

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

3

4 Jory Griffith*¹, Elizabeth Houghton*², Margaret A. Slein*³, Maxime Fraser Franco*⁴, Jhoan
5 Chávez⁵, Amy Forsythe³, Victoria M. Glynn^{6,7}, Egor Katkov¹, Kirsten M. Palmier⁸, Zihaohan
6 Sang⁹, Rolando Trejo-Pérez¹⁰, Bryn Wiley³, Jennifer Sunday⁺¹, Joey R. Bernhardt⁺¹¹

7

8 ¹Department of Biology, McGill University

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

10 ³Department of Zoology, University of British Columbia, Vancouver, BC, Canada

11 ⁴Département des Sciences Biologiques, Université du Québec à Montréal, Montréal, QC,
12 Canada

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

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

16 ⁷Smithsonian Tropical Research Institute, Balboa, Republic of Panama

17 ⁸Department of Biology, University of Regina, Regina, SK, Canada

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

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

20 ¹¹Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

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

22 ⁺JS and JB should be considered joint senior author

23

24 **ABSTRACT**

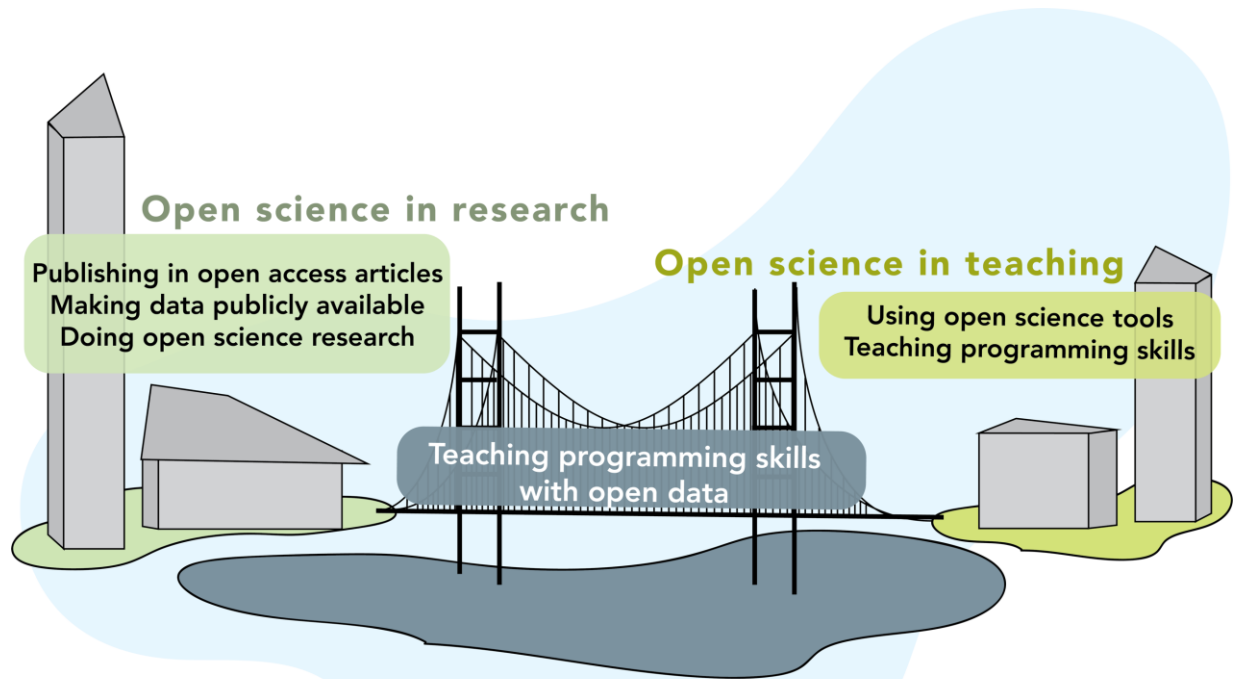
25 Open science skills are increasingly important for a career in ecology and evolution as
26 efforts to make data and analyses publicly available and transparent continue to become more
27 commonplace. However, open science skills are not typically taught in biology undergraduate
28 programs. In learning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

