Fostering the next generation of data-driven open science with R

Karthik Ram [1,2]

1. The rOpenSci Project

2. Berkeley Initiative in Global Change Biology, University of California, Berkeley.

karthik.ram@gmail.com

Abstract

Research is becoming increasingly data intensive and computation driven across various scientific domains, from the social and life sciences all the way to particle physics. Many new scientific insights will likely emerge from vast stores of existing data, rather than from new data collection efforts. In addition, funder and journal mandates now require that researchers share at least their final datasets at the time of publication. As a result of such changes, researchers not only need to continue maintaining their domain expertize, but also be proficient in skills necessary to acquire, manipulate, document and share their data.

rOpenSci is a community driven effort to foster such data driven science among research communities that use R. Our suite of tools (http://ropensci.org/packages/) allow access to these data repositories through a statistical programming environment that is already a familiar part of the workflow of many scientists. Our tools not only facilitate drawing data into an environment where it can readily be manipulated and visualized, but also one in which those analyses and methods can be easily shared, replicated, and extended by other researchers. In this poster and accompanying live demo I highlight some our our recent efforts in advancing open and transparent practices in the sciences.