Dr. Mark van der Loo

Statistics Netherlands and University of Leiden (LIACS)

Course title: Proposal for course on Statistical Data Cleaning with R

Program:

- Processing data and the statistical value chain
- Data Quality
- Techniques for cleaning text data with R
- Data validation with the ‘validate’ R package
- Techniques for deductive correction
- Rule-based data cleaning with R package ‘dcmodify’
- Error localization with R package ‘errorlocate’
- Imputation with R package ‘simputation’
- Process monitoring with the R package ‘lumberjack’

Course format:

The course consists of short lectures interchanged with many exercises. Participants are expected to bring a laptop with a recent version of R and RStudio installed. The following R packages should be installed as well: validate, errorlocate, stringdist, dcmodify, simputation, lumberjack, rspa.

About the speaker

Dr. Mark van der Loo
Senior Researcher, Statistics Netherlands
Research Fellow, LIACS University of Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
pj.vanderloo@cbs.nl

Mark van der Loo received his PhD in molecular physics in 2008 from Radboud University in Nijmegen. He is currently a Senior Methodologist at Statistics Netherlands (CBS) and a Research Fellow at the Leiden Institute for Advanced Computer Science (LIACS) of the University of Leiden. His main research interest is Statistical Computing in the broadest sense. Mark has published popular R packages and scholary articles in the area of statistical data cleaning. Together with Edwin de Jonge he published 'Statistical data cleaning with applications in R' with Wiley Scientific publishers (2018). Mark is currently on the editorial board of the R Journal and co-organizer of the annual 'use of R in official Statistics conference'. At CBS he coordinates the 'hard skills'curriculum for all employees, besides being active in consultancy and research. In Leiden he collaborates with the Network Science group to develop methods and algorithms for disclosure control of (social) network data.