Harnessing open science practices to teach ecology and evolution using interactive tutorials

Jory Griffith*[1], Elizabeth Houghton*[2], Margaret A. Slein*[3], Maxime Fraser Franco*[4], Jhoan Chávez[5], Amy Forsythe[3], Victoria M. Glynn[6,7], Egor Katkov[1], Kirsten M. Palmier[8], Zihaohan Sang[9], Rolando Trejo-Pérez[10], Bryn Wiley[3], Jennifer Sunday[+1], Joey R. Bernhardt[+11]

Corresponding author: Maxime Fraser Franco, fraser_franco.maxime@courrier.uqam.ca

[1]Department of Biology, McGill University

[2]Department of Biology, University of British Columbia (Okanagan), Kelowna, BC, Canada

[3]Department of Zoology, University of British Columbia, Vancouver, BC, Canada

[4]Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC, Canada

[5]Department of Geography, Earth, and Environmental Sciences. University of Northern British Columbia, Prince George, BC, Canada.

[6]Redpath Museum and Department of Biology, McGill University, Montréal, QC, Canada

[7]Smithsonian Tropical Research Institute, Balboa, Republic of Panama

[8]Department of Biology, University of Regina, Regina, SK, Canada

[9]Department of Computer Science, University of Toronto (St.George), Toronto, ON, Canada

[10]Institut de Recherche en Biologie Végétale, Université de Montréal, Montréal, QC, Canada

[11]Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

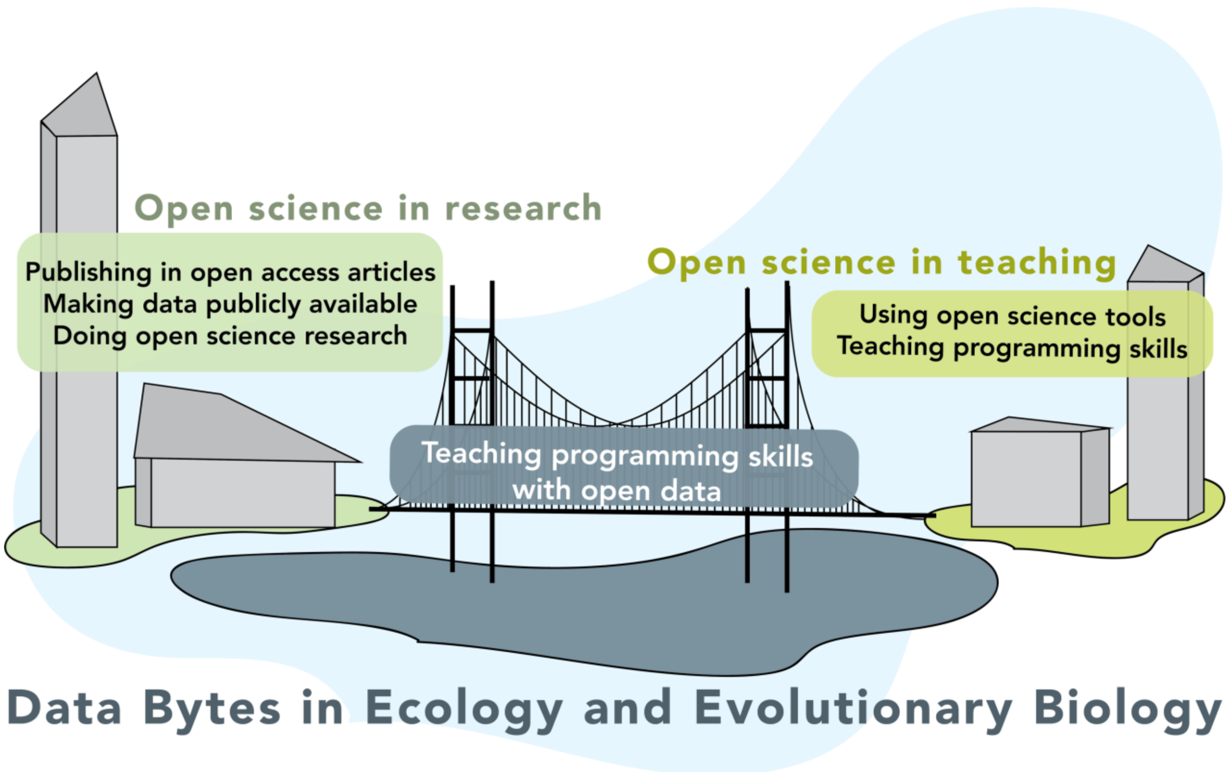
*JG, EH, MAS, and MFF should be considered joint first author.

[+]JS and JB should be considered joint senior author.

ABSTRACT

Open science skills are increasingly important for a career in ecology and evolution as efforts to make data and analyses publicly available and transparent continue to become more commonplace. However, open science skills are not typically taught in biology undergraduate programs. In learning core concepts in ecology and evolutionary biology (EEB), students must also gain skills in open science or they will miss out on opportunities to prepare for future careers. Core open science skills, like programming and practices that promote reproducibility and data sharing, can be taught to undergraduate students alongside core concepts in EEB. Yet, a major challenge in teaching open science skills and EEB concepts simultaneously is the high cognitive load associated with teaching multiple disparate concepts at the same time. One solution is to provide students with easily digestible, scaffolded, pre-formatted code in the form of vignettes and interactive tutorials. Here we present six open-source teaching tutorials for undergraduate students in EEB. These tutorials were developed through a graduate student based working group entitled *Data Bytes in Ecology and Evolutionary Biology*. These tutorials combine teaching data literacy and programming (using *R*) with analyzing publicly available data sets to teach fundamental ecological and evolutionary concepts and by doing so, introduce students to open science concepts and tools. Spanning a variety of EEB topics and skill levels, these tutorials serve as examples of how educators can integrate open science tools, programming, and data literacy into undergraduate teachings in topics of ecology and evolution.

Visual abstract



KEYWORDS

Active learning; ecology and evolutionary biology (EEB); online tutorials; open science; teaching; undergraduate education

INTRODUCTION

The open science movement represents a paradigm shift in the way we do science, emphasizing the sharing of data, analyses, and ideas to promote reproducibility, transparency, and collaboration during the scientific process (Rinaldi, 2014). However, practicing open science requires an additional set of skills, including familiarity with open source programming languages (e.g. *R*, *Python*, *Julia*), reproducible workflows, version control systems (e.g. Git), and data sharing platforms (e.g. GitHub, Open Science Framework, Figshare, etc.) (Hampton et al., 2015). As these skills become increasingly important for future participation in research and the workforce, universities have begun to incorporate them into their curricula (e.g. Hanna et al., 2021; Jekel et al., 2020; Pownall et al., 2023; Read et al., 2023). Initiatives that have developed materials to incorporate these skills with science curriculum content highlight open science practices as an essential part of the scientific process (e.g. CIEE-ICEE, 2023; Openscapes, 2023; Schmidt et al., 2016; Teal et al., 2015). However, teaching open science practices alongside discipline-specific scientific content in undergraduate biology courses can be challenging due to the need to teach programming skills and biology concepts at the same time. An additional challenge in teaching programming skills and open science in undergraduate programs is that instructors must account for varied programming skill levels, which requires a flexible teaching approach.

