gosset: An R package for analysis and synthesis of ranking data in agricultural experimentation

## Kauê de Sousa1,2[\*], David Brown3,4, Jonathan Steinke2,5, Jacob van Etten2

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](mailto:kaue.desousa@inn.no)

# Abstract

Appropriate data management and analysis are necessary to produce practical information from agricultural experimentation data. There is also an ongoing trend advocating for programmatic tools that supports reproducible workflows in scientific research. We developed the R package gosset, providing functionality to support analysis workflows with rank-based models, such as Plackett-Luce and Bradley-Terry. The gosset package facilitates data preparation, modelling and results presentation stages. We demonstrate the functionality of the package with a case of on-farm evaluations of common bean (*Phaseolus vulgaris* L.) genotypes in Nicaragua.

**Keywords**: Bradley-Terry, Plackett-Luce, data science, on-farm trials, participatory-research, tricot

# Required Metadata

# Current code version

*Code metadata*

| **Nr.** | **Code metadata description** | **Please fill in this column** |
| --- | --- | --- |
| C1 | Current code version | 0.4.003 |
| C2 | Permanent link to code/repository used for this code version | <https://github.com/AgrDataSci/gosset> |
| C3 | Code Ocean compute capsule |  |
| C4 | Legal Code License | MIT |
| C5 | Code versioning system used | Git |
| C6 | Software code languages, tools, and services used | R |
| C7 | Compilation requirements, operating environments & dependencies |  |
| C8 | If available Link to developer documentation/manual | <https://agrdatasci.github.io/gosset/> |
| C9 | Support email for questions | [kaue.desousa@inn.no](mailto:kaue.desousa@inn.no) |

# Motivation and significance

Participatory experimentation approaches have been increasingly applied in agricultural research (De Roo et al., 2019). While collecting data in ranking format is uncommon in general agricultural research settings, it is often collected in participatory experiments (Coe, 2002). Recently developed approaches for on-farm experimentation, such as the tricot methodology are based on the collection of data in ranking format (van Etten, Beza, et al., 2019). On the other hand, newly proposed approaches for synthesis of crop variety evaluation data largely depend on the analysis of ranking data (Brown et al., 2020). The analysis of ranking data requires the use of appropriate statistical models such as Plackett-Luce (Luce, 1959; Plackett, 1975) or Bradley-Terry (Bradley & Terry, 1952). Functionality for fitting Bradley-Terry and Plackett-Luce models are available in R with the packages BradleyTerry2 and PlackettLuce respectively (H. L. Turner et al., 2020; H. Turner & Firth, 2012). However, extended functionality was required for the entire data science workflow, which usually includes: (1) Data preparation and cleaning, (2) modelling and validation, and (3) results presentation. For (1) gosset provides functions for converting and preparing data into ranking or pairwise format required by the packages PlackettLuce and BradleyTerry2 respectively. For (2), gosset provides functions for model selection and validation using cross-validation. In the case of (3), enhanced functionality for plotting model results is provided by the gosset package.

# Software description

The R package gosset provides functionality supporting the analysis workflows in agricultural experimentation, especially with rank-based approaches. The package is available in The Comprehensive R Archive Network (CRAN) and can be installed by executing install.packages("gosset"). The package is named in honor to William Sealy Gosset, known by the pen name ‘Student’, a pioneer of modern statistics in small sample experimental design and analysis (Ziliak, 2019).

## Software Architecture

The R package gosset is structured following the guidelines described in the manual for creating R add-on packages (R Core Team, 2022). This structure basically consist of files DESCRTIPTION, LICENSE, NAMESPACE and NEWS, and directories data, dev, docs, inst, man, R, and vignettes. The package functions were developed following the S3 methods style (R Core Team, 2022) and are contained in the R sub-directory.