43

44

45

46



Data Bytes in Ecology and Evolutionary Biology

48

49

50 **KEYWORDS**

51 Active learning; ecology and evolutionary biology (EEB); online tutorials; open science;

52 teaching; undergraduate education

53

54

55

56

57

58

59

60

61 INTRODUCTION

62 The open science movement represents a paradigm shift in the way we do science,
63 emphasizing the sharing of data, analyses, and ideas to promote reproducibility, transparency,
64 and collaboration during the scientific process (Rinaldi, 2014). However, practicing open science
65 requires an additional set of skills, including familiarity with open source programming
66 languages (e.g. *R*, *Python*, *Julia*), reproducible workflows, version control systems (e.g. Git), and
67 data sharing platforms (e.g. GitHub, Open Science Framework, Figshare, etc.) (Hampton et al.,
68 2015). As these skills become increasingly important for future participation in research and the
69 workforce, universities have begun to incorporate them into their curricula (e.g. Hanna et al.,
70 2021; Jekel et al., 2020; Pownall et al., 2023; Read et al., 2023). Initiatives that have developed
71 materials to incorporate these skills with science curriculum content highlight open science
72 practices as an essential part of the scientific process (e.g. CIEE-ICEE, 2023; Openscapes, 2023;
73 Schmidt et al., 2016; Teal et al., 2015). However, teaching open science practices alongside
74 discipline-specific scientific content in undergraduate biology courses can be challenging due to
75 the need to teach programming skills and biology concepts at the same time. An additional
76 challenge in teaching programming skills and open science in undergraduate programs is that
77 instructors must account for varied programming skill levels, which requires a flexible teaching
78 approach.

79 Teaching fundamental ecology and evolutionary biology (EEB) concepts alongside
80 technical open science skills can be challenging but not insurmountable (Guzman et al., 2019). In
81 EEB, instructors need to balance teaching EEB concepts as well as skills such as field sampling,
82 programming, and data analysis (Auker and Barthelmess, 2020; Emery et al., 2021; Geange et
83 al., 2021). Cohesively balancing multiple skills must be done effectively to ensure that students

84 can incorporate multifaceted information easily, an issue addressed by “cognitive load theory”
85 (Sweller et al., 2019, 1998). Cognitive load theory recognizes that learners can only process a
86 limited amount of information at a time and works to mitigate the effort required to understand
87 multiple concepts and commit them to long-term memory (Ou et al., 2022; Sweller et al., 2019).
88 For example, the combination of teaching programming skills and biological concepts
89 simultaneously can decrease student engagement with the course content due to high cognitive
90 load (Engels et al., 2023; Guzman et al., 2019). One strategy to mitigate cognitive load is to use
91 self-paced and scaffolded tutorials where structure is provided to assist with introducing new
92 concepts and then gradually removed for students to apply these concepts themselves (Van De
93 Pol et al., 2010). Using scaffolded interactive tutorials that cater to a variety of student
94 proficiency levels (e.g. programming skills or statistical background) can improve student
95 engagement and comprehension when learning programming in *R* (R Core Team, 2022) by
96 allowing students to engage with material most comfortable to them, whether that means they
97 directly write the associated code or adjust data inputs on an interactive application (Custer et al.,
98 2021; Farrell and Carey, 2018; McGuire et al., 2022; Rissanen and Costello, 2023). A self-paced,
99 scaffolded, multi-proficiency level approach to learning empowers students to challenge
100 themselves in a productive way by tackling complex problems while gaining computational
101 literacy and scientific skills (Dill-McFarland et al., 2021; Emery et al., 2021; Farrell and Carey,
102 2018).

103 Teaching ecological and evolutionary concepts to undergraduate students using publicly
104 available datasets allows students to use real-world data to engage with material and answer
105 scientific questions (Greengrove et al., 2020; O’Reilly et al., 2017; Styers et al., 2021). The
106 ability to answer scientific questions using publicly available data is becoming increasingly

107 important, as the use of this data in ecological research continues to rise (Michener and Jones,
108 2012). Previous work has integrated publicly available data to teach undergraduates and graduate
109 students foundational concepts in their respective fields (e.g. oceanography via oceanographic
110 time series [Greengrove et al., 2020], GIS/remote sensing via satellite data [Styers et al., 2021],
111 earth and environmental science via lacustrine records [O'Reilly et al., 2017]). Students reported
112 feeling more comfortable with ecological concepts and analyzing unstructured data after
113 completing self-directed tutorials featuring real-world ecological data (O'Reilly et al., 2017;
114 Styers et al., 2021). Additionally, students reported feeling a larger appreciation for publicly
115 available data sets and an increased awareness of their potential uses, which could promote
116 future engagement with open data practices (O'Reilly et al., 2017). Together, these studies
117 demonstrate the benefits of integrating publicly available data with ecological and evolutionary
118 concepts in undergraduate courses because the ability to work with these data is an integral skill
119 for the next generation of scientists (LaDeau et al., 2017).

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

129 explore and analyze real-world data. Here, we summarize the tutorials, expanding on two
130 selected tutorials, and suggest how they can be integrated into undergraduate classrooms.