Teaching fundamental ecology and evolutionary biology (EEB) concepts alongside technical open science skills can be challenging but not insurmountable (Guzman et al., 2019). In EEB, instructors need to balance teaching EEB concepts as well as skills such as field sampling, programming, and data analysis (Auker and Barthelmess, 2020; Emery et al., 2021; Geange et al., 2021). Cohesively balancing multiple skills must be done effectively to ensure that students can incorporate multifaceted information easily, an issue addressed by “cognitive load theory” (Sweller et al., 2019, 1998). Cognitive load theory recognizes that learners can only process a limited amount of information at a time and works to mitigate the effort required to understand multiple concepts and commit them to long-term memory (Ou et al., 2022; Sweller et al., 2019). For example, the combination of teaching programming skills and biological concepts simultaneously can decrease student engagement with the course content due to high cognitive load (Engels et al., 2023; Guzman et al., 2019). One strategy to mitigate cognitive load is to use self-paced and scaffolded tutorials where structure is provided to assist with introducing new concepts and then gradually removed for students to apply these concepts themselves (Van De Pol et al., 2010). Using scaffolded interactive tutorials that cater to a variety of student proficiency levels (e.g. programming skills or statistical background) can improve student engagement and comprehension when learning programming in *R* (R Core Team, 2022) by allowing students to engage with material most comfortable to them, whether that means they directly write the associated code or adjust data inputs on an interactive application (Custer et al., 2021; Farrell and Carey, 2018; McGuire et al., 2022; Rissanen and Costello, 2023). A self-paced, scaffolded, multi-proficiency level approach to learning empowers students to challenge themselves in a productive way by tackling complex problems while gaining computational literacy and scientific skills (Dill-McFarland et al., 2021; Emery et al., 2021; Farrell and Carey, 2018).

Teaching ecological and evolutionary concepts to undergraduate students using publicly available datasets allows students to use real-world data to engage with material and answer scientific questions (Greengrove et al., 2020; O’Reilly et al., 2017; Styers et al., 2021). The ability to answer scientific questions using publicly available data is becoming increasingly important, as the use of this data in ecological research continues to rise (Michener and Jones, 2012). Previous work has integrated publicly available data to teach undergraduates and graduate students foundational concepts in their respective fields (e.g. oceanography via oceanographic time series [Greengrove et al., 2020], GIS/remote sensing via satellite data [Styers et al., 2021], earth and environmental science via lacustrine records [O'Reilly et al., 2017]). Students reported feeling more comfortable with ecological concepts and analyzing unstructured data after completing self-directed tutorials featuring real-world ecological data (O'Reilly et al., 2017; Styers et al., 2021). Additionally, students reported feeling a larger appreciation for publicly available data sets and an increased awareness of their potential uses, which could promote future engagement with open data practices (O'Reilly et al., 2017). Together, these studies demonstrate the benefits of integrating publicly available data with ecological and evolutionary concepts in undergraduate courses because the ability to work with these data is an integral skill for the next generation of scientists (LaDeau et al., 2017).

In recognizing the need to provide well-scaffolded tutorials to integrate programming skills with ecological and evolutionary concepts using publicly available data, we developed a set of ecology and evolution *R*-based teaching tutorials that aim to meet these needs. We developed six free and openly accessible teaching tutorials during the working group entitled *Data Bytes in Ecology and Evolutionary Biology* with the following goals: 1) Provide instructors with open-source teaching tutorials and a framework that can be modified to apply to their own courses, 2) Teach EEB concepts while mitigating cognitive load by allowing for students to primarily focus on exploring biological concepts with the option to engage with statistical and/or programming skills secondarily, 3) Use publicly available datasets to teach students how to explore and analyze real-world data. Here, we summarize the tutorials, expanding on two selected tutorials, and suggest how they can be integrated into undergraduate classrooms.

BACKGROUND: OPEN ACCESS TUTORIALS

During the first week of April 2022, a group of twelve graduate students, three post-doctoral facilitators, and six faculty members from universities across Canada participated in an online working group titled *Data Bytes in Ecology and Evolutionary Biology* hosted by the Living Data Project (LDP). The LDP is a nationwide initiative sponsored by the Canadian Institute for Ecology and Evolution/ Institut Canadien d'Écologie et d'Évolution (CIEE/ICEE) to teach and promote open science practices for graduate students through data rescue (i.e. identifying, preserving, and sharing valuable data and metadata at risk of loss; Bledsoe et al., 2022), education, and working groups. As a Canadian initiative, this working group was formed with the goal of developing openly accessible pedagogical tutorials for undergraduate students on foundational concepts in EEB using open source or rescued data sets collected in Canada (Figure 1).

We developed six tutorials using *R* with five data sets from Canadian ecosystems. Tutorials covered various evolutionary and ecological concepts (e.g., population dynamics, disease ecology, biodiversity indices, ecological specialization) and were written for students of a variety of proficiency levels (Table 1). We provided recommendations for undergraduate course levels (from first year to fourth year) these tutorials would best support based on the type of content featured in EEB course descriptions at Canadian universities (Table 1). To address multiple programming skill levels, tutorials also featured varying coding formats and complexity (e.g. Shiny [Chang et al., 2021], Rmarkdown [Allaire et al., 2023], learnR [Aden-Buie et al., 2023]) (Table 1). All tutorials can be accessed online at https://living-data-tutorials.github.io/website/.