## Software Functionalities

### Data management and preparation

* rank\_binomial transforms a ranking object into a binary comparisons, as required by package BradleyTerry2 (H. Turner & Firth, 2012).
* rank\_numeric converts numeric values into rankings. The parameter ascending = indicates if the rankings should be made considering the numeric values in ascending order. The default is asceding = FALSE. This function is useful when the data were collected as numerical observations, for instance, in a experiment measuring crop yield.
* rank\_tricot transforms data in tricot format into PlackettLuce rankings.

### Modelling

* AIC computes the Akaike Information Criterion Akaike (1974)] for a Bradley-Terry model fitted with BradleyTerry2 (H. Turner & Firth, 2012) or a Plackett-Luce model fitted with (H. L. Turner et al., 2020).
* btpermute deviance-based forward variable selection (Lysen, 2009) procedure for Bradley-Terry models.
* compare compare the agreement between methods.
* crossvalidation performs k-fold cross-validation, where k could be specified by the user.The default is 10-fold. Folds can be provided as a vector for a custom cross-validation, such as blocked cross-validation.
* forward executes forward variable selection with cross-validation.
* kendallTau computes the Kendall-tau rank correlation coefficient between two rankings (Kendall, 1938).
* kendallW computes Kendall’s W (coefficient of concordance) among observed rankings and those predicted by the Plackett-Luce model (Kendall & Smith, 1939).
* pseudoR2 computes goodness-of-fit metrics, such as McFadden’s pseudo-R2 (McFadden et al., 1973).

### Visualization and results presentation

* plot provides an alternative method to the default plot function of the PlackettLuce package.
* regret computes the regret (Loomes & Sugden, 1982).
* reliability computes the reliability of crop varieties, the probability to outperform the reference variety (Eskridge & Mumm, 1992).
* worth\_bar creates a bar plot of the estimated worth values for each evaluated variety.
* worth\_map creates a heatmap plot of the estimated worth values for all varieties considering each of the evaluated traits.

# Illustrative example

To demonstrate the functionality of the gosset package, we use the nicabean dataset, which was generated with decentralized on-farm trials of common bean (*Phaseolus vulgaris* L.) varieties in Nicaragua over five seasons (between 2015 and 2016). Following the tricot approach (van Etten, Beza, et al., 2019), farmers were asked to test in their farms three varieties of common bean. The varieties were randomly assigned as incomplete blocks of size three (out of 10 varieties). The farmers assessed which of the three varieties had the best and worst performance in eight traits (vigor, architecture, resistance to pests, resistance to diseases, tolerance to drought, yield, marketability, taste). The farmers also provided their overall appreciation about the varieties, i.e., which variety had the best and the worst performance based on the overall performance considering all the traits. To analyze the data, we use the Plackett-Luce model implemented in the R package *PlackettLuce* (H. L. Turner et al., 2020).

The nicabean is a list with two data frames: i) trial, contains the trial data with farmers’ evaluations, ranked from 1 to 3, with 1 being the higher ranked variety and 3 the lowest ranked variety for the given trait and incomplete block; ii) covar, contains the covariates associated to the on-farm trial plots and farmers. This example will require the packages PlackettLuce (H. L. Turner et al., 2020), climatrends (de Sousa et al., 2020), and ggplot2 (Wickham, 2016).

library("gosset")  
library("PlackettLuce")  
library("climatrends")  
library("ggplot2")  
  
data("nicabean", package = "gosset")  
load("nicabean\_temp\_ind.rda")  
  
dat <- nicabean$trial  
  
covar <- nicabean$covar  
  
traits <- unique(dat$trait)

To start the analysis of the data, we transform the rankings into Plackett-Luce rankings using the function rank\_numeric(). We run iteratively over the traits adding the rankings to a list called R. Since the varieties are ranked in an ascending order, with 1 being the higher ranked and 3 the lower ranked, we use the argument asceding = TRUE to indicate which order should be used.

