

# Kelly L. Sovacool

✉ [sovacool@umich.edu](mailto:sovacool@umich.edu)  
🌐 [sovacool.dev](https://sovacool.dev)  
in [kelly-sovacool](https://www.linkedin.com/in/kelly-sovacool)  
🔗 [kelly-sovacool](https://github.com/kelly-sovacool)

*Bioinformatician building and applying open source tools for microbiome research, and contributing to data science education along the way.*

## Education

- 2018-present **PhD Bioinformatics**, *University of Michigan*,  
Dept. of Computational Medicine and Bioinformatics
- 2014-2018 **BS Biology**, *University of Kentucky*,  
Dept. of Biology, Minor: Computer Science

## Skills

- |                   |  |
|-------------------|--|
| languages & tools | R, Python, Bash, Snakemake, R Markdown, Quarto, LaTeX, conda, git, GitHub, SLURM   |
| research          | gut microbiome research, amplicon sequence analysis, metagenomics, supervised machine learning, data visualization, reproducible manuscripts |
| software          | R package maintenance, continuous integration, high performance computing  |
| teaching          | live coding instruction, curriculum development/maintenance/deployment   |

## Research Experience

- 2019-present **Graduate Student Researcher**, *Schloss Lab*,  
Dept. of Microbiology and Immunology, University of Michigan
- Develop bioinformatics pipelines & software for microbial ecology.
  - Analyze 16S rRNA gene amplicon sequence data.
  - Apply machine learning methods to gut microbiome classification problems in colorectal cancer and *C. difficile* infection.
  - Collaborate with other scientists on microbiome projects and mentor junior lab members.
- 2018-2019 **Rotation Student Researcher**, *Program in Biomedical Sciences*,  
University of Michigan
- 2015-2018 **Undergraduate Lab Assistant**, *Moseley Bioinformatics Lab*,  
Dept. of Molecular and Cellular Biochemistry, University of Kentucky
- Developed a computational tool in Python for identifying sets of orthologous and paralogous gene products in whole genomes to facilitate collinearity analysis and detection of gene duplication events.
- 2016-2018 **BIO395 Independent Research Student**, *Weisrock Lab*,  
Dept. of Biology, University of Kentucky
- Developed bash scripts and a SNP calling pipeline in Snakemake.
  - Population structure analysis of the *Ambystoma tigrinum* species complex.
  - Bayesian species delimitation of the *Desmognathus fuscus* species complex.

- 2015-2016 **Undergraduate Lab Assistant**, *Jaromczyk Lab*,  
Dept. of Computer Science, University of Kentucky
- Maintained the *Epichloë festucae* genome project database.
  - Analyzed RNA-seq data of *Chenopodium quinoa* and coffee ringspot virus.

---

## Open Source Contributions

### Software

- mikropml [User-Friendly R Package for Supervised Machine Learning Pipelines.](#)  
Co-author and maintainer.
- mikropml [Template for running mikropml with Snakemake.](#)  
workflow Co-author and maintainer.
- schttools [Schloss Lab tools for reproducible microbiome research \(R package\).](#)  
Co-author and maintainer.
- mothur [Command-Line Tool for Processing 16S rRNA Gene Sequence Data.](#)  
Contributor.
- mothur [Snakemake template for 16S rRNA gene analysis with mothur.](#)  
workflow Co-author.

### Curricula

- Software [Intro to R, the Unix shell, and git for workshops on reproducible research.](#)  
Carpentry Co-author and maintainer.
- Girls Who [Intro to Python for Data Science for Girls Who Code clubs.](#)  
Code Co-author and maintainer.
- ASM Microbe [Intro to R & RNA-seq Workshop for ASM Microbe attendees.](#)  
2022 Contributor.
- U-M *DANG!* [repro-packs: Organizing projects for reproducibility and headache prevention.](#)  
Author. Seminar for the Data Analysis Networking Group.
- Code Clubs Short tutorials for Code Club meetings in the Schloss Lab.
- [Just enough Python for R users to write advanced Snakemake workflows.](#)
  - [Tips and tricks for making the most of GitHub.](#)
  - [How to maintain R packages.](#)
  - [Introduction to Testing R Code.](#)
  - [Re-creating plots to learn cool tricks and practice problem-solving.](#)
  - [A brief introduction to tidy evaluation.](#)
  - [Materials for a code review on exception handling in R for lab meeting.](#)
  - [A Snakemake tutorial for make users.](#)
  - [Materials for a code review on documenting R code.](#)

