The Journal of Open Source Software

lightr: import spectral data and metadata in R

Hugo Gruson[1], **Thomas White**[2], **and Rafael Maia**[3]

1 CEFE, University Montpellier, CNRS, University Paul Valery Montpellier 3, EPHE, IRD, Montpellier, France **2** School of Life and Environmental Sciences, University of Sidney, Sidney, NSW, Australia **3** Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY

DOI: 10.21105/joss.01857

Software

- Review
- Repository
- Archive

Submitted: 31 October 2019
Published: 01 November 2019

Summary

Living organisms wildly differ in their ability to see colours (Osorio & Vorobyev, 2008). For this reason, colour science relies on the use of objective measurements of reflectance, transmittance, or aborbance spectra rather than human vision (Bennett, Cuthill, & Norris, 1994; Cuthill, Bennett, Partridge, & Maier, 1999; Eaton, 2005). These spectra are then used in vision models that allow scientists to predict how a given object is seen through the eyes of a given species (e.g., how a male bird is seen by a potential mate). This is the basis of all studies in for example the study of the evolution of colours of animals and plants as communication signals

Spectrometers record the amount of captured photons at different wavelengths (usually between 300-700 nm for colour science, as many species are sensitive to ultraviolet radiations). However, there is no standard file format for spectrometry data and different scientific instrumentation companies use wildly different formats to store spectral data. This use of non-standard file formats jeopardises scientific reproducibility (Peng, 2009) as other researchers might not have the (paid) tools to open these files, and it makes us dependent on a third-party which might vanish anytime, leaving a trove of scientific data impossible to access. Vendors' proprietary software sometimes have an option to convert those formats into human readable files such as `csv` but such software are often expensive and they discard most metadata in the process. Yet, those metadata are critical to ensure reproducibility of the measurements, and ultimately of the scientific findings (White, Zeil, & Kemp, 2015).

In this article, we present `lightr`, an R package that aims at offering a unified user-friendly interface for users to read reflectance, transmittance, and absorbance spectra files from various formats in a single line of code. Additionally, it provides for the first time a fully free and open source solution to read proprietary spectra file formats on all operating systems.

`lightr` started as a fork from the popular R package `pavo`, which provides a large suite of colour analysis tools (Maia, Eliason, Bitton, Doucet, & Shawkey, 2013; Maia, Gruson, Endler, & White, 2019).

Package design

`lightr` has been designed to provide two levels on the complexity / customability trade-off:

- Spectral data and metadata for each file format are extracted using specialized parsers. Parsers are also aliased with many different names so that users can often use `lr_parse_$extension()` where `$extension` is the file extension of the file to parse. For convenience, we also provide a generic fallback, named `lr_parse_generic()` that works for many "simple" formats, often derived from `csv` or `tsv`. Specialized parsers

should usually be preferred to `lr_parse_generic()` because `lr_parse_generic()` is not able to parse metadata.

- Because spectrometers store each measurement in a separate file, the number of files for a single study can quickly increase. To ensure easy and efficient processing of those files, `lightr` also provides three high-levels functions that can recursively find files and process them with a parallelized loop using the `parallel` R package: `lr_get_spec()` and `lr_get_metadata()`, which import respectively spectral data and metadata as `data.frame` in R, as well as `lr_convert_tocsv()`, which converts all spectra files in a given folder as `csv`, with the same filename (minus the file extension).

```
library(lightr)
lr_convert_tocsv(where = "yourfolder", ext = "ProcSpec")
```

Recommended workflow

As mentioned earlier, proprietary spectrometry software can also export spectral data into a human-readable format (usually a kind of tabulation separated values, or `tsv`, with a complex header). `lightr` can read files generated by this export step. We however **do not recommend you use the software's built-in export function**, because it will apply possibly unwanted transformation to your data (interpolation and subsetting) and may discard important metadata.

