R and Reproducibility: a Proposal

David Smith[1,*], Joseph Rickert[1]

1. Revolution Analytics
*Contact author: david@revolutionanalytics.com

Keywords: R, reproducibility, packages, CRAN

The R ecosystem is in a state of near constant change. While a new version of the R engine is now released just once a year, 2-3 patches are usually released in the interim. On top of that, new versions of R packages on CRAN are released at rate of several per day (and that's not counting packages that are part of the BioConductor project or hosted elsewhere on the Web).

While this rapid change is a boon for the advancement of R, it can cause problems for package authors[1] and also for scientists and their peers who may need to reliably reproduce the results of an R script (possibly dependent on a number of packages) months or even years down the line. In this talk we propose a downstream distribution of R and CRAN packages that provides for the reproducibility of R scripts and reduces the impact of dependencies for packages authors.

References

[1] Ooms, Jeroen (2013) "Possible Directions for Improving Dependency Versioning in R", *The R Journal* Vol. 5/1, June 2013