

Virtual training on virtual environments: an online open-source introduction to conda

Marisa C.W. Lim^{*1}, Abhijna Parigi^{†1}, Saranya Canchi^{‡1}, Jose Sanchez¹, Jeremy Walter¹, Amanda Charbonneau¹, and C. Titus Brown¹

¹ Lab for Data Intensive Biology, School of Veterinary Medicine, University of California Davis

DOI: [10.21105/jose.00130](https://doi.org/10.21105/jose.00130)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 19 August 2021

Published: 27 December 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

We present our lesson material and resources for teaching how to use conda (<https://conda.io>), a tool that streamlines software installation and version management using isolated environments, while providing multiple methods for reproducing and sharing software set ups. This training material was developed for the NIH Common Fund Data Ecosystem (CFDE), whose primary goal is to teach biologists computational tools that help make their analysis workflows FAIR: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016).

We teach conda because it can be used in any research field that requires software installation, works on multiple computer operating systems (Mac, Windows, Linux), and does not require administrator privileges to configure. We piloted this material in March 2021 and have now taught the 2-hour online workshop five times to a diverse audience. The lesson material and accompanying resources provide information for instructors to adopt our hands-on workshop format and for learners to view past workshop notes or walk-through the lesson at their own pace from our [training website](#).

Statement of Need

As a consequence of switching to fully remote-compatible training due to the COVID-19 pandemic, we designed online material for self-paced and virtual workshop learning modes. Existing conda tutorials typically teach learners how to run commands with the assumption that learners' computer configuration and software versions match those in the lesson, however software and system set-up can be time-consuming, and distract or dissuade learners from wanting to use conda.

For our lesson, we specifically designed our materials to run on a binder cloud computing environment that is free and easy to launch (Jupyter et al., 2018). This structure allows learners to focus on the course content because the computing environment provides standardized interface and software versions; works for Mac, Windows, and Linux users; and reduces remote troubleshooting challenges for learners and instructors. To follow this lesson, learners only need a stable internet connection and web browser (e.g., Firefox, Chrome, Safari). While our lesson does include steps to install and set up conda on learners' own computers, it is taught after hands-on experience with conda, which allows them to decide whether it meets their research needs.

*co-first author

†co-first author

‡co-first author

Curriculum

Our introductory lesson for using conda software teaches learners how to create virtual environments and manage software installations with follow-along demonstrations and practice exercises. The curriculum is aimed at new users, as familiarity with command-line scripting is helpful but not necessary to follow the content. Our website lesson has the following learning objectives, and we focus on the first four objectives in our remote workshops:

1. Learn about conda and how to use it
2. Learn the basics of software installation, software dependencies, and isolated environments
3. Set up a conda environment
4. Learn how to manage virtual environments
5. Install Miniconda

Course Content

The lesson was designed to be self-contained and flexible: learners can go at their own pace or attend a hands-on live workshop. The material is presented with text, screenshots, and short videos to guide learners in different formats. The lesson begins with scenarios for which a virtual environment and package manager, like conda, is helpful, including managing different software versions on one computer, reproducing analysis from other projects, or sharing software configurations with collaborators. Next, we teach several methods for installing software and creating virtual software environments. After practice exercises, we highlight additional applications for conda that are beyond the scope of an introductory lesson, such as packaging software, running workflow steps in separate conda environments, and creating cloud computer interfaces like the one we use for the lesson. The final section shows how to install the lightweight Miniconda distribution of conda to a learner's local computer.

Workshop Instructional Design

We designed the lesson structure to be a 2-hour remote course, in part to accommodate busy schedules and avoid content overload. The workshop is led by two co-instructors who both teach; a moderator who tracks questions, learner participation, and time; and additional helpers who offer one-on-one help in breakout rooms. As a starting point for workshop preparation, we created several resources to provide clear guidance for teachers to follow to reuse these materials for their own workshops, including: - the [CFDE training website Intro to conda lesson](#) for self-guided learning about lesson content - [teaching notes](#) for before, during, and after workshop preparation using CFDE materials - a [workshop notes template](#) to edit and share with workshop participants - the [CFDE Intro to conda workshop notes archive](#) for reference about past events - a [conda Q&A](#) page for questions that have been asked during our workshops - [steps for creating an RStudio binder](#) to edit or create a new one

To teach conda remotely, we use a teleconferencing platform and a cloud computing interface. We chose [Zoom](#) because it is commonly used, but any platform that enables participant feedback (reactions, chat), screen share, and breakout rooms will work. We use [Pangeo Binders](#) ([Eynard-Bontemps et al., 2019](#)) with [RStudio](#) ([RStudio Team, 2019](#)) to provide participants a free cloud computing interface that comes pre-installed with workshop software and an editable workshop commands script, and can be opened from any web browser or computer operating system. This platform consolidates the terminal,

shell script commands, and file system panels to one window, which is especially helpful for managing limited computer screen space. The commands can be executed from the workshop script line by line via point and click or keyboard shortcuts, thus minimizing set-up and syntax errors.

