The growing importance of reproducibility and responsible workflow in the data science and statistics curriculum

Aneta Piekut
Univ. of Sheffield

Colin Rundel
Duke University

Micaela Parker
ADSA

Nicholas Horton
Amherst College

Rohan Alexander
Univ. of Toronto

CAUSE/JSDSE webinar series

Welcome!

Host and Moderator:
Leigh Johnson
Capital University

What's new in the journal?

Latest Article Link

Collections Link

January CAUSE/JSDSE Webinar

SCRATCH to R: Toward an Inclusive Pedagogy in Teaching Coding

Shu-Min Liao
Amherst College

January 17, 2023 at 4:00 PM EST

Consortium for the Advancement of Undergraduate Statistics Education

Important Dates:

Posters & Beyond

- January 29, 2023 (will receive feedback)
- March 5, 2023 (final abstracts due)

Birds-of-a-Feather Discussions

- Proposals due May 6, 2023

Conference runs June 1 - 3, 2023 (with workshops beginning May 30th) at the Penn Stater Hotel in State College, PA

https://www.causeweb.org/cause/uscots/uscots23

Consortium for the Advancement of Undergraduate Statistics Education

USPROC | The Undergraduate Statistics Project Competition

https://www.causeweb.org/usproc/
Submission Deadline: June 23, 2023

Reading Group

Article Link
Zoom Registration Link
Dec. 14, 2023 12:00 - 1:00 PM

Aneta Piekut

Aneta Piekut
a.piekut@sheffield.ac.uk

Aneta Piekut is Senior Lecturer in Quantitative Social Sciences at the Sheffield Methods Institute, University of Sheffield. She is an associate editor of the *Journal of Statistics and Data Science Education*.

Colin Rundel

Colin Rundel
rundel@gmail.com

Colin Rundel is an associate professor of the practice in Statistical Science at Duke University. His research interests include applied spatial statistics with an emphasis on Bayesian statistics and computational methods. Colin was a guest editor of the special issue on teaching reproducibility and responsible workflow.

Micaela Parker

Micaela Parker
micaela@academicdatascience.org

Micaela Parker is the Founder and Executive Director of the Academic Data Science Alliance (ADSA), https://academicdatascience.org. Micaela was a guest editor of the special issue on teaching reproducibility and responsible workflow.

Nicholas J. Horton

Nicholas J. Horton
nhorton@amherst.edu

Nicholas J. Horton is Beitzel Professor of Technology and Society (Statistics and Data Science) at Amherst College. He is the editor of the *Journal of Statistics and Data Science Education* and is the co-chair of the National Academies Committee on Applied and Theoretical Statistics (CATS).

Rohan Alexander

Rohan Alexander
rohan.alexander@utoronto.ca

Rohan Alexander is an assistant professor at the University of Toronto jointly appointed in the Faculty of Information and the Department of Statistical Sciences. He is an associate editor of the *Journal of Statistics and Data Science Education*.

Papers in the issue

- “The growing importance of reproducibility and responsible workflow in the data science and statistics curriculum” (Nicholas J. Horton, Rohan Alexander, Aneta Piekut, Colin Rundel, https://doi.org/10.1080/26939169.2022.2141001)

- “An invitation to teaching reproducible research: lessons from a symposium” (Richard Ball, Norm Medeiros, Nicholas W. Bussberg, and Aneta Piekut, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2099489)

- “Interdisciplinary approaches and strategies from research reproducibility 2020: educating for reproducibility” (Melissa L. Rethlefsen, Hannah F. Norton, Sarah L. Meyer, Katherine A. MacWilkinson, Plato L. Smith II, and Hao Ye, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2104767)

- “Data science ethos lifecycle: interplay of ethical thinking and data science practice” (Margarita Boenig-Liptsin, Anissa Tanweer, and Ari Edmundson, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2089411)

Papers in the issue

- “Opinionated practices for teaching reproducibility: motivation, guided Instruction and practice” (Joel Ostblom and Tiffany Timbers, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2074922)

- “Tools and Recommendations for Reproducible Teaching” (Mine Dogucu and Mine Çetinkaya-Rundel, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2138645)

- “Third Time’s a Charm: A Tripartite Approach for Teaching Project Organization to Students” (Christina Mehta and Renee’ Moore, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2118644)

- “LUSTRE: An online data management and student project resource” (John Towse, Rob Davies, Ellie Ball, Rebecca James, Ben Gooding, and Matthew Ivory, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2118645)

Papers in the issue

- “Teaching for Large-Scale Reproducibility Verification” (Lars Vilhuber, Hyuk Harry Son, Meredith Welch, David N. Wasser, and Michael Darisse, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2074582)
- “Collaborative Writing Workflows in the Data-Driven Classroom: A Conversation Starter” (Sara Stoudt, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2082602)
- “A Journey from Wild to Textbook Data to Reproducibly Refresh the Wages Data from the National Longitudinal Survey of Youth Database” (Dewi Amaliah, Dianne Cook, Emi Tanaka, Kate Hyde, and Nicholas Tierney, https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2094300)
- “Approachable case studies support learning and reproducibility in data science: An example from evolutionary biology” (Luna L. Sanchez Reyes and Emily Jane McTavish https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2099487)

Questions for the panel

1. What insights have you gained from these papers?

Questions for the panel

1. What insights have you gained from these papers?
2. What next steps are you going to take in response to these papers?

Questions for the panel

1. What insights have you gained from these papers?
2. What next steps are you going to take in response to these papers?
3. Audience Q&A

Questions for the panel

1. What insights have you gained from these papers?
2. What next steps are you going to take in response to these papers?
3. Audience Q&A
4. Closing thoughts: data and code sharing

Additional next steps: Data and code sharing

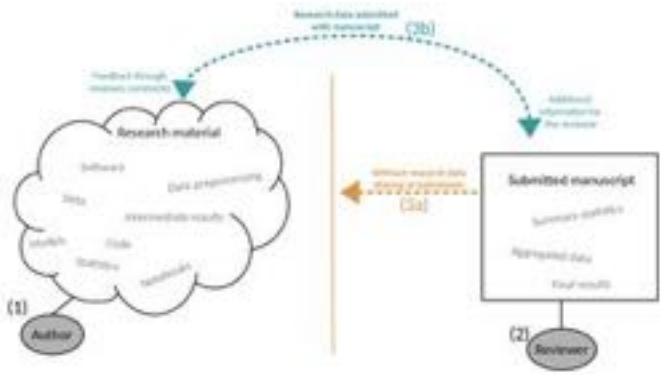
When should data and code be made available?

Significance Magazine (April, 2022)

Rachel Heyard, Leonhard Held

Pages: 4-5 | First Published: 29 March 2022

Sharing data and code as part of a research publication is crucial for ensuring the computational reproducibility of scientific work. But sharing should be done at the article submission stage, not after publication as it is now, say Rachel Heyard and Leonhard Held. Statisticians and data scientists have the skills and tools to make this change and lead by example, though there are obstacles to overcome.



As of September 1, 2022, all submissions to the *Journal of Statistics and Data Science Education* require a "Data Availability Statement" which outlines how code and data underlying a paper have been made available.

See https://www.tandfonline.com/journals/ujse21 for more info