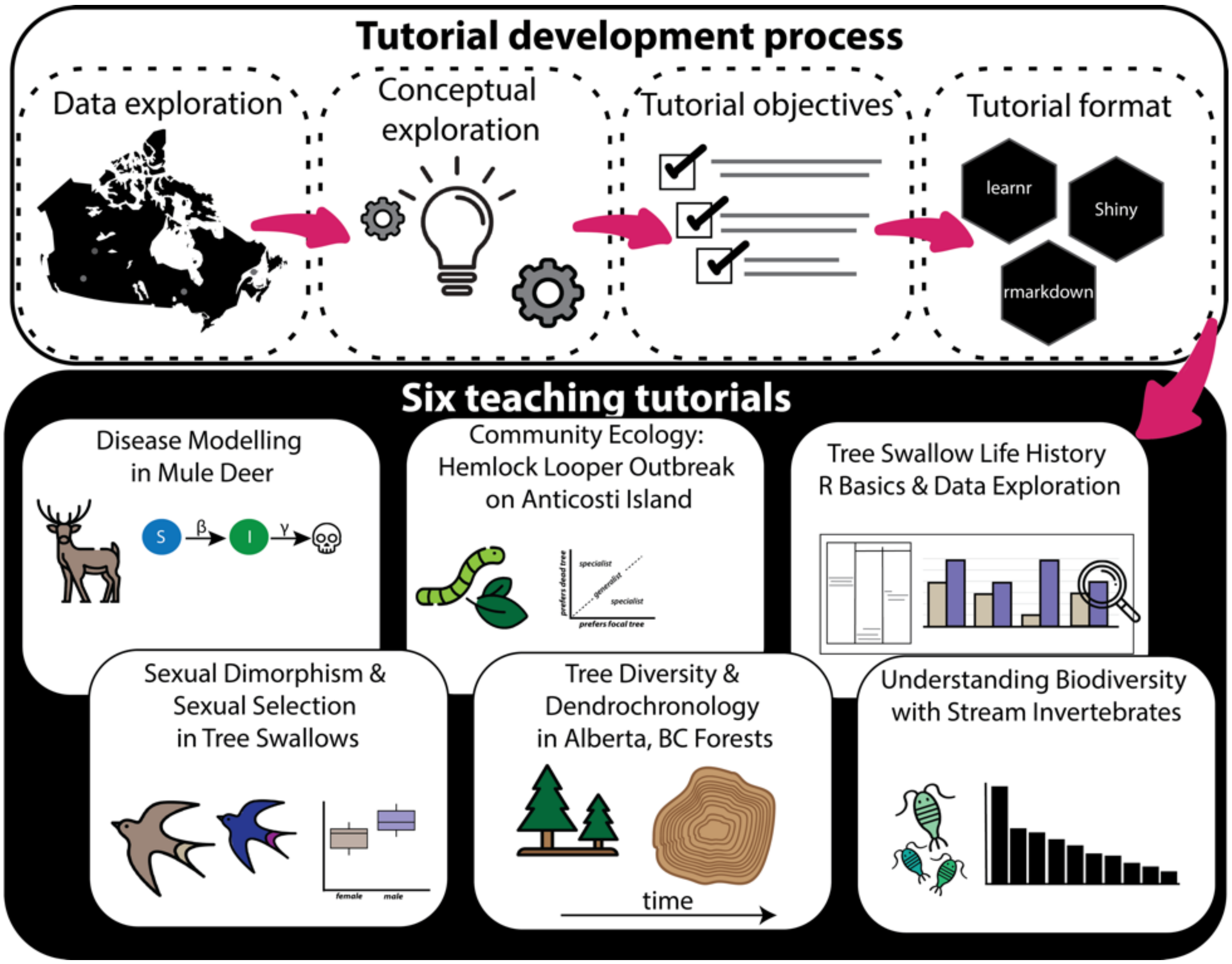


Figure 1. Schematic of tutorials produced from the *Data Bytes in Ecology and Evolutionary Biology* working group. The upper four panels represent the process of tutorial development (data exploration, conceptual exploration, tutorial objectives, and tutorial format) and the lower panels represent the six undergraduate teaching tutorials.

One goal of the tutorials was to mitigate cognitive load when teaching ecological and evolutionary concepts alongside *R* coding. For example, some tutorials offer the option to hide or show *R* code (e.g., the ‘Community Ecology: Hemlock Looper Outbreak on Anticosti Island’ tutorials vs. the ‘Disease Modelling in Mule Deer’ tutorial) and provide students with options to practice coding independently (for an example, see the tutorial titled ‘Tree Swallow Life History: R Basics & Data Exploration’). Other tutorials use a Shiny App to eliminate engagement with the code entirely in order to reduce cognitive load (e.g. in the tutorial titled ‘Disease Modelling in Mule Deer’). These tutorials also introduce a range of *R* tools and supporting packages (e.g., ‘codyn’ [Hallett et al., 2020], ‘tidyverse’ [Wickham et al., 2019], etc.).

Table 1. *Data Bytes in Ecology and Evolutionary Biology* undergraduate level teaching tutorials (https://living-data-tutorials.github.io/website/). Each tutorial was developed using a publicly-availably Canadian dataset to teach ecology and evolution concepts with specific learning objectives. Each tutorial has a suggested level(s) and undergraduate course(s).

Tutorial	Source dataset	Learning objectives	Ecological Topics	Format	Suggested audience	Suggested courses
A) Tree Swallow Life History: R Basics & Data Exploration	*Tree swallow nest box productivity data set from Long-Point, Ontario, Canada (1977-2014) (Diamond et al., 2021)	• Basic commands for handling ecological data in *R* • Graph data in appropriate formats • Observe trends and draw conclusions from figures	Data exploration	RMarkdown	First year	Introduction to Evolution and Ecology Organisms and the Environment
B) Community Ecology: Hemlock Looper Outbreak on Anticosti Island	*Secondary insects and fungi post-hemlock looper on Anticosti in 1973 (Government of Canada, 2021)	• Interpret community ecology plots • Differentiate between a generalist and specialist species using community ecology plots • Make data-driven management recommendations	Species dynamics, coexistence, competition, specialist *vs.* generalist	learnR Shiny	First & second year	Ecology and Evolution Organisms and the Environment
C) Sexual Dimorphism & Sexual Selection in Tree Swallows	*Tree swallow nest box productivity data set from Long-Point, Ontario, Canada (1977-2014) (Diamond et al., 2021)	• Identify sexual dimorphism • Explore how life history traits drive evolutionary trends • Explore population dynamics	Sexual dimorphism, life history, trade-offs	RMarkdown	Second year	Ecology and Evolution
D) Tree Diversity & Dendrochronology in Alberta, BC Forests	*Seasonal and annual dynamics of western Canadian boreal forest plant communities (Hesketh et al., 2021)	• Estimate and visualize plant diversity variation across space and environment • Estimate and visualize tree ring dynamics over time	Biodiversity, tree morphology and growth	learnR Shiny	Second & third year	Ecological Dynamics
E) Disease Modelling in Mule Deer	Chronic wasting disease in deer (history in Alberta) (Government of Alberta, 2023)	• Explore simple SI models and their application predicting disease spread • Explore the role of R_0 in determining disease spread • Learn how vaccination and knowledge of R_0 can be used to prevent the spread of disease	Disease ecology, population dynamics	Shiny	Third year	Population Biology Ecology Mathematical Biology
F) Understanding Biodiversity with Stream Invertebrates	*Turkey lakes watershed study (Government of Canada, 2020)	• Explore different ways biodiversity data can be summarized and compared • Create rank-abundance curves • Calculate biodiversity metrics	Rank-abundance curves, species richness, Shannon & Simpsons diversity index	RMarkdown	Third & fourth year	Ecology and Evolution Community or Aquatic Ecology

* datasets rescued through the Living Data Project (LDP) data rescue internship program

EXAMPLE APPLICATIONS FROM SELECT TUTORIALS

Here we highlight two of these six tutorials to show how we integrate biological concepts into practical examples with open-source data, mitigate cognitive load and accommodate multiple proficiency levels, and exemplify their potential application in the classroom (Table 1). We introduce: 1) 'Community Ecology: Hemlock Looper Outbreak on Anticosti Island' and 2) 'Understanding Biodiversity with Stream Invertebrates' tutorials with insets to demonstrate the variety of layouts in the tutorials (see Figure 1 and Figure 2). In each tutorial example, we include a tutorial overview, background of data collection for datasets featured, description of the tutorial activities, expansion on three tutorial learning objectives, and potential tutorial applications.

Tutorial Example 1. Community Ecology: Hemlock Looper Outbreak on Anticosti Island

Tutorial overview

In this tutorial, we present a broad introduction to consumer-resource interactions using a case study of the hemlock looper (*Lambdina fiscellaria*), a moth that feeds on trees (Table 1, row B). We used a publicly-available data set reporting tree damage during a hemlock looper outbreak in Anticosti Island, Québec, Canada in 1973 (Government of Canada, 2021). This tutorial aims to teach students how to extract ecological information from a data set, such as species abundances and food preferences, using basic data science exploration and visualization techniques (Figure 2a; see tutorial section one: 'Describe the tree community' and two: 'Estimate tree damage'). Additionally, the tutorial asks students to use their gained insights on the hemlock looper's ecology (Figure 2b; see tutorial section three: 'Evaluate the hemlock looper's preference') to make data-driven management recommendations for an outbreak in a new location: Metro Vancouver, British Columbia (Figure 2c; see tutorial section four: ‘Management recommendations’). Together, this tutorial holistically presents the interactions between insect pests and their plant prey, and their socioeconomic ramifications.

