

September 17, 2024

Editor
The R Journal

Dear Dr Mark van der Loo,

Please consider our article titled “Teaching Computers to See Patterns in Scatterplots with Scagnostics” for publication in the R Journal.

The manuscript introduces the new package called **cassowaryr** which provides a user-friendly method to calculate scagnostics in R and examples that illustrate how these scagnostics can be used to explore data. Scagnostics are a set of metrics that describe the shapes in scatter plots, and can be useful as a dimension reduction technique. This package calculates the scagnostics in R, which makes the package more accessible compared to previous scagnostics packages that are simply wrappers for C++ code. Additionally, this package makes several changes to the existing scagnostics (such as the removal of binning as a preprocessing step) and introduces some new scagnostics that have not been discussed previously in the literature. These changes and additions mean the work around the **cassowaryr** package should be documented beyond just function documentation.

We believe the readers of the R Journal will find this article helpful because scagnostics are commonly used in data analysis, and are required for other popular packages such as **tourr** and **ggobi**. This paper provides details on an actively maintained scagnostics package and highlights potential pitfalls that should be avoided when using them. This paper also discusses several different ways that scagnostics can be used, some of which are quite novel and have not been discussed previously in the literature. We illustrate these applications with handful of detailed examples. These examples would be invaluable to those who are unaware of the usefulness of scagnostics as a data exploration tool.

There are some errors returned by `initial_checks.log()`. The nested subdirectories error seems to be caused by `mason-lee-laa-cook_files` which was made automatically by **rjtools**. The `intial_checks.log` files is made when we run `initial_checks.log()`, so it does not seem possible to fix this error. There are the only errors returned by the function.

Regards,

Harriet Mason
Department of Econometrics and Business Statistics
University of Monash
Clayton, VIC, Australia
harriet.mason1@monash.edu

Dianne Cook Department of Econometrics and Business Statistics
University of Monash

Clayton, VIC, Australia
dicook@monash.edu

Ursula Laa Institute of Statistics University of Natural Resources and Life Sciences Vienna.
Vienna, Austria ursula.laa@boku.ac.at

Stuart Lee
Melbourne Data Analytics Platform
University of Melbourne
Melbourne, Australia
stuart.andrew.lee@gmail.com