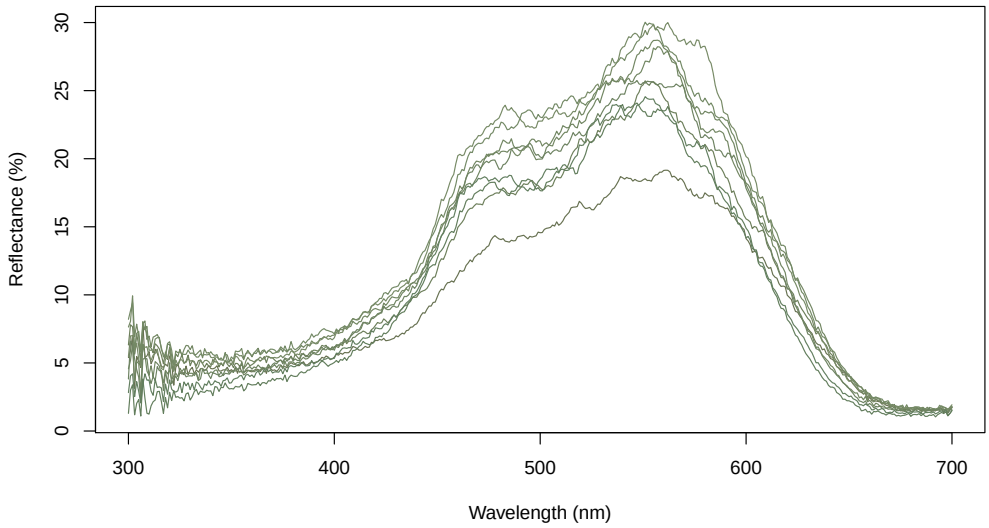
Instead, we recommend you keep the files in the proprietary format (such as Avantes `ABS`, `ROH` and `TRM`, or OceanOptics `ProcSpec` and `jdx`) and that you use `lightr` to convert them into your preferred file format (such as `csv`).

Usage and future directions

`lightr` can serve as a basis for colour analysis R packages to deal with the file import step. Most of them can only read a limited variety of file formats currently. Future versions of `pavo`, for example, will include `lightr` as a dependency. Below is an illustration of a workflow where `lightr` is used to import the spectral data, which is then analysed with `pavo`:

```
library(lightr)
specs <- lr_get_spec(where = "yourfolder", ext = "ProcSpec")

library(pavo)
plot(specs, col = spec2rgb(specs))
```




```
summary(specs, subset = TRUE)
```

B2	S8	H1
9.31682	1.915661	561
11.26643	2.156246	551
12.78053	2.128401	557
13.41558	2.123076	551
13.44852	2.118632	562
12.14774	2.210931	557
11.76633	2.076845	547
10.62519	2.204452	551
10.14280	2.272771	547

The first column indicates the brightness (in % relative to a white reference), the second is the saturation (also called spectral purity) and the last contains the hue (in nm).

`lightr` can also prove useful for developers of other programming languages, providing a free and open source template that can easily be translated to such other languages. We also plan on providing a web application based on shiny (https://github.com/rstudio/shiny), which uses `lightr` in the background, and provides users with limited R or technical knowledge with a simple and convenient way to convert all their proprietary files to `csv`.

Acknowledgements

We thank the two rOpenSci reviewers, Jeroen Ooms and Karthik Ram, for their helpful feedback that improved this package, as well as JOSS editor Daniel S. Katz for his comments on this manuscript.

References

Bennett, A. T. D., Cuthill, I. C., & Norris, K. J. (1994). Sexual selection and the mismeasure of color. *The American Naturalist*, *144*(5), 848–860. doi:10.1086/285711

Cuthill, I. C., Bennett, A. T. D., Partridge, J. C., & Maier, E. J. (1999). Plumage reflectance and the objective assessment of avian sexual dichromatism. *The American Naturalist*, *153*(2), 183–200. doi:10.1086/303160

Eaton, M. D. (2005). Human vision fails to distinguish widespread sexual dichromatism among sexually "monochromatic" birds. *Proceedings of the National Academy of Sciences*, *102*(31), 10942–10946. doi:10.1073/pnas.0501891102

Maia, R., Eliason, C. M., Bitton, P.-P., Doucet, S. M., & Shawkey, M. D. (2013). Pavo: An R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution*, *4*(10), 906–913. doi:10.1111/2041-210X.12069

Maia, R., Gruson, H., Endler, J. A., & White, T. E. (2019). Pavo 2: New tools for the spectral and spatial analysis of colour in R. *Methods in Ecology and Evolution*, *10*(7), 1097–1107. doi:10.1111/2041-210X.13174

Osorio, D. C., & Vorobyev, M. (2008). A review of the evolution of animal colour vision and visual communication signals. *Vision Research*, Vision Research Reviews, *48*(20), 2042–2051. doi:10.1016/j.visres.2008.06.018

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics*, *10*(3), 405–408. doi:10.1093/biostatistics/kxp014

White, T. E., Zeil, J., & Kemp, D. J. (2015). Signal design and courtship presentation coincide for highly biased delivery of an iridescent butterfly mating signal. *Evolution*, *69*(1), 14–25. doi:10.1111/evo.12551