R <- vector(mode = "list", length = length(traits))  
  
for (i in seq\_along(traits)) {  
   
 dat\_i <- subset(dat, dat$trait == traits[i])  
   
 R[[i]] <- rank\_numeric(data = dat\_i,  
 items = "item",  
 input = "rank",   
 id = "id",   
 ascending = TRUE)  
}

Then, using the function kendallTau() we can assess the correlation between rankings. This approach can be support at least two decision making processes by, for example i) indicating the traits

baseline <- which(grepl("OverallAppreciation", traits))  
  
kendall <- lapply(R[-baseline], function(X){  
 kendallTau(x = X, y = R[[baseline]])  
})  
  
kendall <- do.call("rbind", kendall)  
  
kendall$trait <- traits[-baseline]

Kendall correlation between ‘overall performance’ and the other traits assessed in the Nicaragua bean on-farm trials.

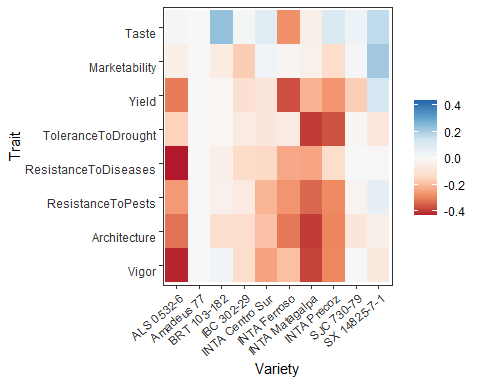
| trait | kendallTau |
| --- | --- |
| Vigor | 0.439 |
| Architecture | 0.393 |
| ResistanceToPests | 0.463 |
| ResistanceToDiseases | 0.449 |
| ToleranceToDrought | 0.411 |
| Yield | 0.749 |
| Marketability | 0.639 |
| Taste | 0.653 |

For each trait, we fit a Plackett-Luce model using the function PlackettLuce from the package of the same name. We use lapply with the list of rankings and the PlackettLuce function as input parameters.

mod <- lapply(R, PlackettLuce)

The worth\_map function can be used to visually assess variety performance based on different traits. The values represented in a worth\_map are worth estimates in the log scale.

worth\_map(mod[-baseline],  
 labels = traits[-baseline],   
 ref = "Amadeus 77") +  
 labs(x = "Variety",  
 y = "Trait")



Trait performance (log-worth) of bean varieties in Niragua. Variety ‘Amadeus’ is set as reference (log-worth = 0). Blue values indicate a superior performance of varieties for a given trait, compared to the reference. Red values indicate a variety with weak performance for the given trait, compared to the reference.

To consider the effect of climatic factors on variety performance, we used temperature-derived covariates to fit a Plackett-Luce tree. For brevity, we used the variables maximum nighttime temperature (maxNT) and maximum daytime temperature, but more variables can be used in a Plackett-Luce tree. Temperature data were obtained from the AgERA5 dataset (Boogaard & Grijn, 2020), using the R package *ag5Tools* (Brown & de Sousa, 2022). The temperature data downloaded from AgERA5 were used for computing the climatic variables with the package *climatrends* (de Sousa et al., 2020). In this case, we only use the rankings corresponding to the overall performance of varieties.

To be coupled with climatic covariates, the rankings should be made formatted as grouped rankings. For this we used the function rank\_numeric with the parameter group = TRUE.

