**Reproducible Research in R: An Introductory Workshop on Reproducible and Collaborative Coding**

**Instructors:**

* **Marissa Dyck,** [marissadyck17@gmail.com](mailto:marissadyck17@gmail.com), University of Victoria
* **Erika Barthelmess,** [barthelmess@stlawu.edu](mailto:barthelmess@stlawu.edu), St. Lawrence University

**Workshop Summary:**

R is a free, open-source statistical programming language widely used in the scientific community. It has become the de facto software for use by ecologists (Auker and Barthelmess 2020) and its use is prominent among Mammalogists as well.

R has several important advantages over other statistical software environments. Key among these are features that lend themselves toward reproducibility in research (Marwick et al. 2018), which increasingly is required by journals and funding agencies. R easily integrates with other tools (e.g., GitHub, RMarkdown) to enhance the rapidity and ease with which results can be collaboratively conducted, presented, and reproduced and thus makes it a highly valuable software for scientists to learn. R markdown can be used to generate both static and interactive products which users can easily share with both R and non-R users to facilitate reproducible science and ease science communication. Git/GitHub applications allow users to collaborate efficiently using R and share reproducible code and data. R is also highly interdisciplinary and extensible, with over 10,000+ packages that allow for complex integrative analyses that span disciplines, from population genetics, to spatial analyses, to image analysis. R offers huge flexibility for working with Big Data and can interface with spreadsheets, databases, as well as other data formats on your computer or on the web. R also offers an excellent platform for the production of high-quality graphics, and using RMarkdown can yield documents, presentations, and apps. As a free, open-source software there are no financial barriers to its use, making it an easily accessible software for all scientists. Furthermore, there is a large and welcoming community of R users that can provide help through platforms such as StackOverflow and RStudio community. For these reasons, among many others, R is an essential, diverse, and flexible statistical software that all scientists should be learning to use.

In this hands-on workshop, we will:

* Introduce participants to the R Studio integrated development environment
* Introduce best practices for reproducible coding
* Learn to install and deploy R packages to increase R’s functionality
* Learn the basics of using R Markdown to generate dynamic documents.
* Explore options for importing and tidying data
* Learn the basics of visualizing data
* Explore extensions and introduce complex techniques
* Learn options for how to get help!
* Learn how to integrate R products with GitHub

**Relevance to the American Society of Mammalogists Membership:**

R can have a steep learning curve because users must write their own code to perform data management or analysis (Lawlor et al. 2022). While many Universities now offer beginner statistics courses in R, either at the undergraduate or graduate levels, these courses are often geared towards statistical theory and fail to introduce basics of coding. Thus, for many students, as well as more experienced scientists, getting started with R remains a formidable and daunting task, and incorporating R with other programs (e.g., GitHub, RMarkdown) can be a convoluted and lengthy process. Furthermore, most courses and online resources fail to introduce concepts of best coding practices which allow for code to be shared and reproduced by others, this is a crucial gap in our field in particular. In our experience, an introductory workshop in R can provide a critical foothold for many scientists who want to begin using R but may not know where to start. Additionally, this workshop will lay essential groundwork, by teaching best coding practices, introducing the available resources to help advance students beyond this course, and integrating with other programs to enhance shareability and reproducibility of code, which can be essential for early career mammalogists and can expedite their research goals. Further, working with others that also are new to the environment, rather than working in solitude, can ease the stress of learning R (Lawlor et al. 2022). Thus, the overarching goals of this workshop will be to assist beginner R users in becoming comfortable learning R and to help R users make the most of the features that facilitate scientific reproducibility.

**Justification of Speakers**:

Marissa Dyck is a Postdoctoral research fellow at University of Victoria where she works with large data sets, collaborating with several other academics, Alberta Government, and first nations all of whom have varying levels of familiarity with R and diverse needs for data and analyses. At the University of Victoria, Marissa also co-teaches an upper level (graduate and undergraduate) elective course, Analyzing Data for Boreal Mammal Ecology, and developed the R labs associated with this course. Marissa completed a PhD at Ohio University and has expertise with various statistical models and data tidying models and has developed a two-day workshop for R that will be adapted for ASM (<https://r-crash-course.github.io/>).

Erika Barthelmess is the Piskor Professor of Biology at St. Lawrence University, a liberal arts college in northern New York State. In addition to teaching Mammalogy, she teaches biostatistics (focused on using R) and is also co-founder and organizer of the “North Country R Users Group.” She has been teaching her colleagues across the sciences and social sciences as well as her undergraduate students in the use of R for data management and statistical analysis for the last decade or so. She is also co-author of a recent paper on teaching R in the undergraduate ecology classroom. She uses R for all sorts of data management and analysis applications which have recently been focused around large amounts of camera trap data.

**Preferred room set-up (classroom, circular tables, presentation mode, etc.):**

We would require AV equipment and a meeting room with tables that participants can use to support their laptops and lots of outlets for participants to keep their laptops charged.

**Preferred number (range) of participants:** [is there a maximum number of participants you can support]?

Max 30 – or however many the room can comfortably accommodate.

**Literature Cited:**

Auker, LA and EL Barthelmess. 2020. Teaching R in the undergraduate ecology classroom: approaches, lessons learned, and recommendations. Ecosphere <https://doi.org/10.1002/ecs2.3060>

Lawlor J, Banville F, Forero-Muñoz N-R, Hébert K, Martínez-Lanfranco JA, Rogy P, et al. (2022) Ten simple rules for teaching yourself R. PLoS Comput Biol 18(9): e1010372. https://doi.org/10.1371/journal.pcbi.1010372

Machado, FA and P Teta. 2020. Morphometric analysis of skull shape reveals unprecedented diversity of African Canidae. Journal of Mammalogy 101(2): 349-360.

Marwick, B, C. Boettiger and L. Mullen. 2018. Packaging data analytical work reproducibly using R (and friends). The American Statistician 72(1): 80-88