---

## Service

- 2018-present **Executive Committee Member: Open Source Technical Lead, Girls Who Code at U-M DCMB**
- Plan, apply for funding, develop curriculum, and maintain resources to teach introductory Python programming & data science to young women+.
  - Facilitate collaborative development and maintenance of our open source teaching resources.
  - Organize our year-round Club and annual Data Science Summer Experience for high school women+.
- 2019-present **Organizer and Maintainer, U-M Carpentries**
- Co-lead development & maintenance of a curriculum for workshops teaching programming skills for reproducible research.
  - Maintain the website, develop curriculum, and organize workshops.
  - Collaborate with U-M Women in Science and Engineering to organize workshops for women+.
- 2021-present **Mentor, Schloss Lab**
- Mentoring an undergraduate student in building reproducible machine learning models to predict *C. difficile* infection severity from gut microbiome composition.
- 2021-present **CoderSpaces co-host, U-M ISR Data Science Hub**
- Hold office hours at a weekly virtual help session for data science practitioners.
- 2019-2021 **Graduate Student Coordinator, U-M Data Analysis Networking Group**
- Organize monthly meetings & a one-day symposium for researchers to sharpen their data analysis skills.
  - Apply for funding through a Rackham Interdisciplinary Workshop grant.
- 2009-present **Live Sound Engineer, *for various churches and non-profit organizations.***
- Set up, maintain, repair, and operate front of house, monitor, and recording systems during sound checks, rehearsals, services, and concerts.
  - Record and mix live performances for streaming.
  - Train new sound techs in the art and science of live sound, live stream mixing, and recording.
- 2021-present **Peer review for scientific journals, PLOS ONE (1)**

## Teaching Experience

- 2019-2022 **Facilitator & Capstone Project Mentor, Girls Who Code at U-M DCMB**
- Aug 2021 - May 2022. Weekly virtual Club for high schoolers.
  - 12-22 July 2021. Virtual Data Science Summer Experience.
  - Aug 2020 - May 2021. Weekly virtual Club for high schoolers.
  - 06-16 July 2020. Virtual Data Science Summer Experience.
  - Aug 2019 - May 2020. Weekly Club for high schoolers.
  - 15-19 July 2019. Data Science Summer Experience in Detroit, MI.
  - Jan-May 2019. Weekly Club for high schoolers.

- 2018-present **Workshop instructor & helper**, U-M Carpentries
- 20-21 Oct 2022. Workshop instructor.
  - 28 Feb 2022. Workshop helper. Virtual workshop led by the U-M Bioinformatics core.
  - 11-12 Jan 2021. Lead instructor for virtual workshop sponsored by U-M WISE.
  - 06-07 Jan 2020. Lead instructor for in-person workshop sponsored by U-M WISE.
  - 01-02 July 2019. Workshop instructor.
  - 22-23 May 2019. Workshop helper.
  - 01 Mar 2019. Workshop helper.
  - 17-18 Dec 2018. Workshop helper.
- 01 Jun 2022 **Instructor**, *Intro to R & RNA-seq Workshop*, for ASM Microbe Conference attendees  
Virtual
- 18 Feb 2022 **Invited Speaker/Instructor**, *Git & GitHub seminar*, Norwegian University of Science and Technology, KG Jebsen Center for Genetic Epidemiology  
Virtual
- 25 Apr 2019 **DNA Day Ambassador**, *Epigenetics & scientific journeys*, MI DNA Day  
Ann Arbor, MI
- 20 Mar 2019 **Workshop helper**, *Data visualization with Python*, Graduate Society of Black Engineers and Scientists  
University of Michigan
- 16 Mar 2019 **Capstone Activity Leader**, *Binary numbers through Ozobots with GWC at U-M DCMB*, Females Excelling More in Math, Engineering, & Science  
University of Michigan
- 2012-2018 **Tutor**, for high school and college students in Biology, Calculus, Chemistry, Computer Science, and Bioinformatics.