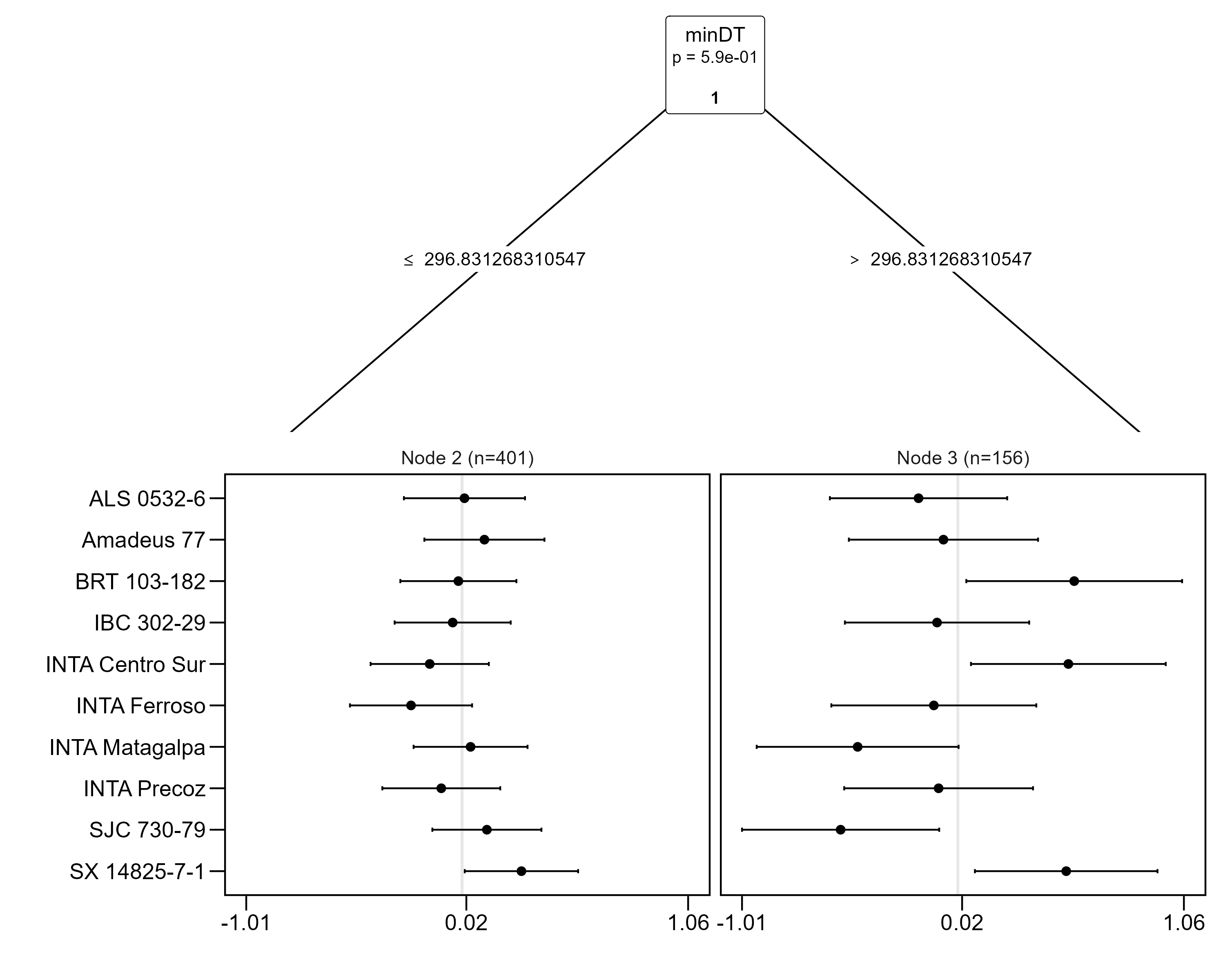
R <- vector(mode = "list", length = length(traits))  
  
for (i in seq\_along(traits)) {  
   
 dat\_i <- subset(dat, dat$trait == traits[i])  
   
 R[[i]] <- rank\_numeric(data = dat\_i,  
 items = "item",  
 input = "rank",   
 id = "id",   
 ascending = TRUE,  
 group = TRUE)  
}

Now we can fit the Plackett-Luce tree with climate covariates. Notice that we used the 9th element in the list of rankings as we are only using the overall appreciation ranking.

