

A TAO for Data Wrangling: A Practical Routine for Getting Past the ‘Janitor Work’

Memo to Editors

We thank the editors for their valuable and constructive comments that have helped us improve this manuscript. Below, we summarize how we have addressed each of the key suggestions.

The primary revision concerns the R package recommendation section in the supplementary material. Instead of the long and heterogeneous list of packages, the editors hoped we would provide more practical recommendations, such as a top-five “must-use” list. We were actually uncertain how best to rank the top packages, given that our recommendations target a general audience whose research questions and data formats vary widely. Instead, in this version, we provide more detailed introductions to four packages that play crucial roles in the data wrangling processes of the DCPO and SWIID projects. We hope these illustrations give readers a better sense of how packages can improve the automation of data wrangling. Meanwhile, we retain the package tables based on the R search function as a supplementary resource. We include this table not to recommend that readers “must” try each package, but to show which packages most researchers use in the R community.

Apart from the revision of the package recommendation section, we have also corrected the mismatched section numbers and further revised the language throughout this new version. The replication file (a Quarto document) is also attached with the revision.

We would like to thank the editors again for their valuable feedback. We hope you find this revision helpful.