The Riffomonas YouTube Channel: An Educational Resource To Foster Reproducible Research Practices

Patrick D. Schloss

Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT Methods for analyzing data in a reproducible manner are often viewed as impenetrable to scientists more familiar with laboratory research. The Riffomonas YouTube channel is committed to teaching these scientists and others how to engage in reproducible research using modern data science tools.

As high-throughput data generation becomes more common in microbiology and other disciplines, there is a significant need for laboratory scientists to develop data science skills (1). Unfortunately, traditional undergraduate and graduate biology training programs are often deficient in opportunities for scientists to develop the skills necessary to analyze large datasets in a reproducible and robust manner (2, 3). Numerous organizations seek to fill this void, including the Carpentries, Codecademy, and DataCamp (4). There are also numerous video tutorials available on YouTube. Although the content available through these platforms is popular, there has been a gap in content that emphasizes project-based learning.

The Riffomonas YouTube channel (https://www.youtube.com/c/RiffomonasProject) seeks to fill this gap. I started consistently posting videos at the beginning of the coronavirus disease 2019 (COVID-19) pandemic in April 2020. As of the end of November 2022, the channel had 11,327 subscribers and included 285 videos that had been viewed 635,947 times. The majority of these are 264 videos in the “Code Club” playlist (5) (Table 1). Other videos are related to a previously described tutorial series on reproducible research (6) and series in which reproducible research practices are used to address topical questions. Code Club videos are typically between 20 and 30 min long. The code that is developed in the videos is available through a website (https://riffomonas.org/code_club/) and the channel’s GitHub-hosted account (https://github.com/riffomonas).

The channel name, Riffomonas, comes from the concept of riffling, in which musical themes are adapted to achieve a similar sound, albeit perhaps in different contexts (6). This is to emphasize the value of reproducibility not only for recreating a set of results but for applying a method with a different data set (7). The channel covers topics related to reproducible data analysis practices, including R programming, data visualization, project organization, version control, command line programming, workflow tools, and scientific publishing (Table 1). Each video includes a brief introduction followed by me live coding to achieve a goal. I emphasize the use of live coding to modulate the rate of instruction and to show viewers my own coding practices. Observing an experienced analyst make mistakes normalizes some level of failure and demonstrates the strategies they can use to resolve their own mistakes. Viewers are encouraged to follow along with each video and to apply the new information to their own project.

Each video emphasizes a specific topic but includes other content that is selected to review topics covered in recent videos. Although videos can be watched individually, they often form a project arc (Table 1). For example, between July 2020 and July 2021, I formulated a research question, obtained and analyzed data to answer the
question, and wrote a paper that was published in *mSphere* (8). This series of 67 videos covered every topic from creating the initial directory on my computer to house the project files through reviewing the proofs of the published manuscript. Other project arcs have included visualizing microbiome data, modeling microbiome data using machine learning tools, analyzing the impacts of rarefying microbiome data, and other topics. Going forward, the Riffomonas channel will continue to post project-based content to help researchers develop their reproducible research skills.

**Data availability.** The Riffomonas YouTube channel is available at [https://www.youtube.com/c/RiffomonasProject](https://www.youtube.com/c/RiffomonasProject). The code developed in the Code Club videos is available at [https://riffomonas.org/code_club/](https://riffomonas.org/code_club/) and the channel’s GitHub-hosted account ([https://github.com/riffomonas](https://github.com/riffomonas)).

**ACKNOWLEDGMENTS**

I am grateful to the audience of the Riffomonas channel for their feedback on topics that I should cover in future episodes.

**REFERENCES**


---

**TABLE 1** Playlists found on the Riffomonas YouTube channel

<table>
<thead>
<tr>
<th>Topic</th>
<th>Playlist title</th>
<th>No. of videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data science</td>
<td>Data visualization with R’s tidyverse and allied packages</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Data manipulation within R’s tidyverse and other packages</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>Data analysis with base R</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Tools for reproducible data analysis</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Working at the command line</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Literate programming with R markdown</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Machine learning with mikropml R package</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Version control with Git and GitHub</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Scientific writing</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Project organization</td>
<td>3</td>
</tr>
<tr>
<td>Project-based series</td>
<td>All Code Club videos since 2 April 2020</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>Microbiome data analysis and visualization</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>ASV/OTU sensitivity and specificity analyses</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Visualizing COVID-19 vaccination attitudes</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Climate change data visualization</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Evaluating rarefaction and its alternatives</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Drought index visualization</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Reproducible research tutorial series</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Commemorating Juneteenth 2022 with a visualization</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2018 MLB All Star Break data analysis sprint</td>
<td>4</td>
</tr>
</tbody>
</table>

---

a Because most videos cover more than one topic, they are found in multiple playlists. Playlists and counts were current as of 1 December 2022. Playlists can be found under the Playlists tab at [https://www.youtube.com/c/RiffomonasProject](https://www.youtube.com/c/RiffomonasProject).

b ASV, amplicon sequence variant; OTU, operational taxonomic unit.