

My original contributions and additional information

Harriet Mason

Overview

The thesis is written in paper format, so that it will be ready to submit to the R Journal, after completion of the Honours degree.

This work has also been used as a test case for the new HTML format for the R Journal. The new format delivers the journal content natively in HTML, to accommodate interactive graphics, and in the traditional PDF format. As such, there may still be some small errors in the typesetting and figure sizing, in this thesis, that are beyond my control. However, this has been enormously useful for the journal's editorial staff to identify problems with the new styling and to fix them.

The new format uses Rmarkdown as the source file, which includes all the code needs to produce the results reported in the paper. The full set of files can be found at <https://github.com/harriet-mason/paper-cassowaryr/tree/main/paper-RJ>.

My original contributions

This project has involved the work of several people, but I have lead the development since taking on the project at the start of Honours. The initial package structure was created by Stuart Lee. Most of the scagnostics definitions originated from previous work in the binostics package collated by Ursula Laa, but this package has to be abandoned because the code base in incomprehensible C makes it impossible to maintain.

While this work is primarily my own under the supervision of Di Cook, Ursula Laa and Stuart Lee, as a joint work some parts are more my own than others. While Stuart set up the skeleton of the package, I am the package maintainer and most of the code in the package, was written by me, as well as all the adjustments made to the scagnostics. Some of the initial descriptions of scagnostics were written by Ursula, but I have added in visual diagrams and additional explanations. We brainstormed possible examples together, but the AFLW example is my own work, the physics example was started in previous work by Ursula, so this example is mostly her work although I did make adjustments to the writing and figures in multiple rewrites. The time series example was started by Di, but I took it over, and contributed a similar amount to the WBI example as I did the physics example. The writing of the rest of the paper is primarily my own work.

Limitations and Caveats

The greatest limitation in this project was the 1 year time limit on the research. When the presenting the research proposal, I thought there would be time to code up all the original scagnostics, fix any issues with them, implement binning as an option, and have time left over to create scagnostics completely of my own design. This clearly did not occur. Ultimately coding up the previous scagnostics was more time consuming than I originally thought it would be, especially because there was a fair bit of room for interpretation on some of the scagnostics. For example, skinny and convex do not specify what to do in the event the alpha hull has no area (this occurs when the data is on a perfect straight line), and this edge case is now handled by setting them to 1 and 0 respectively. Additionally, a large portion of the project was invisible to me at the outset, such as the testing to ensure the scagnostics were working according to definition, rounds of debugging, and meeting CRAN requirements. These elements produced very little output but were required

for the rigour of the project. On top of this, to reduce dependencies most of the functions were written only using base R, which made the project more challenging. For these reasons, the software aspect of this thesis narrowed in scope throughout the year, however the number of examples and applications increased, as we recognised new ways the scagnostics could be used. Ultimately the final project is significantly different to its original goals, but it contains an equal amount of work.

Software

The package is not yet accepted on CRAN, so instructions for CRAN installation in the paper will not yet work. Installation needs to follow the GitHub instructions.