Data collection background

During the outbreak on Anticosti Island, researchers visually surveyed the degree of damage caused by the hemlock looper. The species of each tree affected by the hemlock looper was recorded in addition to classifying each individual tree by herbivory status. Trees were classified as “Good” if they did not show any signs of damage from the outbreak, “Dead” if the hemlock looper had attacked and killed the tree, or “Questionable” if the surveyors were unsure. The data are available on the Government of Canada’s open data portal (Government of Canada, 2021).

Description of the activity

Using the hemlock looper consumer-resource case study as a framework, the tutorial covers basic ecological concepts such as species' absolute and relative abundances, tree community composition after an outbreak, and ecological specialization (i.e. generalist and specialist species). These concepts are covered across four sections in the tutorial that address three distinct learning objectives (featured below). Throughout the tutorial, students are able to view well annotated *R* code used to explore the data, create plots, and calculate summary statistics. At the end of each section, students can test their knowledge by answering questionnaires based on the section’s material.

Learning Objectives

Objective 1 - Interpreting community ecology plots: In section one, students are guided to produce boxplots that summarize the abundances of tree species that were impacted by the outbreak (Figure 2a). In section two, students are asked to make bar plots visualizing the health status of the various tree species during the outbreak (Figure 2a). These exercises in data visualization reveal which tree species were impacted the most by the hemlock looper.

Objective 2 - Assess specialist vs generalist species status with data visualization: In section three, students are encouraged to develop hypotheses and predictions about the hemlock looper's tree preferences and whether it is a specialist or generalist species (i.e. whether it feeds on all trees equally or prefer one or two species) based on the plots they produced in the previous sections. Students are then introduced to a graphical method to test their hypothesis, which consists of a scatterplot with a 1:1 line displaying the relative abundance of trees in each species (x) and the proportion of trees of each species that are dead (y). The position of the points along the 1:1 line is used to determine whether the hemlock looper has preferences for certain species and conclude whether it is a specialist insect. Tree species that fall above the 1:1 line indicate that the hemlock looper attacks proportionately more of these trees even if their relative abundance is lower than other trees (Figure 2b).

Objective 3 - Data interpretation for data-driven management recommendations: In section four, students are encouraged to make data-driven management recommendations using inferences drawn from the data (Figure 2c). This section includes three videos focusing on a contemporary hemlock looper outbreak in Metro Vancouver, with a few guiding questions for students to reflect upon while they watch the clips. To provide flexibility for discussions on the socio-ecological ramifications of a hemlock looper outbreak, the tutorial asks students to reply to a prompt, which is to be submitted to their instructor. The tutorial closes with examples of big-picture ideas on the biological, social, and ethical factors to consider in management, and provides some main takeaways and additional references for students to pursue at their leisure.

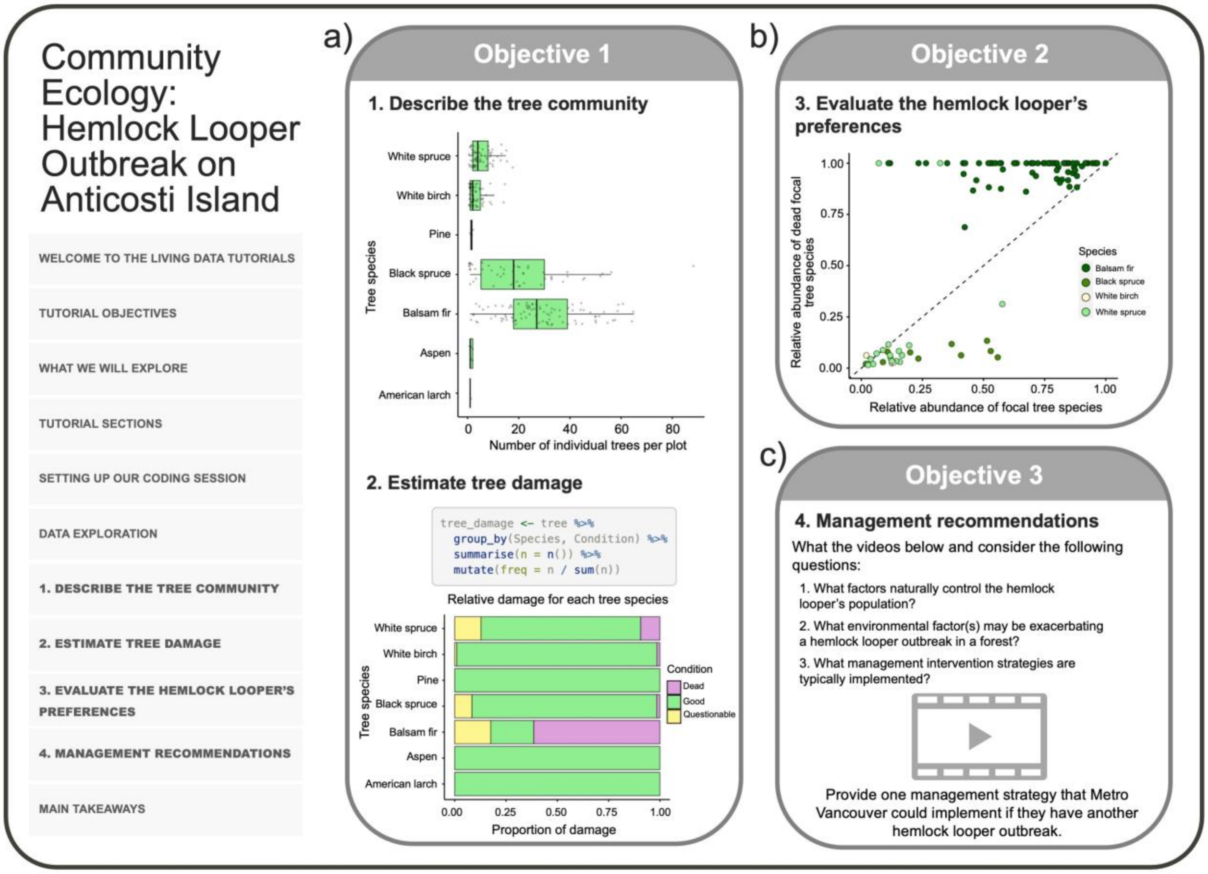


Figure 2. 'Community Ecology: Hemlock Looper Outbreak on Anticosti Island' tutorial content and examples of the material covered and knowledge testing questions. a) Students are taught to interpret community ecology plots by describing tree communities using box plots of the abundance of different species and using bar graphs to estimate the proportion of each tree species that was damaged by the hemlock looper (Objective 1). b) Students evaluate the hemlock looper's tree species preferences by plotting tree species' mortality against tree species' abundance (Objective 2). c) Students can make management recommendations in other

geographical regions based on information covered throughout the tutorial and suggested videos provided in this section (Objective 3).

Potential tutorial applications

This tutorial was designed for use in a first- or second-year ecology course/curriculum that has previously discussed species interactions, in particular consumer-resource interactions, and aims to provide students with a real-life, local example of the impact pests can have on ecological communities. The tutorial can also be used to expose students to data-driven hypothesis testing, as we use the data set to ask the question "does the hemlock looper have a preferred tree prey species?," allowing students to apply the scientific method in a more computationally-intensive environment as opposed to a lab environment. This tutorial provides students with the opportunity to develop their analytical skills as they learn to produce plots and interpret their data, and then apply these skills to make data-driven management recommendations. Overall, this tutorial aims to encourage undergraduate students to see consumer-resource dynamics from a more holistic lens, both from an ecological and a socio-ecological perspective, while introducing students to important open science practices such as open data and technical skills such as the use of open-source programming languages (*R*).

