Ranking data analysis and synthesis of crop science experiments in R: the `gosset` package

Kauê de Sousa[1,2[*]], Jacob van Etten[2], David Brown[3,4], Jonathan Steinke[2,5]
[1] Department of Agricultural Sciences, Inland Norway University of Applied Sciences, 2318 Hamar, Norway
[2] Digital Inclusion, Bioversity International, Parc Scientifique Agropolis II, 34397, Montpellier Cedex 5, France
[3] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB, Wageningen, The Netherlands
[4] Digital Inclusion, Bioversity International, 30501, Turrialba, Costa Rica
[5] Humboldt University Berlin, Berlin, Germany
[*] Correspondence should be addressed to: kaue.desousa@inn.no

Abbreviations: Data-driven Agriculture; Plackett-Luce model

Core ideas

- We developed the freely available and open-source R package gosset
- The package functionality support rank-based analysis and synthesis of data from crop variety evaluations.
- The package facilitates tasks in the main stages of the data analysis workflow.

Abstract

Reproducible and efficient workflows are fundamental in scientific research (Lowndes et al. 2017). Data analysis workflows roughly include the following stages: (1) Data preparation and cleaning, (2) modelling and validation, and (3) results presentation. It is also common that those stages would be repeated iteratively until a solution is found which satisfies the initial objectives. Every of these stages presents different difficulties and constraints for the researchers. Digital tools can facilitate the tasks within those stages, but it is necessary to choose the right tool, if any, for the intended work. While not frequently used, experimental approaches using ranking data are being applied for the evaluation of crop varieties. Recently developed rank-based approaches for on-farm experimentation, such as the tricot methodology (van Etten et al. 2019), required customized tools for all the aforementioned stages. On the other, new rank-based data synthesis approaches also required tailored tools to facilitate all the involved stages. Along with those experiences, we developed the R package gosset, supporting several activities in the analysis of experiments in agronomy and crop science. The main objective of this paper is to describe and demonstrate the functionality of the package.

Introduction

Materials and Methods

```
library("climatrends")
library("sf")
library("nasapower")
```

```r
# create a polygon within the coordinates 7, 17, 59, 63
e <- matrix(c(7, 59, 17, 59, 17, 63,
              7, 63, 7, 59),
            nrow = 5, ncol = 2, byrow = TRUE)

e <- st_polygon(list(e))

# sample 100 points in the hexagonal type
p <- st_sample(e, 100, type = "hexagonal")
p <- st_as_sf(p, crs = 4326)

# compute the temperature indices using the random points
temp <- temperature(p, day.one = "2000-01-01", last.day = "2019-12-31",
                    timeseries = TRUE, intervals = 365)
```


Acknowledgements

This work was supported by The Nordic Joint Committee for Agricultural and Food Research (grant num. 202100-2817). We thank Julian Ramirez-Villegas and Marcel Schrijvers-Gonlag for the useful insights and discussion that helped in the development of this study.

References

Lowndes, Julia S. Stewart, Benjamin D. Best, Courtney Scarborough, Jamie C. Afflerbach, Melanie R. Frazier, Casey C. O'Hara, Ning Jiang, and Benjamin S. Halpern. 2017. "Our Path to Better Science in Less Time Using Open Data Science Tools." *Nature Ecology & Evolution* 1 (6). https://doi.org/10.1038/s41559-017-0160.

van Etten, Jacob, Kauê de Sousa, Amílcar Aguilar, Mirna Barrios, Allan Coto, Matteo Dell'Acqua, Carlo Fadda, et al. 2019. "Crop variety management for climate adaptation supported by citizen science." *Proceedings of the National Academy of Sciences* 116 (10): 4194–99. https://doi.org/10.1073/pnas.1813720116.