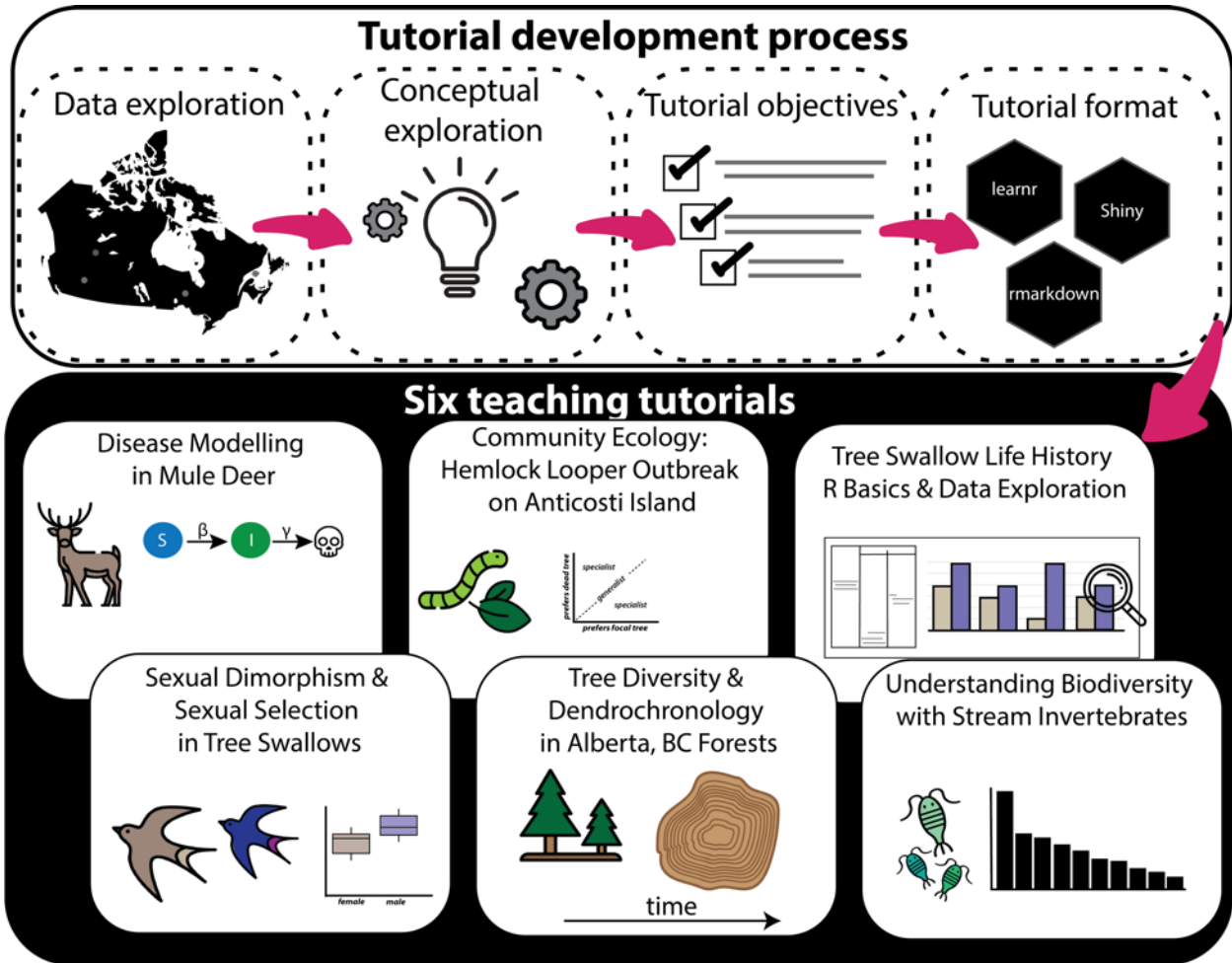
131

132 **BACKGROUND: OPEN ACCESS TUTORIALS**

133 During the first week of April 2022, a group of twelve graduate students, three post-
134 doctoral facilitators, and six faculty members from universities across Canada participated in an
135 online working group titled *Data Bytes in Ecology and Evolutionary Biology* hosted by the
136 Living Data Project (LDP). The LDP is a nationwide initiative sponsored by the Canadian
137 Institute for Ecology and Evolution/ Institut Canadien d'Écologie et d'Évolution (CIEE/ICEE) to
138 teach and promote open science practices for graduate students through data rescue (i.e.
139 identifying, preserving, and sharing valuable data and metadata at risk of loss; Bledsoe et al.,
140 2022), education, and working groups. As a Canadian initiative, this working group was formed
141 with the goal of developing openly accessible pedagogical tutorials for undergraduate students
142 on foundational concepts in EEB using open source or rescued data sets collected in Canada
143 (Figure 1).

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

152 2023]) (Table 1). All tutorials can be accessed online at [https://living-data-](https://living-data-tutorials.github.io/website/)
153 [tutorials.github.io/website/](https://living-data-tutorials.github.io/website/).



154

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

159

160 One goal of the tutorials was to mitigate cognitive load when teaching ecological and
161 evolutionary concepts alongside *R* coding. For example, some tutorials offer the option to hide or

162 show *R* code (e.g., the ‘Community Ecology: Hemlock Looper Outbreak on Anticosti Island’
163 tutorials vs. the ‘Disease Modelling in Mule Deer’ tutorial) and provide students with options to
164 practice coding independently (for an example, see the tutorial titled ‘Tree Swallow Life History:
165 R Basics & Data Exploration’). Other tutorials use a Shiny App to eliminate engagement with
166 the code entirely in order to reduce cognitive load (e.g. in the tutorial titled ‘Disease Modelling
167 in Mule Deer’). These tutorials also introduce a range of *R* tools and supporting packages (e.g.,
168 ‘codyn’ [Hallett et al., 2020], ‘tidyverse’ [Wickham et al., 2019], etc.).

