

Welcome

Welcome

This is the website for the Autumn 2014 course “Reproducible Research Methods” taught by Eric C. Anderson at NOAA’s Southwest Fisheries Science Center. The course meets on Tuesdays and Thursdays from 3:30 to 4:30 PM in Room 188 of the Fisheries Ecology Division. It runs from Oct 7 to December 18.

The goal of this course is for students to learn:

1. to write programs in the [R](#) language to manipulate and analyze data,
2. to integrate data analysis with report generation and article preparation using [knitr](#),
3. to work fluently within the [Rstudio](#) integrated development environment for R,
4. to use [git](#) version control software and [GitHub](#) to effectively manage source code, collaborate efficiently with other researchers, and neatly package their research.

By the end of the course, the hope is that students will have mastered strategies that allow them to use the above-listed, freely-available and open-source tools for conducting their research in a reproducible fashion. The ideal we will be striving for is to be able to start from a raw data set and then write a computer program that conducts all the cleaning, manipulation, and analysis of the data, and presentation of the results, in an automated fashion. Carrying out analysis and report-generation in this way carries a number of advantages to the researcher:

1. Newly-collected data can be integrated easily into your analysis.
2. If a mistake is found in one section of your analysis, it is not terribly onerous to correct it and then re-run all the downstream analyses.
3. Revising a manuscript to address referee comments can be done quickly.
4. Years after publication, the exact steps taken to analyze the data will still be available should anyone ask you how, exactly, you did an analysis!
5. If you have to conduct similar analyses and produce similar reports on a regular basis with new data each time, you might be able to do this readily by merely updating your data and then automatically producing the entire report.
6. If someone finds an error in your work, they can fix it and then easily show you exactly what they did to fix it.

Additionally, packaging one’s research in a reproducible fashion is beneficial to the research community. Others that would like to confirm your results can do so easily. If someone has concerns about *exactly* how a particular analysis was carried out, they can find the precise details in the code that you wrote to do it. Someone wanting to apply your methods to their own data can easily do so, and, finally, if we are all transparent and open about the methods that we use, then everyone can learn more quickly from their colleagues.

In many fields today, publication of research requires the submission of the original data to a publicly-available data repository. Currently, several journals require that all analyses be packaged in a clear and transparent fashion for easy reproduction of the results, and I predict that trend will continue until most, if not all, journals will require that data analyses be available in easily reproduced formats. This course will help scientists prepare themselves for this eventuality. In the process, you will probably find that conducting your research in a reproducible fashion helps you work more efficiently (and perhaps even more enjoyably!)

For more information about reproducible research you might be interested to read:

- [rep-res on wikipedia](#)

- [Reproducible Research with R and Rstudio](#) This is a book on the topic. You can buy it, and you can also build and compile it yourself from the source code at its [GitHub site](#) with the tools you will learn in this course.
- [A Report from a Stanford roundtable on reproducible research](#)

About this website

Nearly all of the content for this website was authored using [Rmarkdown](#). It was then rendered to html using [knitr](#) wrapped in a [jekyll](#) plugin written by Hadley Wickham. The lecture notes appear in outline format and also in slide format.

The slide format was rendered using the `slidy_presentation` format of the `render` function from the `rmarkdown` package. I build the site locally, and then push it up to [GitHub](#) where it is hosted for free.

I point this out to stress that a single format—the easily-learned and easily-read Rmarkdown markup language (see what the raw source for this page looks like [here](#))—can be rendered to many different formats. In fact, for demonstration, here is a [PDF](#) of this page automatically generated by converting Rmarkdown to LaTeX using [pandoc](#).

And, though I hope that you will have entirely weaned yourself from using Microsoft products for anything research-related by the time you are done with this course, it is worth knowing that you can render the same Rmarkdown source to [Word format](#). For final submission of an article, converting to MS-Word format might be expedient if you don't know LaTeX, and until the journals get their act together and start accepting markdown-formatted articles.

All of which just goes to demonstrate what can be done with the techniques you will learn in this course: in addition to learning how to program and analyze data in R, and collaborate through GitHub, you will learn easy, reproducible, portable ways of disseminating your research results in multiple output formats from a single input format.

Links

You should be able to navigate quickly anywhere in the site using the navbar at the top of the page (Once that is completed!). You can also find links here (eventually). To the left you will find links to headings on whatever page you are on.