Tutorial Example 2. Understanding Biodiversity Change with Stream Invertebrates

Tutorial Overview

In this tutorial, we present an introduction to biodiversity indices and biodiversity change using a benthic stream invertebrates data set collected in the Turkey Lakes Watershed in Ontario, Canada. We use a publicly available dataset that measures how logging activities affected benthic invertebrate abundance and biodiversity in nearby streams (Government of Canada, 2020). This tutorial aims to teach students how to understand biodiversity (see tutorial section one: 'Why biodiversity matters' and section two: 'Types of data typically collected'), quantify biodiversity change (see tutorial section three: 'Real world biodiversity data' and section four: 'Metrics of biodiversity'), and compare biodiversity change across time and environmental conditions (see tutorial section five: 'Visualizing biodiversity'). Additionally, this tutorial teaches students how to calculate biodiversity indices via the 'codyn' *R* package (Hallett et al., 2020) and provides an activity for students to calculate biodiversity indices themselves in different logging areas (see tutorial section six: 'Measure and visualize biodiversity with codyn'). This tutorial presents ways to quantify and understand biodiversity with real world applications for how anthropogenic activities could influence biodiversity change.

Data collection background

In 1997, an experiment was conducted in Turkey lakes watershed where sites were logged with varying intensities (low, medium, and high) to test how logging affected stream ecosystems. From 1995 to 2009, scientists collected, identified and counted benthic invertebrates from streams near these logged areas around the Turkey Lakes watershed. Benthic invertebrates are small, often microscopic organisms that form an important link between aquatic and terrestrial habitats and can act as an indicator of ecosystem health. The data are available on the Government of Canada's open data portal (Government of Canada, 2020).

Description of the activity or structure of the tutorial

Using the benthic stream invertebrates from the Turkey Lakes Watershed as a case study, the tutorial covers ecological concepts such as species richness, evenness, and diversity, rank abundance curves, and comparing biodiversity metrics. These concepts are covered across six sections in the tutorial that address three distinct learning objectives (featured below). Throughout the tutorial, students are able to reveal well annotated *R* code used to explore the data, create plots, and calculate indices using the 'show code' button. At the end of the tutorial students can test their knowledge with an additional self-guided activity that applies the skills they learned to new questions and hypotheses using Turkey Lakes Watershed dataset.

Learning objectives

Objective 1 - Identify, calculate, and implement commonly used metrics to quantify biodiversity: In section one, students are given an introduction for how understanding biodiversity and biodiversity change is important for conservation decisions and resource management as well as potential ecological implications for disrupting ecosystem dynamics. In section two and three students are introduced to the raw data typically collected in order to quantify biodiversity as well as an introduction to the Turkey Lakes Watershed data. In section four students are then presented with a suite of metrics to quantify biodiversity change (e.g. species richness, Shannon-Wiener Index (H'), Simpson's Diversity Index (D), evenness (E)), using the Turkey Lakes Watershed data to do so (Figure 3a). These exercises in calculating biodiversity change with these different indices reveal annual changes in community composition and suggest potential effects of logging on these changes in community composition.

Objective 2 - Understand and create figures used to display biodiversity data: In section five, students are presented with ways to visualize changes in biodiversity using rank-abundance curves (RAC), using select data from the Turkey Lakes Watershed (Figure 3b). Students are then able to compare RACs to trace species distributions across study sites and logging intensity treatments. Visualizing the different ways in which RACs can be used to inform changes in ecosystem dynamics gives students exposure to the tools that could be used to inform important management and conservation decisions.

Objective 3 - Analyze biodiversity change in benthic invertebrates over multiple years and disturbance levels: In section six, students are introduced to an *R* package with functions built in to quantify changes in biodiversity (i.e. 'codyn' [Hallett et al., 2020]) and apply this analysis pipeline for an in-depth analysis of the Turkey Lakes Watershed dataset. Calculating a suite of biodiversity indices primes students for an additional activity at the end of the tutorial to explore additional in depth questions about the impact logging intensity on invertebrate biodiversity in the Turkey Lakes Watershed (Figure 3c).

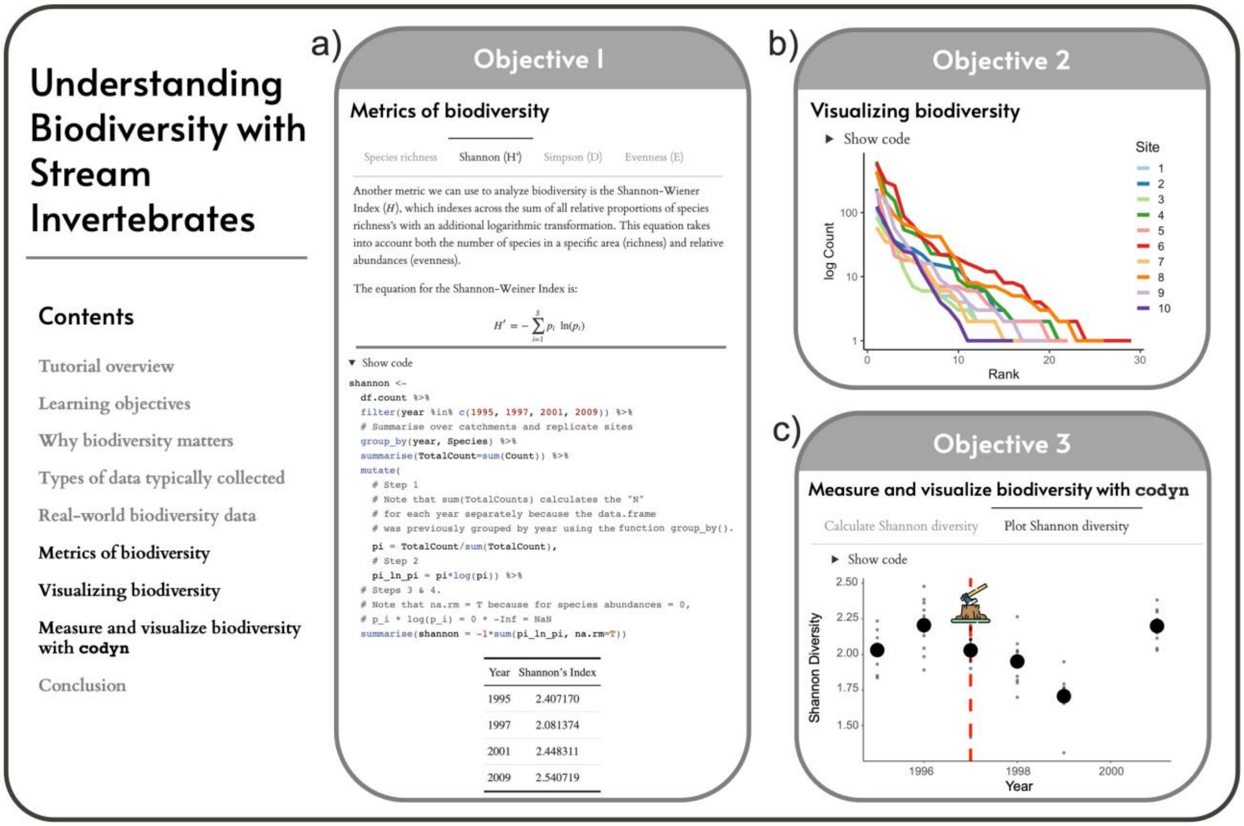


Figure 3. 'Understanding Biodiversity with Stream Invertebrates' tutorial content and examples of the material and exercises. Students are taught to a) identify, calculate, and implement commonly used metrics (richness, Shannon and Simpson diversity, and evenness) to quantify biodiversity (Objective 1), b) understand and create rank abundance curves to visualize biodiversity (Objective 2), and c) use the *R* package 'codyn' to calculate diversity metrics and plot over time to visualize biodiversity change following disturbance (Objective 3).