Workshop Experience

We asked participants to fill out short pre- and post-workshop surveys to help instructors prepare, gauge participant learning, and collect feedback for workshop improvements. After five introductory conda workshops using these materials, we have trained 120 participants (1: 20, 2: 17, 3: 17, 4: 40, 5: 26) from a range of career positions (i.e., undergraduate students to faculty) and biology and technology fields.

To address the challenge of gauging learning and participation in remote workshops, we employed several successful teaching strategies: checking in with yes/no questions and exercises (responses via chat or Zoom raised hand reaction), using a standardized computing interface with workshop commands script, and providing a help breakout room. With an instructional team of 4-5, we recommend workshop sizes of 20-30 participants, as these strategies are less effective for lower instructor to participant ratios.

The survey feedback helped to refine material by reducing information overload that obscured learning objectives and to improve pacing. For example, we ask participants to launch binder environments upon arrival, so they are ready by the time we reach the hands-on portion. In addition, we now send [general resources for attendees](#) before the workshop, along with specific [workshop notes](#) that are streamlined from our website lesson to guide instructors and participants. Our surveys show that 86.8% of conda workshop participants read and found the general resources useful (n=53 responses). Participants also suggested adding time for more exercises and intermediate topics, highlighting the efficacy of hands-on workshops to help new users learn introductory material and develop confidence in exploring conda functionality further.

In summary, these remote workshops have been broadly successful, as participants reported that the pacing was not too slow or fast (Figure 1A) and a larger percentage of respondents agreed they would like to use conda for future projects in post- compared to pre-workshop surveys (Figure 1B).

Figure

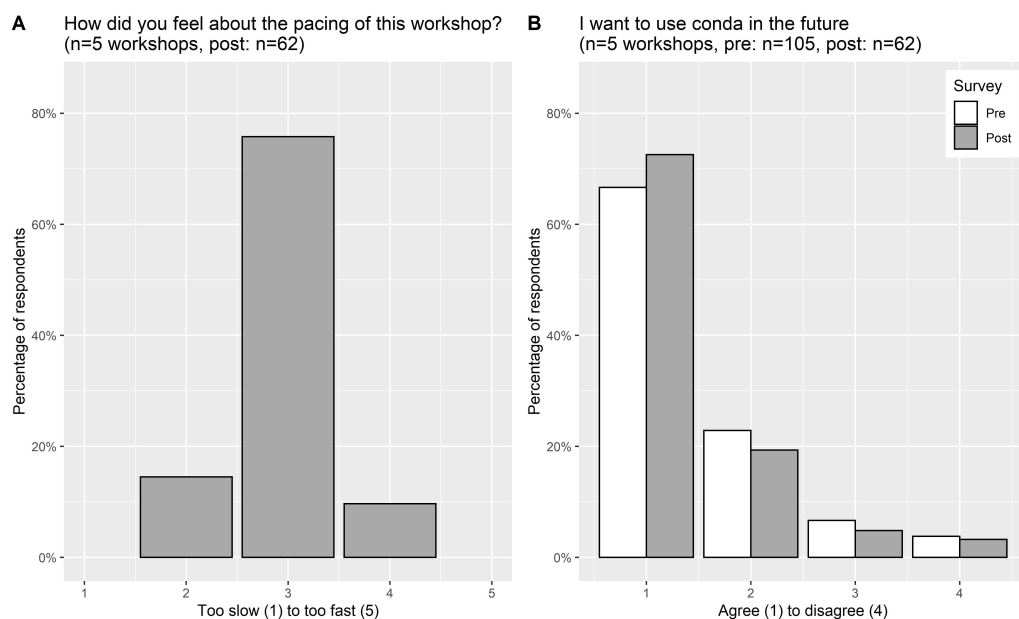


Figure 1: Survey results for (A) workshop pacing and (B) interest in using conda for future projects.

Acknowledgements

We thank our Introduction to Conda workshop participants for attending and providing feedback that improved our training material. Our lesson was inspired from material written for the [GGG298 course](#) at the University of California at Davis.

Funding

Our training program is funded by the National Institutes of Health (1OT3OD025459-01).

Author Contributions

ML, AP, and SC developed and taught the workshops. JS and JW helped coordinate the workshops. All team members provided feedback for lesson and workshop materials and this paper.

Conflicts of Interest

None.

References

- Eynard-Bontemps, G., Abernathey, R., Hamman, J., Ponte, A., & Rath, W. (2019). *The pangeo big data ecosystem and its use at CNES* [Proceedings paper]. <https://archimer.ifremer.fr/doc/00503/61441/>
- Jupyter, Project, Bussonnier, Matthias, Forde, Jessica, Freeman, Jeremy, Granger, Brian, Head, Tim, Holdgraf, Chris, Kelley, Kyle, Nalvarte, Gladys, Osheroff, Andrew, Pacer, M., Panda, Yuvi, Perez, Fernando, Ragan-Kelley, Benjamin, & Willing, Carol. (2018). Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. In Fatih Akici, David Lippa, Dillon Niederhut, & M. Pacer (Eds.), *Proceedings of the 17th Python in Science Conference* (pp. 113–120). <https://doi.org/10.25080/Majora-4af1f417-011>
- RStudio Team. (2019). *RStudio: Integrated development environment for r*. RStudio, Inc. <http://www.rstudio.com/>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., & al., et. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>