169

170

171

172

173

174

175

176

177

178

179

180

181

182

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-available 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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

183 **EXAMPLE APPLICATIONS FROM SELECT TUTORIALS**

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

193

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

195 *Tutorial overview*

196 In this tutorial, we present a broad introduction to consumer-resource interactions using a
197 case study of the hemlock looper (*Lambdina fiscellaria*), a moth that feeds on trees (Table 1, row
198 B). We used a publicly-available data set reporting tree damage during a hemlock looper
199 outbreak in Anticosti Island, Québec, Canada in 1973 (Government of Canada, 2021). This
200 tutorial aims to teach students how to extract ecological information from a data set, such as
201 species abundances and food preferences, using basic data science exploration and visualization
202 techniques (Figure 2a; see tutorial section one: ‘Describe the tree community’ and two: ‘Estimate
203 tree damage’). Additionally, the tutorial asks students to use their gained insights on the hemlock
204 looper’s ecology (Figure 2b; see tutorial section three: ‘Evaluate the hemlock looper’s
205 preference’) to make data-driven management recommendations for an outbreak in a new

206 location: Metro Vancouver, British Columbia (Figure 2c; see tutorial section four: ‘Management
207 recommendations’). Together, this tutorial holistically presents the interactions between insect
208 pests and their plant prey, and their socioeconomic ramifications.

209

210 *Data collection background*

211 During the outbreak on Anticosti Island, researchers visually surveyed the degree of
212 damage caused by the hemlock looper. The species of each tree affected by the hemlock looper
213 was recorded in addition to classifying each individual tree by herbivory status. Trees were
214 classified as “Good” if they did not show any signs of damage from the outbreak, “Dead” if the
215 hemlock looper had attacked and killed the tree, or “Questionable” if the surveyors were unsure.
216 The data are available on the Government of Canada’s open data portal (Government of Canada,
217 2021).

