

HoboR: An R package to manipulate weather stations data

Alcalá Briseño RI, Carson AR, Lang S, Peterson E, and LeBoldus, J.

January 18, 2024

Summary

HoboR is an R package for efficiently processing extensive datasets obtained from HOBO weather stations and data loggers, supporting various weather station formats. Multiple tools were designed for streamlined weather data management, **HoboR** enables users to load CSV files into a tibble format, eliminate duplicates, summarize data by time intervals (minutes, hours, and days), subset files by date ranges, and address common data quality issues such as sensor failures, out-of-range entries, and time zone discrepancies. Additionally, the package incorporates guidelines for weather data analysis (REF), advocating for adherence to standard practices in handling weather variables. Despite its name, **HoboR** is adaptable to other weather station formats sharing a similar data structure

Weather station data can be logged by the minute from any point in time and from different types of sensors, such as the standard time, rain, relative humidity (RH), light, and more. **HoboR** main functions implement dynamic interpretation programming, allowing to process the spreadsheet independently of the number of sensors, adapting to different initial headers. Among the challenges of recording and collecting data, replacing batteries and download data could create multiple entries that might be challenging and time consuming to handle in graphic user interface spreadsheet programs. Duplicate entries might vary from seconds to minutes, and with the help of this package, it can be merged and summarized for further data analysis.

HoboR was tested on log files with hundreds to thousands entries, facilitating the post-processing of weather station and data loggers, loading csv files regardless of the header column order and dimensions, and summarised within seconds, the summary statistics can be rounded to minutes, hours and days, yielding minimums and maximum, mean and standard deviation of your data. Additional functions can help to identify and replace impossible values and correct the variation within your loggers. As a proof of concept applied to these weather data, we implement a couple of functions to calculate disease trends of the sudden oak death epidemiology affecting tanoek (*Notholiticarpus densiflorus*) in the Pacific Northwest.

Statement of need

Developing automated software for preprocessing weather stations and data logger information may facilitate the analysis of epidemiological surveillance, microbiome, and multiple disciplines (Dahl et al., 2023; Nikolaou et al., 2023; Wu et al., 2023). Traditional spreadsheet interfaces pose a challenge in handling extensive and complex studies that are difficult to manage, time-consuming to organize, error-prone if done by hand, and might not handle whole datasets. By automating these tasks, **hoboR** enhances accuracy and significantly reduces the time and effort required for data preparation, leaving more time for robust epidemiological modeling. The integration of advanced algorithms and user-friendly software makes it accessible to both experienced researchers and program beginners, addressing the current potential of implementing weather variables for plant pathology and disease ecology for effective management (Garrett et al., 2023).

To our knowledge, no packages in R are available to analyze weather station and data logger files. A graphic user interface for HOBO exists but is incompatible with data postprocessing and summary statistics.

Package workflow

The workflow of the **hoboR** package consists of three consecutive steps and six assisting functions:

- **hobinder**: Load multiple csv files regardless of the order and number of columns from a single directory; the files must come from the same weather station or data logger model.
- **hobocleaner**: Averages duplicate entries from the large csv file.
- **meanhobo**: Summary statistic (min, max, mean, and standard deviation) for the different weather station and data logger sensors
- **hobotime**: Allows aggregating your data by minutes, hours, or days.
- **horange**: Allows to parse your data by date ranges,
- **impossiblevalues**: Identify the min and max values in the data set, the user should consider what are the minimum and maximum values for the region.
- **NAsensorfailures**: Allows to replace with NAs impossible values in your data set using logical statements.
- **timestamp**: Select a time and gives you an interval
- **horrelation**: Plot what weather variable correlates among them.
- **hoboplot**: Plot weather variable trends.

samplingrates() sampling.trends()

Example

Installation

This package requires R version 4.1.3 or later. It also requires the following packages: `data.table`, `dplyr`, `ggplot2`, `lubridate`, `plyr`, `purrr`. These dependencies should be installed automatically when `dependencies = TRUE` is set in the command used to install the package.

```
> if (!require("devtools")) \\  
> install.packages("devtools")\\  
> devtools::install_github("leboldus_lab/hoboR", dependencies = TRUE)
```

Authors contribution

Ricardo I. Alcalá authored and developed the original version of the package, maintained the package, wrote the documentation, debugged the code, and wrote the manuscript. Adam R. Carson collected the data, wrote code implemented in the package's main functions, and debugged the code. Sky Lang collected the data and assisted in the user-functionality of the code functions. Ebba Peterson assisted in best practices for post-processing. Jared LeBoldus supervised the project and participated in the manuscript drafting process.

Acknowledgements

Garrett et al., 2023 <https://doi.org/10.1146/annurev-phyto-021021-042636>

Dahl et al., 2023, <https://doi.org/10.1111/1462-2920.16347>

Nikolaou et al., 2023, <https://doi.org/10.1016/j.envres.2023.117173>

Wu et al., 2023, <https://doi.org/10.1093/aob/mcad195>