Potential tutorial applications

This tutorial was designed for use in a third- or fourth-year ecology course/curriculum that has previously discussed community ecology, in particular biodiversity indices, and aims to provide students with a real-life, local example of how biodiversity can change over time and with anthropogenic activities. The tutorial exposes students to how discipline-specific *R* packages like ‘codyn’ (Hallett et al., 2020) can be used to analyze publicly available datasets like the Turkey Lakes Watershed stream dataset. This tutorial teaches foundational biodiversity metrics used in community ecology, applies these indices to quantify important changes in community dynamics, and considers potential implications for management when it comes to logging impacts on biodiversity. Overall, this tutorial uses well-documented R Markdown files and a Shiny application to teach important skills in biodiversity science and encourages application of these skills by working with open data and open-source programming languages like *R*.

DISCUSSION AND CONCLUSION

We designed open educational resources in the form of online tutorials to support teaching undergraduates a variety of evolutionary and ecological concepts while integrating open science skills. The tutorials provide hands-on experience in data manipulation and statistical analysis, and introduce students to open science practices such as real-world open data and open-source software. The two highlighted tutorials demonstrate how these data can be used to guide decisions, both in pest and biodiversity management, presenting the impact and role of open science outside the classroom. These tutorials aim to increase student’s enthusiasm and confidence in using and interpreting open data and understanding essential EEB concepts, all while reducing cognitive load. The tutorials cater to a variety of student proficiency levels and provide a scaffolding for instructors to incorporate teaching programming into their own coursework.

To reduce cognitive load, these tutorials begin by presenting relevant information to improve comprehension of the biological concepts covered and clearly present tutorial learning objectives. These tutorials were designed in a way that breaks information down into smaller ‘chunks’. Breaking information into smaller ‘chunks’ can help improve absorption and retention of information which may improve deep comprehension and student application of knowledge (Kossen and Ooi, 2021). For example, the tutorials highlighted in this paper are formatted with different tabs or headings that allow the students to work through it in shorter sections. Additionally, many of these tutorials provide students with scaffolding where they first learn about a biological concept or a programming skill and then have the opportunity to apply this new information by answering questions or practice coding in an R Markdown file. It has been suggested that this scaffolding approach to teaching can help reduce the learner’s cognitive load and allows them to perform tasks that they may not have been able to perform otherwise (Myhill and Warren, 2005; Turner and Berkowitz, 2005; Van Merrienboer et al., 2003). Furthermore, these tutorials were designed to move from more simple to more complex tasks, use design elements to help draw attention to key elements (e.g. using bolded text), incorporate visuals and interactive plots to work to keep learners engaged, make non-essential content available but not the focus of the tutorial (e.g. by providing supplementary video or article insets or website links), and allow for students to engage with the programming elements to different degrees (e.g. view or hide code segments) (Clark et al., 2006; Van Merrienboer et al., 2003).

These tutorials were developed to cater to different programming and statistical proficiency levels often encountered among students in undergraduate EEB programs. Since students in EEB programs may have minimal background in programming and statistics, there can be a trade-off between time spent learning to code *versus* learning ecological concepts (Auker and Barthelmess, 2020). To address this, the tutorials vary in exposure to and interaction with programming from no coding requirements in a Shiny App, to options to view code when visualizing data or conducting statistical analyses, to downloading a R Markdown file and following prompts to practicing coding. The opportunity to engage with and learn programming at different levels varies both between and within many of the tutorials presented in this paper. Furthermore, the tutorials use a range of statistical concepts from simple summary statistics to more complex statistical modeling. Self-directed tutorials combining data science practices (i.e. data visualization and statistics) with concrete questions and working examples have been shown to help students integrate ecological and statistical concepts alongside programming skills (O'Reilly et al., 2017; Styers et al., 2021). By accommodating a variety of proficiency levels and conceptual backgrounds both within and between tutorials, we provide an entry point for a wide range of EEB undergraduate students to learn foundational concepts in EEB and open science with ease.

Skills such as the ability to manipulate and analyze large datasets with open-source software are becoming increasingly important within and outside of academia (Lai et al., 2019; Michener and Jones, 2012). However, designing assignments involving programming requires a significant time investment, such that incorporating programming in undergraduate curricula can be challenging for instructors (Auker and Barthelmess, 2020; Morrison et al., 2020). Preparing undergraduate students for research, including discussions about open science, can require additional planning and in-class time (Morrison et al., 2020). The tutorials presented here may help reduce this time investment for instructors by both providing user-friendly resources to directly incorporate into their curriculum as well as a scaffolding to expand on or develop additional tutorials that are shaped to the needs of individual courses.

In conclusion, the tutorials presented here teach both key EEB concepts via publicly available datasets and statistical analyses using *R* statistical software in a pre-packaged interactive way. This approach may help reduce cognitive load for students and the time needed for instructors to create course content. In doing so, these tutorials help to bridge the gap between open science research and open science teaching practices by using publicly available datasets and open-source statistical software.

DATA AVAILABILITY

Data used in the tutorial presented in this paper are available as follows: Tree Swallow Life History: R Basics & Data Exploration and Tree swallow and sexual dimorphism and sexual selection tutorial - https://doi.org/10.6084/m9.figshare.14156801.v1; Anticosti historical data tutorial - Government of Canada. Record ID: 9dda09b0-649f-4002-b207-7b204eb81cbb, https://open.canada.ca/data/en/dataset/9dda09b0-649f-4002-b207-7b204eb81cbb; Alberta trees tutorial - Borealis. https://doi.org/10.5683/SP3/PZCAVE; Chronic wasting disease tutorial - Government of Alberta: https://www.alberta.ca/chronic-wasting-disease-history-in-alberta.aspx#jumplinks-0; Understanding Biodiversity with Stream Invertebrates - Government of Canada. Record ID: f2ac0ae9-dd2f-4a70-b059-f8a49d9f5982, https://open.canada.ca/data/en/dataset/f2ac0ae9-dd2f-4a70-b059-f8a49d9f5982

COMPETING INTERESTS

The authors declare that there are no competing interests.

AUTHOR CONTRIBUTIONS

Jory Griffith: tutorial development (equal); writing – original draft (lead); writing – review and editing (lead). **Elizabeth Houghton:** tutorial development (equal); writing – original draft (lead); writing – review and editing (lead). **Margaret A. Slein:** tutorial development (equal); writing – original draft (lead); writing – review and editing (lead). **Maxime Fraser Franco:** tutorial development (equal); writing – original draft (lead); writing – review and editing (lead). **Jhoan Chávez:** tutorial development (equal); writing – review and editing (supporting). **Amy Forsythe:** tutorial development (equal); writing – original draft (supporting); writing – review and editing (supporting). **Victoria M. Glynn:** tutorial development (equal); writing – original draft (supporting); writing – review and editing (supporting). **Egor Katkov:** tutorial development (equal); writing – review and editing (supporting). **Kirsten Palmier:** tutorial development (equal); writing – original draft (supporting); writing – review and editing (supporting). **Zihaohan Sang:** tutorial development (equal); writing – review and editing (supporting). **Rolando Trejo:** tutorial development (equal); writing – review and editing (supporting). **Bryn Wiley:** tutorial development (equal); writing – original draft (supporting); writing – review and editing (supporting). **Jennifer Sunday:** conceptualization (equal); project administration (equal); supervision (equal); writing – review and editing (equal). **Joey Bernhardt:** conceptualization (equal); project administration (equal); supervision (equal); writing – review and editing (equal).