R <- R[[9]]  
  
pld <- cbind(R, nicabean\_temp\_ind)  
  
pl <- pltree(R ~ .,   
 alpha = .6,  
 data = pld)

The following is an example of the plot made with the function plot in the gosset package.

plot(pl)



“Plackett-Luce tree plot with gosset function *plot*”

We can use the function reliability to compute the reliability of the evaluated common bean varieties in each of the resulting nodes of the Plackett-Luce tree (Table 3).

Reliability of evaluated varieties in each of the resulting nodes of the Plackett-Luce tree

|  | node 2 | node 3 |
| --- | --- | --- |
| ALS 0532-6 | 0.9019670 | 0.8760321 |
| Amadeus 77 | 1.0000000 | 1.0000000 |
| BRT 103-182 | 0.8701957 | 1.4583517 |
| IBC 302-29 | 0.8395336 | 0.9694489 |
| INTA Centro Sur | 0.7077457 | 1.4433065 |
| INTA Ferroso | 0.5891942 | 0.9538329 |
| INTA Matagalpa | 0.9329531 | 0.5046691 |
| INTA Precoz | 0.7758749 | 0.9769946 |
| SJC 730-79 | 1.0115239 | 0.3803085 |
| SX 14825-7-1 | 1.1593903 | 1.4376290 |

With the function crossvalidation we can validate the model.

Results of 10-fold random cross-validation

| AIC | deviance | logLik | logLikNull | McFadden | kendallTau |
| --- | --- | --- | --- | --- | --- |
| 2438.664 | 2038.664 | -1019.332 | -8982.09 | 0.8865054 | 0.1478579 |

# Impact

Reproducible and efficient workflows are fundamental in scientific research (Lowndes et al. 2017). The gosset package provides functionality that was not previously available from other R packages and which enabled scientific studies based on the analysis of ranking data. This functionality enables making the entire workflow to be reproducible and more efficient. The utility of the gosset package has been demonstrated by enabling studies based on the analysis of ranking data. For instance, van Etten, de Sousa, et al. (2019), Moyo et al. (2021) and de Sousa et al. (2021) applied the Plackett-Luce model in combination with recursive partitioning (H. L. Turner et al., 2020; Zeileis et al., 2008). In these studies, the gosset package supported data preparation, model validation and results presentation tasks.

# Conclusions

We described the functionality of the R package gosset to support the synthesis and analysis of ranking data. The package provide functions not available in existing R packages for analyzing ranking data. We provided an illustrative example covering the main functionality across the stages involved in the analysis workflow.

# Conflict of Interest

No conflict of interest exists: We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

# Acknowledgements

We acknowledge Olga Spellman (Science Writing Service of the Alliance of Bioversity International and CIAT) for English editing of this manuscript

# References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Boogaard, H., & Grijn, G. van der. (2020). *Agrometeorological indicators from 1979 to present derived from reanalysis* (W. E. Research, Ed.).

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), 324–345.

Brown, D., & de Sousa, K. (2022). *ag5Tools: Toolbox for downloading and extracting copernicus AgERA5 data*. <https://agrdatasci.github.io/ag5Tools/>

Brown, D., Van den Bergh, I., de Bruin, S., Machida, L., & van Etten, J. (2020). Data synthesis for crop variety evaluation. A review. *Agronomy for Sustainable Development*, *40*(4), 1–20.

Coe, R. (2002). Analyzing ranking and rating data from participatory on-farm trials. In M. R. Bellon & J. Reeves (Eds.), *Quantitative Analysis of Data from Participatory Methods in Plant Breeding* (pp. 44–65). CIMMYT.

