

The practice of open science in archaeology using R and online digital repositories in case study research



Liying Wang and Ben Marwick
University of Washington

Introduction

- The typical publication process in archaeology involves data analysis, visualization, and interpretation. These are normally done separately, in different software, and the outputs combined during the final writing stages.
- Research transparency is limited due to the disconnected nature of this process, and the availability of only the final product in publication.

Open science

- Open science emphasizes the free, public availability of research data, methods, and publications.
- It enables others to more fully examine the research process to enhance research credibility and reproducibility, and improve equity and inclusiveness in participation in research.
- We use a case study from an Iron Age site in northeastern Taiwan to demonstrate an open science workflow that is relevant and practical for archaeology.

Workflow in a case study

- We used the R software for analysis and visualization, and organized our files into a research compendium. This standard compendium structure enables other researchers to easily inspect our data and analytical processes.
- We wrote the manuscript with R Markdown (Rmd) to transparently combine our manuscript text and R code for data analysis and visualization. This Rmd file can be 'knitted' (executed) to typical document formats (Word, PDF, etc) for paper submission.
- During the writing process, we used Git for collaboration and version control. We continuously uploaded updates of our work to a repository created for this project on Github.
- Once the writing was complete, we deposited a preprint and archived all our research materials for long-term availability on the Open Science Framework (OSF).

Our workflow of open science is practical, easily learned, and applicable to many other types of archaeological research.

Knit Rmd file to create an output

Git for version control

Rmd file showing a manuscript that includes code and text

Fig 1. Our case study using R and R Markdown for publication writing in RStudio

Data is read into the R environment for analysis and visualization

A research compendium separates research materials into data and code, and other folders as we customized

Data and Code Availability
The data and R code demonstrated here are openly available online at Github: <https://github.com/LiYingWang/kwl.pottery>, and at OSF: <https://doi.org/10.17605/OSF.IO/ABVGF>

Research compendium



```
405 #making the cv label
406 cv_label <-
407   cv_pcs_all %>%
408   mutate(cv_pc1 = paste("CV =", cv_pc1, "%"),
409          cv_pc2 = paste("CV =", cv_pc2, "%"),
410          cv_pc3 = paste("CV =", cv_pc3, "%"))
411
412 cv_label <- c(cv_label$cv_pc1, cv_label$cv_pc2, cv_label$cv_pc3)
413
414 ggplot(pc_plot,
415        aes(phase,
416            pc)) +
417   geom_violin() +
418   ggforce::geom_sina() +
419   xlab("") +
420   facet_wrap(~ pcn, ncol = 1) +
421   geom_text(data = cv_pcs_all,
422            mapping = aes(x = c(2.1, 1.1, 3.1, 2.1, 1.1, 3.1, 2.1, 1.1, 3.1),
423                          y = 8.5,
424                          label = cv_label),
425            size = 2,
426            color = "red",
427            hjust = -0.1,
428            vjust = -1) +
429   theme_minimal()
```

Fig 2. R code in the manuscript using R markdown

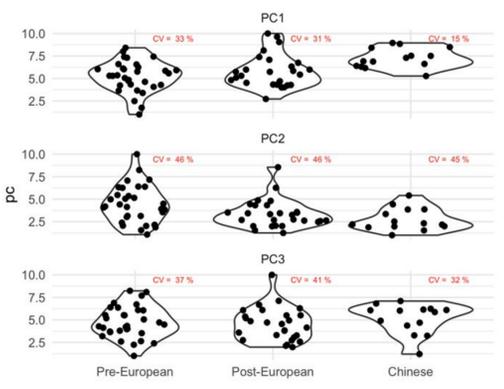


Fig 3. The plot produced by the code shown in Fig 2

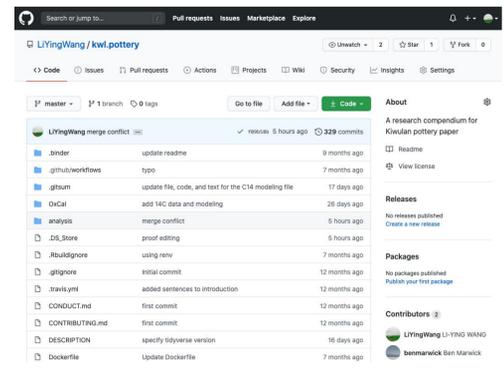


Fig 4. A Github repo that hosts research materials and any updates

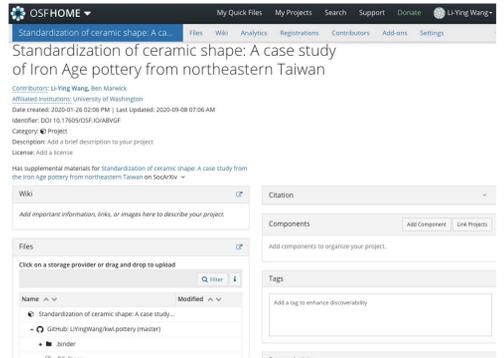


Fig 5. An OSF repo for sharing preprint and materials for long-term availability