

Wide-to-tall Data Reshaping Using Regular Expressions and the nc Package

by Toby Dylan Hocking

Abstract Regular expressions are powerful tools for extracting tables from non-tabular text data. Capturing regular expressions that describe the information to extract from column names can be especially useful when reshaping a data table from wide (few rows with many regularly named columns) to tall (fewer columns with more rows). We present the R package `nc` (short for named capture), which provides functions for wide-to-tall data reshaping using regular expressions. We describe the main new ideas of `nc`, and provide detailed comparisons with related R packages (`stats`, `utils`, `data.table`, `tidyr`, `tidyfast`, `tidyfst`, `reshape2`, `cdata`).

Toby Dylan Hocking