218

219 *Description of the activity*

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

228

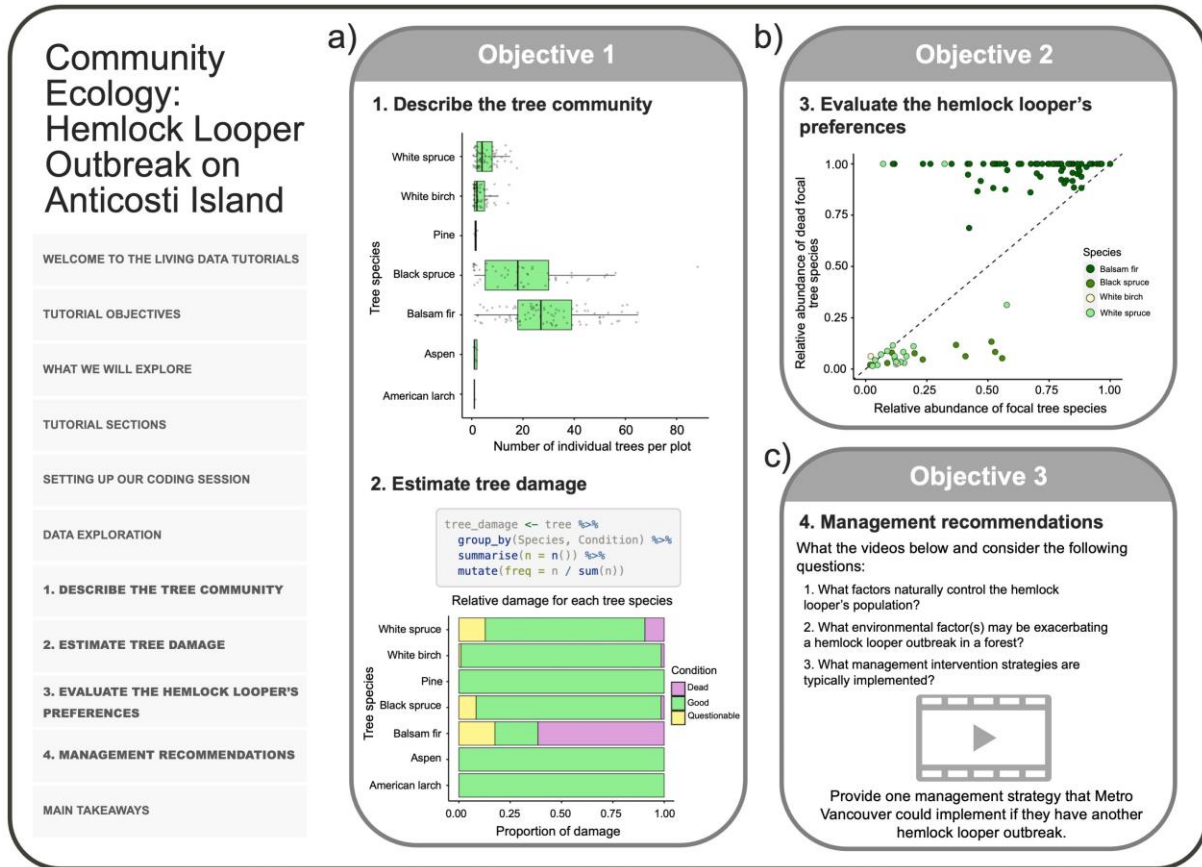
229 *Learning Objectives*

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

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

246 *Objective 3 - Data interpretation for data-driven management recommendations:* In
247 section four, students are encouraged to make data-driven management recommendations using
248 inferences drawn from the data (Figure 2c). This section includes three videos focusing on a
249 contemporary hemlock looper outbreak in Metro Vancouver, with a few guiding questions for
250 students to reflect upon while they watch the clips. To provide flexibility for discussions on the
251 socio-ecological ramifications of a hemlock looper outbreak, the tutorial asks students to reply to

252 a prompt, which is to be submitted to their instructor. The tutorial closes with examples of big-
 253 picture ideas on the biological, social, and ethical factors to consider in management, and
 254 provides some main takeaways and additional references for students to pursue at their leisure.



255

256 Figure 2. 'Community Ecology: Hemlock Looper Outbreak on Anticosti Island' tutorial content

257 and examples of the material covered and knowledge testing questions. a) Students are taught to

258 interpret community ecology plots by describing tree communities using box plots of the

259 abundance of different species and using bar graphs to estimate the proportion of each tree

260 species that was damaged by the hemlock looper (Objective 1). b) Students evaluate the hemlock

261 looper's tree species preferences by plotting tree species' mortality against tree species'

262 abundance (Objective 2). c) Students can make management recommendations in other

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

265

266 *Potential tutorial applications*

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

280

281 **Tutorial Example 2. Understanding Biodiversity Change with Stream Invertebrates**

282 *Tutorial Overview*

283 In this tutorial, we present an introduction to biodiversity indices and biodiversity change
284 using a benthic stream invertebrates data set collected in the Turkey Lakes Watershed in Ontario,
285 Canada. We use a publicly available dataset that measures how logging activities affected

286 benthic invertebrate abundance and biodiversity in nearby streams (Government of Canada,
287 2020). This tutorial aims to teach students how to understand biodiversity (see tutorial section
288 one: ‘Why biodiversity matters’ and section two: ‘Types of data typically collected’), quantify
289 biodiversity change (see tutorial section three: ‘Real world biodiversity data’ and section four:
290 ‘Metrics of biodiversity’), and compare biodiversity change across time and environmental
291 conditions (see tutorial section five: ‘Visualizing biodiversity’). Additionally, this tutorial
292 teaches students how to calculate biodiversity indices via the ‘codyn’ R package (Hallett et al.,
293 2020) and provides an activity for students to calculate biodiversity indices themselves in
294 different logging areas (see tutorial section six: ‘Measure and visualize biodiversity with
295 codyn’). This tutorial presents ways to quantify and understand biodiversity with real world
296 applications for how anthropogenic activities could influence biodiversity change.