## Presentations

### Talks

- Nov 2022 **Bioinformatics Student Research Hour**, *Predicting the severity of C. difficile infections from the taxonomic composition of the gut microbiome*, Talk  
University of Michigan
- Mar 2021 **Bioinformatics Student Research Hour**, *OptiFit: a fast method for fitting amplicon sequences to existing OTUs*, Talk  
University of Michigan (virtual)
- Apr 2018 **Systems Biology & Omics Integration Seminar**, *Developing a Global Homology Analysis for Comparative Genomics*, Talk  
University of Kentucky

### Posters

- Jun 2022 **ASM Microbe**, *Predicting the severity of C. difficile infections from the taxonomic composition of the gut microbiome*, Poster  
Washington, DC
- Jun 2020 **ASM Microbe**, *OptiFit: a fast method for fitting amplicon sequences to existing OTUs*, Poster, Cancelled due to COVID-19.

- Apr 2018 **Showcase for Undergraduate Scholars**, *Developing a Global Homology Analysis for Comparative Genomics*, Poster  
University of Kentucky
- Apr 2018 **National Conference on Undergraduate Research**, *Developing a Global Homology Analysis for Comparative Genomics*, Poster  
University of Central Oklahoma
- Apr 2016 **Showcase for Undergraduate Scholars**, *Processing RNA-seq Reads of Plants Infected with the Coffee Ringspot Virus*, Poster  
University of Kentucky
- Apr 2016 **UT-KBRIN Bioinformatics Summit**, *Processing RNA-seq Reads of Plants Infected with the Coffee Ringspot Virus*, Poster  
Cadiz, KY
- Apr 2015 **Showcase for Undergraduate Scholars**, *The Effect of Meditation on Performance*, Poster  
University of Kentucky

## Awards

### Grants & Fellowships

- 2022 **Conference Travel Grant**, *Rackham Graduate School*, (\$900)  
University of Michigan
- 2020-2021 **Rackham Interdisciplinary Workshop Grant**, (\$500)  
University of Michigan
- 2020 **Conference Travel Grant**, *Rackham Graduate School*, (\$800)  
University of Michigan
- 2019-2020 **Rackham Interdisciplinary Workshop Grant**, (\$500)  
University of Michigan
- 2019-2021 **NIH T32 Bioinformatics Training Program Fellow**  
University of Michigan
- Dec 2017 **Oswald Research & Creativity Award**, *UK Office of Undergraduate Research*, 2nd place in the Biological Sciences category (\$200)  
University of Kentucky
- May-Aug 2017 **Summer Research Grant**, *UK Office of Undergraduate Research*, (\$2,000)  
University of Kentucky
- 2014-2018 **Presidential Scholarship**, (out-of-state full tuition)  
University of Kentucky

### Honors

- May 2018 **Graduated Cum Laude with Departmental Honors in Biology**  
University of Kentucky
- May 2018 **Biology Undergraduate Research Award Nominee**  
University of Kentucky
- 2014-2018 **Lewis Honors College**  
University of Kentucky

---

## Continuing Education

- 27-30 Jan 2020 **Building Tidy Tools workshop at rstudio::conf**, *RStudio, PBC*  
San Francisco, CA
- 17-19 Dec 2019 **Winter School in Research Software Engineering**, *US Research Software Sustainability Institute*  
Seattle, WA
- 03-04 Jun 2019 **Software Carpentry Instructor Training**, *The Carpentries*  
University of Michigan
- 01-05 May 2019 **PyCon Education Summit & Conference**, *Python Software Foundation*  
Cleveland, OH
- 24-26 Apr 2019 **MICROBIOL 612.2**, *Riffomonas minimalR workshop*  
University of Michigan

