Learning R for Wildlife

Jon Weissman [jweissma@gmail.com](mailto:jweissma@gmail.com)

4/28/2022

Table of Contents

This document compiles R resources that I recommend. R is not the most straightforward programming language, and different styles of R code can appear drastically different. Almost every style will look quite different from more conventional programming languages like C or Python.

Given that, it is easy to head down the wrong path in R. The resources I list here are largely chosen to avoid those pitfalls.

## Basics

Do you want to learn R?

If you have little to no coding experience and want to get started, or are new to R and data processing, this online book is great:

**Statistical Inference via Data Science** <https://moderndive.com/1-getting-started.html>

If you have are kinda familiar with code, or have even coded in old school base R extensively, but are new to tidy R or data science with R, [Hadley’s](https://github.com/hadley) book is the definitive guide to analyzing data with R:

**R for Data Science** <https://r4ds.had.co.nz/index.html>

Both of these resources are listed on the [tidyverse](https://www.tidyverse.org/learn/) website.

Once you’ve started using these, you’ll quickly need more reference info on tidyverse tools. Thankfully, these are extremely well documented with vignettes and examples. Many of the most commonly used data manipulation functions come from the package dplyr:

**Introduction to dplyr** <https://dplyr.tidyverse.org/index.html>

You’ll also want to plot data, and hence learn ggplot2:

**Learning ggplot2** <https://ggplot2.tidyverse.org/#learning-ggplot2>

If you want to get started plotting quickly, you can find what you need here:

**R Graphics Cookbook** <https://r-graphics.org/>

## Statistics

Do you want to get some basic descriptive statistics (mean, variance, etc)? Do you want to make some plots or fit a basic model?

**R for Data Science**, again, has you covered.

* [Ch 7 Exploratory Data Analysis](https://r4ds.had.co.nz/exploratory-data-analysis.html)
* [Ch 22 Modeling](https://r4ds.had.co.nz/model-intro.html)

Do you want to model your data for legit scientific use? Understand linear or logistic regression? Use (and actually understand) generalized additive models?

By far the best resource for statistical data analysis I’ve ever seen is [Cosma Shalizi’s](http://bactra.org/):

**Advanced Data Analysis from an Elementary Point of View** <https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ADAfaEPoV.pdf>

Most texts either focus on the principles and theory of statistics, or the nitty gritty of data analysis code. This book stays focussed on the basics while also demonstrating everything with actual R code and plots. The one downside is that it uses base R instead of tidyverse stuff. However, once you are familiar with tidyverse stuff it is not hard to go between the two. The largest difference is in the base plot versus the tidy ggplot.

Some of the content in this book is probably not of interest. I certainly haven’t made it through the whole thing. So I’ll highlight some sections I’ve found particularly useful here:

* 1.1 *Regression: Predicting and Relating Quantitative Features*
* 2.4 *Linear Regression Is Not the Philosopher’s Stone*
* 3 *Model Evaluation*
* 8 *Additive Models*
* 12 *Generalized Linear Models and Generalized Additive Models*

I think if you were to work through that whole book you would have a better statistics education than the majority of professional data scientists.

## Spatial Data

Do you want to make maps? Do spatial statistics? Explore movements?

Spatial data comes in two forms: Vector and Raster.

* Vector: This is data associated with precise coordinates. Think points, lines, gps fixes, roads.
* Raster: This is gridded data that is mapped to a spatial surface. The data is in images, and each pixel of that image covers a general area rather than precise coordinates. Think satellite imagery or elevation models.

For vector data, you will want to use sf:

**Simple Features for R** <https://r-spatial.github.io/sf/>

This is an R package, not a book. However, like many good R packages, it has excellent documentation in the *Articles* or *vignettes*.

sf is at the heart of all the code I’ve written to process telemetry data. I can’t recommend it enough, and [Edzer](https://github.com/edzer) is the most responsive developer I’ve encountered.

For more info on how to use sf:

**Spatial Analysis with R** <https://chiajung-yeh.github.io/Spatial-Analysis/>

This text by [Yeh Chia Jung](https://github.com/chiajung-yeh) looks nice. It’s basically and in-depth how to on using the sf package. It illustrates how to use ggplot in combination with spatial data, and also explains what goes into making data spatial.

For raster data, the corresponding package is stars:

**Spatiotemporal Arrays: Raster and Vector Datacubes** (<https://r-spatial.github.io/stars/>)

However, this is a younger package and has a strong competitor in [terra](https://rspatial.org/terra/pkg/index.html), which I haven’t used but I know is popular.

More info:

The underlying libraries driving all of these spatial data tools come from [Open Geospatial Consortium](https://www.osgeo.org/partners/ogc/). These tools, like [GDAL](https://gdal.org/) and [PROJ](https://proj.org/) also power software like [QGIS](https://www.qgis.org/en/site/) and [PostGIS](https://postgis.net/).

The good thing about that is that it’s easy to exchange data between any of these tools, and often an analysis done in one can be easily replicated in another. For example, I’ve often done something using the QGIS graphical interface, and then once I’ve figured out what I wanted, copied the steps over into R or Python using the same underlying GDAL library.

Unfortunately, ESRI does it’s own thing and often doesn’t play well with OGC tools.

## Advanced R topics

At some point writing tidy R code you are going to get frustrated and feel limited with what you can do. Most likely you will then start trying to write Python using R, with for loops and such. This can work but will also get really messy, and if you choose this path, you will regret ever using R because things would have been easier and faster had you started with Python.

Fortunately there is another path, and once again Hadley is there to guide us:

**Advanced R** <https://adv-r.hadley.nz/>

## RMarkdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

For more info on how to write R Markdown and do really fancy stuff with it:

**R Markdown: The Definitive Guide** <https://bookdown.org/yihui/rmarkdown/>

Source: <https://github.com/rgzn/r_guides>