Survey on the R tracking packages

Rocio Joo

October 14, 2018

A large amount of R packages for movement and the need for a survey

The use of tracking devices is increasing as a consequence of the development of new and cheaper loggers. With large amounts of available data, more sophisticated tools to process, visualize and analyze data are being developed. One of the most used programming softwares for this matter is the free open source R software. We have identified more than 70 R packages created for that purpose! The large amount of existent packages makes it difficult to keep track of the spectrum of choices. For that reason, we aimed at assessing the current state of the art of the movement-related R packages regarding

1. How popular those packages are;
2. How well documented they are;
3. How relevant they are for users.

Packages included in the survey

In theory, any package could be potentially useful for movement analysis; either a time series package or a spatial analysis one or even `ggplot2` to make more beautiful graphics! For this survey, we considered only packages created with the main purpose of 1) processing or analyzing data obtained from devices used to track organisms, or 2) analyze tracking data in general. Packages created for eye or computer-mouse movement were not considered here (but you are welcome to make your own survey about them!). These packages are:

are working on a review of the packages intended as a road map for users, to know what exists and their purpose, and discuss new challenges for package developers.

Survey Questions

The purpose of this anonymous survey (see the call to participate in [LINK]) was an assessment about the current state of the art of movement-related R packages about three aspects:

focused on tracking or trajectory data,

All of the packages in the main manuscript were considered in the survey except for trajr, which was identified for the review after the survey started.

The survey got exemption from the Institutional Review Board at University of Florida. IRB02 Office, Box 112250, University of Florida, Gainesville, FL 32611-2250.

Since the only question analyzed for the purpose of this review concerns package documentation, we only detail that question here.

Package documentation

How helpful is the documentation provided for each of the packages you've used for your work? Documentation includes what is contained in the manual and help pages, vignettes, published manuscripts, and other material about the package provided by the authors. Please answer using one of the following options:

- `Not enough: It's not enough to let me know how to do what I need;`
- `Basic: It's enough to let me get started with simple use of the functions but not to go further (e.g. use all arguments in the functions, or put extra variables);`
- `Good: I did everything I wanted and needed to do with it;`
- `Excellent: I ended up doing even more than what I planned because of the excellent information in the documentation.`
- `Don't remember: I honestly can't remember...`

Survey representativity

There was no previous selection of the participants and no probabilistic sampling was involved. The survey was advertised by Twitter, mailing lists (r-sig-geo and r-sig-ecology), individual emails to researchers and the mablab website: http://matlab.org. We analyzed only completed surveys (all questions answered); 225 participants completed the survey.

To get an idea of how representative the survey was of the population of R-tracking-packages users, we compared the number of participants that used each package to the number of monthly downloads that each package has.

The number of downloads were calculated using the R package cran.stats. It calculates the number of independent downloads by each package (substracting downloads by dependencies) by day. It only works for downloads using Rstudio, and for packages on CRAN. So for the tracking packages on CRAN, we computed the average of downloads per month, from September 2017 to August 2018; less months were considered for packages that were less than one year old.

There is no perfect match between the number of users and the number of downloads per package, but a correlation 0.88 for the 42 packages on CRAN, provides evidence of a good representativity of the users of tracking packages in the survey. A log-log plot for both metrics is shown in the figure below.

Description of survey analysis

Only responses of participants who remembered the documentation, and packages with more than 10 respondents were considered in the analysis.

We counted the number of participants who expressed that the documentation was either good or excellent, and divided by the total of answers on documentation given about the package. The packages that had more than 0.75 (or 75%), were considered to have “adequate documentation”.