De Roo, N., Andersson, J. A., & Krupnik, T. J. (2019). On-farm trials for development impact? The organisation of research and the scaling of agricultural technologies. *Experimental Agriculture*, *55*(2), 163–184.

de Sousa, K., van Etten, J., Poland, J., Fadda, C., Jannink, J.-L., Kidane, Y. G., Lakew, B. F., Mengistu, D. K., Pè, M. E., Solberg, S. Ø.others. (2021). Data-driven decentralized breeding increases prediction accuracy in a challenging crop production environment. *Communications Biology*, *4*(1), 1–9.

de Sousa, K., van Etten, J., & Solberg, S. Ø. (2020). *climatrends: Climate variability indices for ecological modelling*. <https://CRAN.R-project.org/package=climatrends>

Eskridge, K. M., & Mumm, R. F. (1992). Choosing plant cultivars based on the probability of outperforming a check. *Theoretical and Applied Genetics*, *84*(3), 494–500. <https://doi.org/10.1007/BF00229512>

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1-2), 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>

Kendall, M. G., & Smith, B. B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, *10*(3), 275–287. <http://www.jstor.org/stable/2235668>

Loomes, G., & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, *92*(368), 805–824. <https://doi.org/10.2307/2232669>

Luce, R. D. (1959). *Individual Choice Behavior* (p. 153). Courier Corporation.

Lysen, S. (2009). Permuted inclusion criterion: A variable selection technique. *Publicly Accessible Penn Dissertations*, 28.

McFadden, D.others. (1973). *Conditional logit analysis of qualitative choice behavior*.

Moyo, M., Ssali, R., Namanda, S., Nakitto, M., Dery, E. K., Akansake, D., Adjebeng-Danquah, J., van Etten, J., de Sousa, K., Lindqvist-Kreuze, H., Carey, E., & Muzhingi, T. (2021). Consumer Preference Testing of Boiled Sweetpotato Using Crowdsourced Citizen Science in Ghana and Uganda. *Frontiers in Sustainable Food Systems*, *5*(3), 6. <https://doi.org/10.3389/fsufs.2021.620363>

Plackett, R. L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *24*(2), 193–202. <https://doi.org/10.2307/2346567>

R Core Team. (2022). *Writing r extensions*. R Foundation for Statistical Computing Vienna, Austria. <https://cran.r-project.org/doc/manuals/r-release/R-exts.html>

Turner, H. L., van Etten, J., Firth, D., & Kosmidis, I. (2020). Modelling rankings in R: the PlackettLuce package. *Computational Statistics*. <https://doi.org/10.1007/s00180-020-00959-3>

Turner, H., & Firth, D. (2012). Bradley-Terry Models in R: The BradleyTerry2 Package. *Journal of Statistical Software*, *48*(9), 1–21. <https://www.jstatsoft.org/v48/i09/>

van Etten, J., Beza, E., Calderer, L., Van Duijvendijk, K., Fadda, C., Fantahun, B., Kidane, Y. G., Van De Gevel, J., Gupta, A., Mengistu, D. K., Kiambi, D. A. N., Mathur, P. N., Mercado, L., Mittra, S., Mollel, M. J., Rosas, J. C., Steinke, J., Suchini, J. G., & Zimmerer, K. S. (2019). First experiences with a novel farmer citizen science approach: Crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Experimental Agriculture*, *55*(S1), 275–296. <https://doi.org/10.1017/S0014479716000739>

van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell’Acqua, M., Fadda, C., Gebrehawaryat, Y., van de Gevel, J., Gupta, A., Kiros, A. Y., Madriz, B., Mathur, P., Mengistu, D. K., Mercado, L., Nurhisen Mohammed, J., Paliwal, A., Pè, M. E., Quirós, C. F., … Steinke, J. (2019). Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences*, *116*(10), 4194–4199. <https://doi.org/10.1073/pnas.1813720116>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514. <https://doi.org/10.1198/106186008X319331>

Ziliak, S. T. (2019). How Large Are Your G-Values? Try Gosset’s Guinnessometrics When a Little “p” Is Not Enough. *The American Statistician*, *73*(sup1), 281–290. <https://doi.org/10.1080/00031305.2018.1514325>