FUNDING

The Living Data Project is funded by a Collaborative Research and Training Experience (CREATE) grant to the Canadian Institute of Ecology and Evolution from the Natural Sciences and Engineering Research Council of Canada.

ACKNOWLEDGEMENTS

We are grateful to the Living Data Project postdoctoral researchers Drs. Gracielle Higino, Mike Lavender, and Ellen Bledsoe as well as Dr. Jason Pither, Dr. Melissa Guzman, and Jake Lawlor for their assistance with the *Data Bytes in Ecology and Evolutionary Biology* working group. We thank Nicole Knight, Nicole Moore, and William Ou for their constructive feedback on this manuscript.

REFERENCES

Aden-Buie, G., Schloerke, B., Allaire, J., Rossell Hayes, A., 2023. learnr: Interactive Tutorials for R. URL https://CRAN.R-project.org/package=learnr

Allaire, J.J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., Iannone, R., 2023. rmarkdown: Dynamic Documents for R.

Auker, L.A., Barthelmess, E.L., 2020. Teaching R in the undergraduate ecology classroom: approaches, lessons learned, and recommendations. Ecosphere 11, e03060. https://doi.org/10.1002/ecs2.3060

Bledsoe, E.K., Burant, J.B., Higino, G.T., Roche, D.G., Binning, S.A., Finlay, K., Pither, J., Pollock, L.S., Sunday, J.M., Srivastava, D.S., 2022. Data rescue: saving environmental data from extinction. Proceedings of the Royal Society B: Biological Sciences 289, 20220938. https://doi.org/10.1098/rspb.2022.0938

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. shiny: web application framework for R [WWW Document]. URL https://CRAN.R-project.org/package=shiny (accessed 8.1.21).

CIEE-ICEE, 2023. canadian institute of ecology and evolution | institut canadien d’écologie et d’évolution. Courses & Workshops. URL https://www.ciee-icee.ca/training.html (accessed 8.14.23).

Clark, R.C., Nguyen, F., Sweller, J., 2006. Efficiency in learning: evidence-based guidelines to manage cognitive load. Pfeiffer, San Francisco, CA, USA.

Custer, G.F., Diepen, L.T.A., Seeley, J., 2021. Student perceptions towards introductory lessons in R. Nat. Sci. Educ. 50, e20073. https://doi.org/10.1002/nse2.20073

Diamond, J., Bradley, D., Burant, J.B., 2021. Tree Swallow Nest Box Productivity Dataset from Long-Point, Ontario, Canada (1977-2014). URL https://figshare.com/articles/dataset/Tree_Swallow_Nest_Box_Productivity_Dataset_from_Long-Point_Ontario_Canada_1977-2014_/14156801/1 (accessed 4.15.22).

Dill-McFarland, K.A., König, S.G., Mazel, F., Oliver, D.C., McEwen, L.M., Hong, K.Y., Hallam, S.J., 2021. An integrated, modular approach to data science education in microbiology. PLoS Comput Biol 17, e1008661. https://doi.org/10.1371/journal.pcbi.1008661

Emery, N.C., Crispo, E., Supp, S.R., Farrell, K.J., Kerkhoff, A.J., Bledsoe, E.K., O'Donnell, K.L., McCall, A.C., Aiello-Lammens, M.E., 2021. Data science in undergraduate life science education: a need for instructor skills training. BioScience 71, 1274–1287. https://doi.org/10.1093/biosci/biab107

Engels, G., Grosjean, P., Artus, F., Numerical Ecology Department, C. and I.I., University of Mons, Avenue du Champ de Mars, 8, 7000 Mons, Belgium, Pedagogical Support and Quality Assurance Department, U. of M., Place du Parc, 20,700 Mons, Belgium, 2023. Teaching data science to students in biology using R, RStudio and Learnr: Analysis of three years data. Found. data sci. 5, 266–285. https://doi.org/10.3934/fods.2022022

Farrell, K.J., Carey, C.C., 2018. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. Ecol Evol 8, 7744–7751. https://doi.org/10.1002/ece3.4363

Geange, S.R., Oppen, J., Strydom, T., Boakye, M., Gauthier, T.J., Gya, R., Halbritter, A.H., Jessup, L.H., Middleton, S.L., Navarro, J., Pierfederici, M.E., Chacón-Labella, J., Cotner, S., Farfan-Rios, W., Maitner, B.S., Michaletz, S.T., Telford, R.J., Enquist, B.J., Vandvik, V., 2021. Next-generation field courses: Integrating open science and online learning. Ecol Evol 11, 3577–3587. https://doi.org/10.1002/ece3.7009

Government of Alberta, 2023. CWD surveillance - Annual summaries since 2009. URL https://www.alberta.ca/chronic-wasting-disease-history-in-alberta.aspx (accessed 5.27.22).

Government of Canada, 2021. Secondary insects and fungi post-hemlock looper on Anticosti in 1973. URL https://open.canada.ca/data/en/dataset/9dda09b0-649f-4002-b207-7b204eb81cbb (accessed 4.8.22).

Government of Canada, 2020. Turkey Lakes watershed study. URL https://open.canada.ca/data/en/dataset/f2ac0ae9-dd2f-4a70-b059-f8a49d9f5982 (accessed 4.7.22).

Greengrove, C., Lichtenwalner, S., Palevsky, H., Pfeiffer-Herbert, A., Severmann, S., Soule, D., Murphy, S., Smith, L., Yarincik, K., 2020. Using authentic data from NSF's Ocean Observatories Initiative in undergraduate teaching: an invitation. Oceanog 33, 62–73. https://doi.org/10.5670/oceanog.2020.103

Guzman, L.M., Pennell, M.W., Nikelski, E., Srivastava, D.S., 2019. Successful Integration of Data Science in Undergraduate Biostatistics Courses Using Cognitive Load Theory. LSE 18, ar49. https://doi.org/10.1187/cbe.19-02-0041.

Hallett, L., Avolio, M.L., Carroll, I.T., Jones, S.K., MacDonald, A.A.M., Flynn, D.F.B., Slaughter, P., Ripplinger, J., Collins, S.L., Gries, C., Jones, M.B., 2020. codyn: Community Dynamics Metrics. https://doi.org/10.5063/F1N877Z6. URL https://github.com/NCEAS/codyn

Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K., Mudge, J., Pourmokhtarian, A., Schildhauer, M.P., Woo, K.H., Zimmerman, N., 2015. The Tao of open science for ecology. Ecosphere 6, art120. https://doi.org/10.1890/ES14-00402.1

Hanna, S., Pither, J., Vis-Dunbar, M., 2021. Implementation of an Open Science instruction program for undergraduates. Data Intell. 3, 150–161. https://doi.org/10.1162/dint_a_00086

Hesketh, A., Loesberg, J., Bledsoe, E., Karst, J., Macdonald, 2021. Seasonal and annual dynamics of western Canadian boreal forest plant communities: a legacy dataset spanning four decades (V2 No. UNF:6:VXnwHUM0NAxB7w/Ix+Lp0g== [fileUNF]). Borealis.

