation-based research ... in the social sciences, with a focus on common scenarios in system dynamics research” and additional “guidelines to improve visualization of research to reduce the costs of reproduction” (p. 396).

Summary: Salganik outlines reproducible research’s tasks and methods as they relate to sociology.

Summary: The website describes the project’s approach to promoting “the integration of principles and practices related to transparency and replicability in the research training of social scientists.”