297

298 *Data collection background*

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

306

307 *Description of the activity or structure of the tutorial*

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

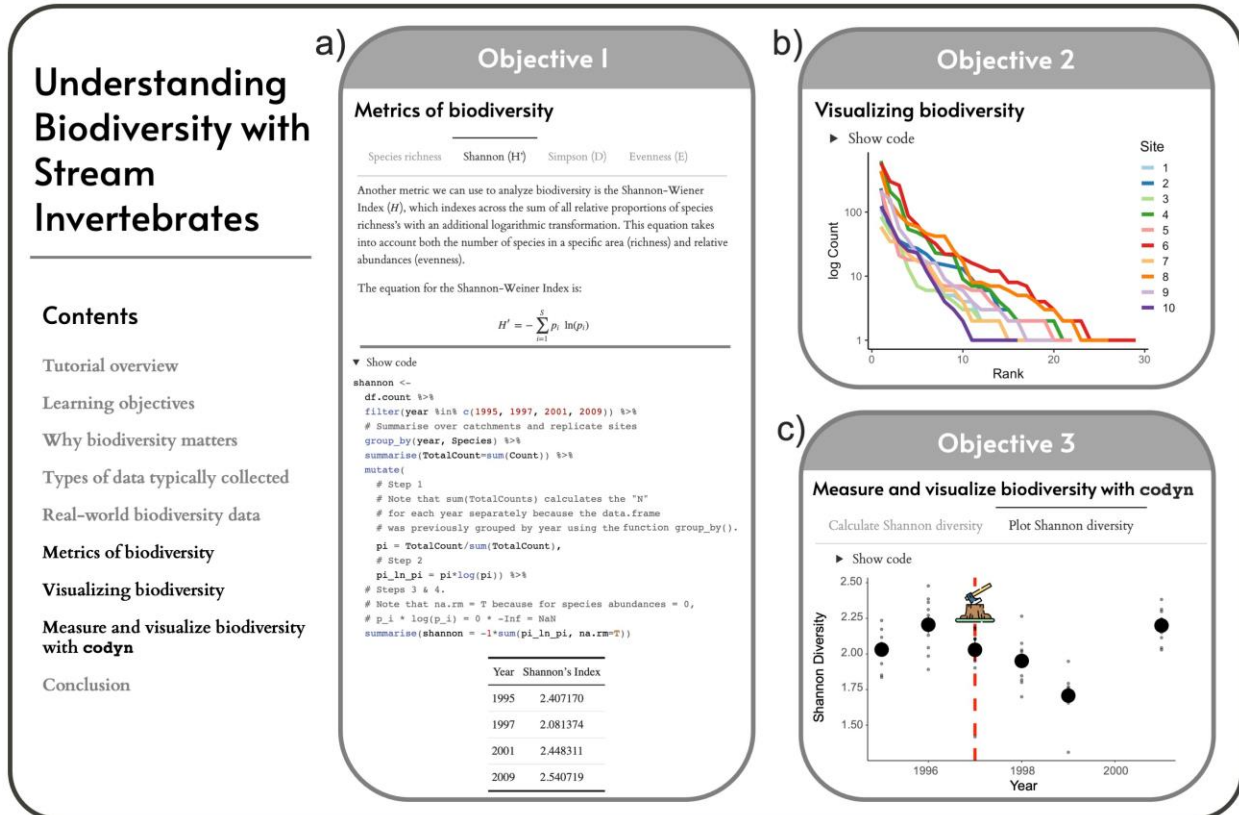
316

317 *Learning objectives*

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

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

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



344

345 Figure 3. 'Understanding Biodiversity with Stream Invertebrates' tutorial content and examples

346 of the material and exercises. Students are taught to a) identify, calculate, and implement

347 commonly used metrics (richness, Shannon and Simpson diversity, and evenness) to quantify

348 biodiversity (Objective 1), b) understand and create rank abundance curves to visualize

349 biodiversity (Objective 2), and c) use the R package 'codyn' to calculate diversity metrics and

350 plot over time to visualize biodiversity change following disturbance (Objective 3).

351

352 *Potential tutorial applications*

353 This tutorial was designed for use in a third- or fourth-year ecology course/curriculum

354 that has previously discussed community ecology, in particular biodiversity indices, and aims to

355 provide students with a real-life, local example of how biodiversity can change over time and

356 with anthropogenic activities. The tutorial exposes students to how discipline-specific R

357 packages like ‘codyn’ (Hallett et al., 2020) can be used to analyze publicly available datasets like
358 the Turkey Lakes Watershed stream dataset. This tutorial teaches foundational biodiversity
359 metrics used in community ecology, applies these indices to quantify important changes in
360 community dynamics, and considers potential implications for management when it comes to
361 logging impacts on biodiversity. Overall, this tutorial uses well-documented R Markdown files
362 and a Shiny application to teach important skills in biodiversity science and encourages
363 application of these skills by working with open data and open-source programming languages
364 like *R*.

365

366 **DISCUSSION AND CONCLUSION**

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

378 To reduce cognitive load, these tutorials begin by presenting relevant information to
379 improve comprehension of the biological concepts covered and clearly present tutorial learning

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

397 These tutorials were developed to cater to different programming and statistical
398 proficiency levels often encountered among students in undergraduate EEB programs. Since
399 students in EEB programs may have minimal background in programming and statistics, there
400 can be a trade-off between time spent learning to code *versus* learning ecological concepts
401 (Auker and Barthelmess, 2020). To address this, the tutorials vary in exposure to and interaction
402 with programming from no coding requirements in a Shiny App, to options to view code when

403 visualizing data or conducting statistical analyses, to downloading a R Markdown file and
404 following prompts to practicing coding. The opportunity to engage with and learn programming
405 at different levels varies both between and within many of the tutorials presented in this paper.
406 Furthermore, the tutorials use a range of statistical concepts from simple summary statistics to
407 more complex statistical modeling. Self-directed tutorials combining data science practices (i.e.
408 data visualization and statistics) with concrete questions and working examples have been shown
409 to help students integrate ecological and statistical concepts alongside programming skills
410 (O'Reilly et al., 2017; Styers et al., 2021). By accommodating a variety of proficiency levels and
411 conceptual backgrounds both within and between tutorials, we provide an entry point for a wide
412 range of EEB undergraduate students to learn foundational concepts in EEB and open science
413 with ease.

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

424 In conclusion, the tutorials presented here teach both key EEB concepts via publicly
425 available datasets and statistical analyses using *R* statistical software in a pre-packaged

426 interactive way. This approach may help reduce cognitive load for students and the time needed
427 for instructors to create course content. In doing so, these tutorials help to bridge the gap between
428 open science research and open science teaching practices by using publicly available datasets
429 and open-source statistical software.