Jekel, M., Fiedler, S., Allstadt Torras, R., Mischkowski, D., Dorrough, A.R., Glöckner, A., 2020. How to teach open science principles in the undergraduate curriculum—the Hagen Cumulative Science Project. Psychol Learn Teach 19, 91–106. https://doi.org/10.1177/1475725719868149

Kossen, C., Ooi, C.-Y., 2021. Trialling micro-learning design to increase engagement in online courses. AAOUJ 16, 299–310. https://doi.org/10.1108/AAOUJ-09-2021-0107

LaDeau, S.L., Han, B.A., Rosi-Marshall, E.J., Weathers, K.C., 2017. The next decade of big data in ecosystem science. Ecosystems 20, 274–283. https://doi.org/10.1007/s10021-016-0075-y

Lai, J., Lortie, C.J., Muenchen, R.A., Yang, J., Ma, K., 2019. Evaluating the popularity of R in ecology. Ecosphere 10, e02567. https://doi.org/10.1002/ecs2.2567

McGuire, R.M., Hayashi, K.T., Yan, X., Caritá Vaz, M., Cinoğlu, D., Cowen, M.C., Martínez-Blancas, A., Sullivan, L.L., Vazquez-Morales, S., Kandlikar, G.S., 2022. EcoEvoApps:

Interactive apps for theoretical models in ecology and evolutionary biology. Ecol Evol 12, e9556. https://doi.org/10.1002/ece3.9556

Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol Evol 27, 85–93. https://doi.org/10.1016/j.tree.2011.11.016

Morrison, M.E., Lom, B., Buffalari, D., Chase, L., Fernandes, J.J., McMurray, M.S., Stavnezer, A.J., 2020. Integrating research into the undergraduate curriculum: 2. Scaffolding research skills and transitioning toward independent research. J Undergrad Neurosci Educ. 19, A64–A74. https://doi.org/PMID: 33880093; PMCID: PMC8040851.

Myhill, D., Warren, P., 2005. Scaffolds or straitjackets? Critical moments in classroom discourse. Educ. Rev. 57, 55–69. https://doi.org/10.1080/0013191042000274187

Openscapes, 2023. Initiatives. URL https://openscapes.org/initiatives (accessed 8.10.23).

O'Reilly, C.M., Gougis, R.D., Klug, J.L., Carey, C.C., Richardson, D.C., Bader, N.E., Soule, D.C., Castendyk, D., Meixner, T., Stomberg, J., Weathers, K.C., Hunter, W., 2017. Using large data sets for open-ended inquiry in undergraduate science classrooms. BioScience 67, 1052–1061. https://doi.org/10.1093/biosci/bix118

Ou, W.J.-A., Henriques, G.J.B., Senthilnathan, A., Ke, P.-J., Grainger, T.N., Germain, R.M., 2022. Writing accessible theory in ecology and evolution: insights from cognitive load theory. BioScience 72, 300–313. https://doi.org/10.1093/biosci/biab133

Pownall, M., Azevedo, F., König, L.M., Slack, H.R., Evans, T.R., Flack, Z., Grinschgl, S., Elsherif, M.M., Gilligan-Lee, K.A., de Oliveira, C.M.F., Gjoneska, B., Kalandadze, T., Button, K., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Děchtěrenko, F., Alzahawi, S., Baker, B.J., Pittelkow, M.-M., Riedl, L., Schmidt, K., Pennington, C.R., Shaw, J.J., Lüke, T., Makel, M.C., Hartmann, H., Zaneva, M., Walker, D., Verheyen, S., Cox, D., Mattschey, J., Gallagher-Mitchell, T., Branney, P., Weisberg, Y., Izydorczak, K., Al-Hoorie, A.H., Creaven, A.-M., Stewart, S.L.K., Krautter, K., Matvienko-Sikar, K., Westwood, S.J., Arriaga, P., Liu, M., Baum, M.A., Wingen, T., Ross, R.M., O'Mahony, A., Bochynska, A., Jamieson, M., Tromp, M.V., Yeung, S.K., Vasilev, M.R., Gourdon-Kanhukamwe, A., Micheli, L., Konkol, M., Moreau, D., Bartlett, J.E., Clark, K., Brekelmans, G., Gkinopoulos, T., Tyler, S.L., Röer, J.P., Ilchovska, Z.G., Madan, C.R., Robertson, O., Iley, B.J., Guay, S., Sladekova, M., Sadhwani, S., null, null, 2023. Teaching open and reproducible scholarship: a critical review of the evidence base for current pedagogical methods and their outcomes. R. Soc. Open Sci. 10, 221255. https://doi.org/10.1098/rsos.221255

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Read, K.B., Lieffers, J., Massie, M., 2023. Integrating open science education into an undergraduate health professional research program. J Med Libr Assoc 110, 429–437. https://doi.org/10.5195/jmla.2022.1457

Rinaldi, A., 2014. Spinning the web of open science: Social networks for scientists and data sharing, together with open access, promise to change the way research is conducted and communicated. EMBO Reports 15, 342–346. https://doi.org/10.1002/embr.201438659

Rissanen, A., Costello, J.M., 2023. The effectiveness of interactive online tutorials in first-year large biology course. J. Appl. Res. High. Educ. 15, 632–649. https://doi.org/10.1108/JARHE-09-2020-0312

Schmidt, B., Orth, A., Franck, G., Kuchma, I., Knoth, P., Carvalho, J., 2016. Stepping up open science training for European research. Publications 4, 16. https://doi.org/10.3390/publications4020016

Styers, D.M., Schafer, J.L., Kolozsvary, M.B., Brubaker, K.M., Scanga, S.E., Anderson, L.J., Mitchell, J.J., Barnett, D., 2021. Developing a flexible learning activity on biodiversity and spatial scale concepts using open-access vegetation datasets from the National Ecological Observatory Network. Ecol. Evol. 11, 3660–3671. https://doi.org/10.1002/ece3.7385

Sweller, J., Van Merriënboer, J.J.G., Paas, F., 2019. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 31, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J.J.G., Paas, F.G.W.C., 1998. Cognitive architecture and instructional design. Educational Psychology Review 10, 251–296.

Teal, T.K., Cranston, K.A., Lapp, H., White, E., Wilson, G., Ram, K., Pawlik, A., 2015. Data Carpentry: Workshops to Increase Data Literacy for Researchers. IJDC 10, 135–143. https://doi.org/10.2218/ijdc.v10i1.351

Turner, V.D., Berkowitz, M.W., 2005. Scaffolding morality: Positioning a socio-cultural construct. New Ideas Psychol. 23, 174–184. https://doi.org/10.1016/j.newideapsych.2006.04.002

Van De Pol, J., Volman, M., Beishuizen, J., 2010. Scaffolding in teacher–student interaction: a decade of research. Educ Psychol Rev 22, 271–296. https://doi.org/10.1007/s10648-010-9127-6

Van Merrienboer, J.J.G., Kirschner, P.A., Kester, L., 2003. Taking the load off a learner's mind: Instructional design for complex learning. Educ. Psychol. 38, 5–13. https://doi.org/10.1207/S15326985EP3801_2

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686