

Documentation and reproducibility with R and L^AT_EX

Monikrishna Roy,^{1*} Perry J. Williams,²

¹Department of Computer Science and Engineering, University of Nevada, Reno

²Department of Natural Resources and Environmental Science, University of Nevada, Reno

*To whom correspondence should be addressed; E-mail: monikrishna@nevada.unr.edu.

The reproducible research workshop covers basic usage of R, RStudio, LaTeX, and knitr to develop a workflow for research documentation and reproduction. The reproducible research workshop was developed from the book Reproducible Research with R and RStudio, by Christopher Gandrud (*I*).

Introduction

GRAD 778 introduces broadly applicable skills relevant for understanding the rapidly changing technological landscape of the 21st century. Further, skills developed in this course are directly supportive to Reno's research technology economy. This course builds human capital with research computing skills that are timely and necessary for current, local employment. Computational research is essential to understanding the complexities facing the modern workforce, and graduates of GRAD 788 will be required to handle sample datasets and learn to navigate and choose amongst various research computing skill sets.

This course is an overview of computational research as well as a skills-based introduction to programming and shell scripting for automating computational tasks. This course is offered for variable credit depending on the number of weekend workshops a student selects at the start

of the semester.

Anyone interested in using computational tools for research in any graduate department is encouraged to attend. Students will have the opportunity to work hands-on with large datasets with various real-world examples and write basic programs in more than one programming language. Classes will be part lecture, part open discussion, and part hands-on practice, held in mini-session format and lessons will be fully inclusive within each workshop.

Class materials in a mini-workshop format will be posted online in video format, and then discussion sessions held live on Saturdays. Students will select from the mini-workshops of interest. Mandatory attendance is required at course introduction during the first week (August 29th) of classes to establish proper equipment and enrollment.

To receive one graduate credit, a minimum of three-weekend workshops must be attended in full. To receive two graduate credits, a minimum of six-weekend workshops must be attended in full. To receive three graduate credits, a minimum of nine-weekend workshops must be attended in full. To receive course credit, students must display a working piece of code and pass a quiz at the end of each workshop.

“Documentation and reproducibility with R and L^AT_EX ” is a workshop of GRAD 778 courses.

1 Course Overview

1.1 Video Lectures

- Introduction
- Installation
- LaTeX
- knitr
- Conclusion

1.2 Overview of Files

1.2.1 Rep-Res-Workshop

The parent directory for the workshop is a file called Rep-Res-Workshop. Within this file are three files, including Analysis, Data, and Presentation.

Analysis The analysis file contains an R script file called **MainAnalysis.R** that was used to analyze avian species richness in the United States. It also contains a sub-file called Results-Figures, which includes .pdf output of figures from the **MainAnalysis.R** file.

Data The data file contains the avian species richness **data.csv** file used in the **MainAnalysis.R** script.

Presentation The presentation file contains seven sub-files. Four files contain examples used during the workshop. Two files contain exercises used during the workshop. One file contains the documents required to compile the workshop lecture pdf.

2 Materials

2.1 Installing the Main Software for the Workshop

Before coming to the workshop you should install the main software. All of the software programs covered in the workshop are open source and can be easily downloaded for free. They are available for Windows, Mac, and Linux operating systems. They should run well on most modern computers.

You should install R before installing RStudio. You can download the programs from the following websites:

- R (2)

- RStudio (3)

The download webpages for these programs have comprehensive information on how to install them, so please refer to those pages for more information. After installing R and RStudio you will want to install a number of user-written packages that are covered in this workshop. To install all of these user-written packages, run the following command in R:

```
required.packages=c("formatR", "here", "knitr", "Matrix")
lapply(required.packages, install.packages, character.only=TRUE)
lapply(required.packages, library, character.only=TRUE)
```

2.2 Installing markup language

To create \LaTeX documents you need to install a \TeX distribution. They are available for Windows, Mac, and Linux systems. They can be found at: <http://www.latex-project.org/ftp.html>. Please refer to that site for more installation information.

2.3 Session Info

```
toLatex(sessionInfo())
```

- R version 3.6.0 (2019-04-26),x86_64-apple-darwin15.6.
- Locale: en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-
- Running under:macOS Mojave 10.14.
- Matrix products: default
- BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
- LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: formatR 1.7, here 0.1, knitr 1.25, Matrix 1.2-
- Loaded via a namespace (and not attached): backports 1.1.4, compiler 3.6.0, evaluate 0.14, grid 3.6.0, highr 0.8, lattice 0.20-38, magrittr 1.5, rprojroot 1.3-2, stringi 1.4.3, stringr 1.4.0, tcltk 3.6.0, tools 3.6.0, xfun 0.

3 Example of Tables and Figures

Following are a few example handling the table and figures.

Tables. The Table 1 is showing head the data provided in the workshop directory.

rank	state	st	spp	area	temp	precip	statename
17	Alabama	AL	326	135765	62.8	58.3	alabama
46	Alaska	AK	269	1717854	26.6	22.5	alaska
3	Arizona	AZ	435	295254	60.3	13.6	arizona
22	Arkansas	AR	312	137732	60.4	50.6	arkansas
4	California	CA	415	423970	59.4	22.2	california

Table 1: Head of the dataset provided in the workshop.

Figures. Figures 1 & 2 are taken from the example figure of the workshop.



Figure 1: Final Doc Picture



Figure 2: Git Comments Example

4 Conclusion

The purpose of the workshop is developing a workflow to maximize reproducible research and research impact in computational science: managing data, computer code, and analyses output. Examples will include combining LaTeX, R, and knitr (4).

References

1. R. Glennie, *Journal of Agricultural, Biological and Environmental Statistics* **26**, 128 (2021).
2. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2020).

3. RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, PBC., Boston, MA (2020).
4. Y. Xie, *knitr: A General-Purpose Package for Dynamic Report Generation in R* (2021). R package version 1.36.

Acknowledgments

Thanks to the instructor for the wonderful workshop.

Supplementary materials

All code and lecture sheets will be found in the Canvas.