

# Parallel computing and analysis of large data in R

Simon Urbanek<sup>1,2,\*</sup>

1. AT&T Labs - Research

2. R Core development team

\* Contact author: [simon.urbanek@r-project.org](mailto:simon.urbanek@r-project.org)

**Keywords:** parallel computing, large data

The emergence of affordable multi-core computers provides great opportunity to speed up and scale statistical computing. However, for historical reasons R is limited in the range of possibilities to leverage this technology. In this talk we will illustrate several new options in utilizing R for the analysis of large data by parallelization, including in-depth look at the *multicore* package which allows immediate parallelization in pure R code as well as review and comparison to native-code options and other approaches such as *snow*. We will also show examples how R can be used in practical setting with huge data with sizes in terabytes.