430

431 **DATA AVAILABILITY**

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

442

443 **COMPETING INTERESTS**

444 The authors declare that there are no competing interests.

445

446 **AUTHOR CONTRIBUTIONS**

447 **Jory Griffith:** tutorial development (equal); writing – original draft (lead); writing –
448 review and editing (lead). **Elizabeth Houghton:** tutorial development (equal); writing – original

449 draft (lead); writing – review and editing (lead). **Margaret A. Slein:** tutorial development
450 (equal); writing – original draft (lead); writing – review and editing (lead). **Maxime Fraser**
451 **Franco:** tutorial development (equal); writing – original draft (lead); writing – review and
452 editing (lead). **Jhoan Chávez:** tutorial development (equal); writing – review and editing
453 (supporting). **Amy Forsythe:** tutorial development (equal); writing – original draft (supporting);
454 writing – review and editing (supporting). **Victoria M. Glynn:** tutorial development (equal);
455 writing – original draft (supporting); writing – review and editing (supporting). **Egor Katkov:**
456 tutorial development (equal); writing – review and editing (supporting). **Kirsten Palmier:**
457 tutorial development (equal); writing – original draft (supporting); writing – review and editing
458 (supporting). **Zihaohan Sang:** tutorial development (equal); writing – review and editing
459 (supporting). **Rolando Trejo:** tutorial development (equal); writing – review and editing
460 (supporting). **Bryn Wiley:** tutorial development (equal); writing – original draft (supporting);
461 writing – review and editing (supporting). **Jennifer Sunday:** conceptualization (equal); project
462 administration (equal); supervision (equal); writing – review and editing (equal). **Joey**
463 **Bernhardt:** conceptualization (equal); project administration (equal); supervision (equal);
464 writing – review and editing (equal).

465

466 **FUNDING**

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

470

471 **ACKNOWLEDGEMENTS**

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

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491 **REFERENCES**

- 492 Aden-Buie, G., Schloerke, B., Allaire, J., Rossell Hayes, A., 2023. learnr: Interactive Tutorials
493 for R. URL <https://CRAN.R-project.org/package=learnr>
- 494 Allaire, J.J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham,
495 H., Cheng, J., Chang, W., Iannone, R., 2023. rmarkdown: Dynamic Documents for R.
- 496 Auken, L.A., Barthelmess, E.L., 2020. Teaching R in the undergraduate ecology classroom:
497 approaches, lessons learned, and recommendations. *Ecosphere* 11, e03060.
498 <https://doi.org/10.1002/ecs2.3060>
- 499 Bledsoe, E.K., Burant, J.B., Higino, G.T., Roche, D.G., Binning, S.A., Finlay, K., Pither, J.,
500 Pollock, L.S., Sunday, J.M., Srivastava, D.S., 2022. Data rescue: saving environmental
501 data from extinction. *Proceedings of the Royal Society B: Biological Sciences* 289,
502 20220938. <https://doi.org/10.1098/rspb.2022.0938>
- 503 Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J.,
504 Dipert, A., Borges, B., 2021. shiny: web application framework for R [WWW
505 Document]. URL <https://CRAN.R-project.org/package=shiny> (accessed 8.1.21).
- 506 CIEE-ICEE, 2023. canadian institute of ecology and evolution | institut canadien d'écologie et
507 d'évolution. Courses & Workshops. URL <https://www.ciee-icee.ca/training.html>
508 (accessed 8.14.23).
- 509 Clark, R.C., Nguyen, F., Sweller, J., 2006. Efficiency in learning: evidence-based guidelines to
510 manage cognitive load. Pfeiffer, San Francisco, CA, USA.
- 511 Custer, G.F., Diepen, L.T.A., Seeley, J., 2021. Student perceptions towards introductory lessons
512 in R. *Nat. Sci. Educ.* 50, e20073. <https://doi.org/10.1002/nse2.20073>

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

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

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

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

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

533 Geange, S.R., Oppen, J., Strydom, T., Boakye, M., Gauthier, T.J., Gya, R., Halbritter, A.H.,
534 Jessup, L.H., Middleton, S.L., Navarro, J., Pierfederici, M.E., Chacón-Labela, J., Cotner,
535 S., Farfan-Rios, W., Maitner, B.S., Michaletz, S.T., Telford, R.J., Enquist, B.J., Vandvik,

536 V., 2021. Next-generation field courses: Integrating open science and online learning.
537 *Ecol Evol* 11, 3577–3587. <https://doi.org/10.1002/ece3.7009>

538 Government of Alberta, 2023. CWD surveillance - Annual summaries since 2009. URL
539 <https://www.alberta.ca/chronic-wasting-disease-history-in-alberta.aspx> (accessed
540 5.27.22).

541 Government of Canada, 2021. Secondary insects and fungi post-hemlock looper on Anticosti in
542 1973. URL [https://open.canada.ca/data/en/dataset/9dda09b0-649f-4002-b207-](https://open.canada.ca/data/en/dataset/9dda09b0-649f-4002-b207-7b204eb81cbb)
543 [7b204eb81cbb](https://open.canada.ca/data/en/dataset/9dda09b0-649f-4002-b207-7b204eb81cbb) (accessed 4.8.22).

