



Using R for large spatiotemporal data sets

Edzer Pebesma

University of Münster, Institute for Geoinformatics, Germany (edzer.pebesma@uni-muenster.de)

Writing and sharing scientific software is a means to communicate scientific ideas for finding scientific consensus, no more and no less than writing and sharing scientific papers is. Important factors for successful communication are adopting an open source environment, and using a language that is understood by many. For many scientist, R's combination of rich data abstraction and highly exposed data structures makes it an attractive communication tool. This paper discusses the development of spatial and spatiotemporal data handling and analysis with R since 2000, and will point to some of R's strengths and weaknesses in a historical perspective. We will also discuss a new, S3-based package for feature data ("Simple Features for R"), and point to a way forward into the data science realm, where pipeline-based workflows are assumed. Finally, we will discuss how, in a similar vein, massive satellite or climate model data sets, potentially held in a cloud environment, can be handled and analyzed with R.