---

## Publications

1. Armour CR, **Sovacool KL**, Close WL, Topçuoğlu BD, Wiens J, and Schloss PD, (2022). Streamlined implementation of a machine learning model to classify screen relevant neoplasia using reference-based OTU clustering. *bioRxiv*. <http://dx.doi.org/10.1101/2022.09.01.506299>.
2. Barron MR, **Sovacool KL**, Abernathy-Close L, Vendrov KC, Standke AK, Bergin IL, Schloss PD, and Young VB, (2022). Intestinal Inflammation Reversibly Alters the Microbiota to Drive Susceptibility to *Clostridioides Difficile* Colonization in a Mouse Model of Colitis. *mBio*, 0(0):e01904–22. <http://dx.doi.org/10.1128/mbio.01904-22>.
3. \*Lapp Z, **\*Sovacool KL**, Lesniak N, King D, Barnier C, Flickinger M, Krüger J, Armour CR, Lapp MM, Tallant J, Diao R, Oneka M, Tomkovich S, Anderson JM, Lucas SK, and Schloss PD, (2022). Developing and Deploying an Integrated Workshop Curriculum Teaching Computational Skills for Reproducible Research. *JOSE*. <http://dx.doi.org/10.21105/jose.00144>.
4. **Sovacool KL**, Westcott SL, Mumphrey MB, Dotson GA, and Schloss PD, (2022). OptiFit: An Improved Method for Fitting Amplicon Sequences to Existing OTUs. *mSphere*. <http://dx.doi.org/10.1128/msphere.00916-21>.
5. \*Duda M, **\*Sovacool KL**, Farzaneh N, Nguyen VK, Haynes SE, Falk H, Furman KL, Walker LA, Diao R, Oneka M, Drotos AC, Woloshin A, Dotson GA, Kriebel A, Meng L, Thiede SN, Lapp Z, and Wolford BN, (2021). Teaching Python for Data Science: Collaborative Development of a Modular & Interactive Curriculum. *JOSE*, 4(46):138. ISSN 2577-3569. <http://dx.doi.org/10.21105/jose.00138>.
6. Everson KM, Gray LN, Jones AG, Lawrence NM, Foley ME, **Sovacool KL**, Kratovil JD, Hotaling S, Hime PM, Storfer A, Parra-Olea G, Percino-Daniel R, Aguilar-Miguel X, O'Neill EM, Zambrano L, Shaffer HB, and Weisrock DW, (2021). Geography Is More Important than Life History in the Recent

Diversification of the Tiger Salamander Complex. PNAS, 118(17). ISSN 0027-8424, 1091-6490. <http://dx.doi.org/10.1073/pnas.2014719118>.

7. \*Topçuoğlu BD, \*Lapp Z, \***Sovacool KL**, Snitkin E, Wiens J, and Schloss PD, (2021). Mikropml: User-Friendly R Package for Supervised Machine Learning Pipelines. JOSS, 6(61):3073. ISSN 2475-9066. <http://dx.doi.org/10.21105/joss.03073>.
8. Hagan AK, Lesniak NA, Balunas MJ, Bishop L, Close WL, Doherty MD, Elmore AG, Flynn KJ, Hannigan GD, Koumpouras CC, Jenior ML, Kozik AJ, McBride K, Rifkin SB, Stough JMA, **Sovacool KL**, Sze MA, Tomkovich S, Topcuoglu BD, and Schloss PD, (2020). Ten Simple Rules to Increase Computational Skills among Biologists with Code Clubs. PLOS Computational Biology, 16(8):e1008119. ISSN 1553-7358. <http://dx.doi.org/10.1371/journal.pcbi.1008119>.
9. Goodin MM, Farman M, Inocencio H, Jang C, Jaromczyk JW, Moore N, and **Sovacool KL**, (2016). Processing RNA-Seq Data of Plants Infected with Coffee Ringspot Virus. In *Proceedings of the 15th Annual UT-KBRIN Bioinformatics Summit 2016: Cadiz, KY, USA. 8-10 April 2016*. BMC Bioinformatics. <http://dx.doi.org/10.1186/s12859-016-1154-y>.

\* Indicates co-first author.