544 Government of Canada, 2020. Turkey Lakes watershed study. URL
545 <https://open.canada.ca/data/en/dataset/f2ac0ae9-dd2f-4a70-b059-f8a49d9f5982> (accessed
546 4.7.22).

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

551 Guzman, L.M., Pennell, M.W., Nikelski, E., Srivastava, D.S., 2019. Successful Integration of
552 Data Science in Undergraduate Biostatistics Courses Using Cognitive Load Theory. *LSE*
553 18, ar49. <https://doi.org/10.1187/cbe.19-02-0041>.

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

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

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

565 Hesketh, A., Loesberg, J., Bledsoe, E., Karst, J., Macdonald, 2021. Seasonal and annual
566 dynamics of western Canadian boreal forest plant communities: a legacy dataset spanning
567 four decades (V2 No. UNF:6:VXnwhUM0NAXB7w/Ix+Lp0g== [fileUNF]). *Borealis*.

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

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

574 LaDeau, S.L., Han, B.A., Rosi-Marshall, E.J., Weathers, K.C., 2017. The next decade of big data
575 in ecosystem science. *Ecosystems* 20, 274–283. [https://doi.org/10.1007/s10021-016-](https://doi.org/10.1007/s10021-016-0075-y)
576 [0075-y](https://doi.org/10.1007/s10021-016-0075-y)

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

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

581 Interactive apps for theoretical models in ecology and evolutionary biology. *Ecol Evol*
582 12, e9556. <https://doi.org/10.1002/ece3.9556>

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

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

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

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

592 O'Reilly, C.M., Gougis, R.D., Klug, J.L., Carey, C.C., Richardson, D.C., Bader, N.E., Soule,
593 D.C., Castendyk, D., Meixner, T., Stomberg, J., Weathers, K.C., Hunter, W., 2017. Using
594 large data sets for open-ended inquiry in undergraduate science classrooms. *BioScience*
595 67, 1052–1061. <https://doi.org/10.1093/biosci/bix118>

596 Ou, W.J.-A., Henriques, G.J.B., Senthilnathan, A., Ke, P.-J., Grainger, T.N., Germain, R.M.,
597 2022. Writing accessible theory in ecology and evolution: insights from cognitive load
598 theory. *BioScience* 72, 300–313. <https://doi.org/10.1093/biosci/biab133>

599 Pownall, M., Azevedo, F., König, L.M., Slack, H.R., Evans, T.R., Flack, Z., Grinschgl, S.,
600 Elsherif, M.M., Gilligan-Lee, K.A., de Oliveira, C.M.F., Gjoneska, B., Kalandadze, T.,
601 Button, K., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Děchtěrenko, F.,
602 Alzahawi, S., Baker, B.J., Pittelkow, M.-M., Riedl, L., Schmidt, K., Pennington, C.R.,
603 Shaw, J.J., Lüke, T., Makel, M.C., Hartmann, H., Zaneva, M., Walker, D., Verheyen, S.,

604 Cox, D., Mattscheck, J., Gallagher-Mitchell, T., Branney, P., Weisberg, Y., Izydorczak,
605 K., Al-Hoorie, A.H., Creaven, A.-M., Stewart, S.L.K., Krautter, K., Matvienko-Sikar, K.,
606 Westwood, S.J., Arriaga, P., Liu, M., Baum, M.A., Wingen, T., Ross, R.M., O'Mahony,
607 A., Bochynska, A., Jamieson, M., Tromp, M.V., Yeung, S.K., Vasilev, M.R., Gourdon-
608 Kanhukamwe, A., Micheli, L., Konkol, M., Moreau, D., Bartlett, J.E., Clark, K.,
609 Brekelmans, G., Gkinopoulos, T., Tyler, S.L., Röer, J.P., Ilchovska, Z.G., Madan, C.R.,
610 Robertson, O., Iley, B.J., Guay, S., Sladekova, M., Sadhwani, S., null, null, 2023.
611 Teaching open and reproducible scholarship: a critical review of the evidence base for
612 current pedagogical methods and their outcomes. *R. Soc. Open Sci.* 10, 221255.
613 <https://doi.org/10.1098/rsos.221255>

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

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

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

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

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

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

633 Sweller, J., Van Merriënboer, J.J.G., Paas, F., 2019. Cognitive architecture and instructional
634 design: 20 years later. *Educ Psychol Rev* 31, 261–292. [https://doi.org/10.1007/s10648-](https://doi.org/10.1007/s10648-019-09465-5)
635 [019-09465-5](https://doi.org/10.1007/s10648-019-09465-5)

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

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

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

644 Van De Pol, J., Volman, M., Beishuizen, J., 2010. Scaffolding in teacher–student interaction: a
645 decade of research. *Educ Psychol Rev* 22, 271–296. [https://doi.org/10.1007/s10648-010-](https://doi.org/10.1007/s10648-010-9127-6)
646 [9127-6](https://doi.org/10.1007/s10648-010-9127-6)

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

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

655