

Using R in the Statistical Office

Alexander Kowarik and Mark van der Loo

uRos 2017 | Bucharest

Statistics Austria

Introducing R

At first

- ▶ Started very “unofficial”
- ▶ Self installed R versions floating around in the office
- ▶ No support

First improvements

- ▶ Standardized R installation
- ▶ Specific units allowed to use R
- ▶ First R server

Support and policy

- ▶ Official support infrastructure (Jira) and responsible unit (Methods)
- ▶ RStudio on server and desktop
- ▶ Presentation of R projects (twice a year)

Infrastructure

- ▶ \pm 40 weekly active users (over 100 installations)

Current situation

- ▶ Installation package for the Windows desktops including R, RStudio, Latex, SVN and RTools
- ▶ RStudio Server on a Linux (Ubuntu) server with 16 cores and 128 GB memory

Future

- ▶ No more desktop R
- ▶ RStudio Server Pro on a Linux machine as default client for everyone

Statistics Netherlands

Introducing R

Typical hurdles (2010):

- ▶ How to install FOSS?
- ▶ OMG everybody can write CODE now!

Approach

- ▶ Project with dedicated project leader
- ▶ Standardized 3 installation types geared to different user types.
- ▶ Set up code/documentation standards

Currently

- ▶ ± 200 users (± 100 active)
- ▶ One single central installation
- ▶ Refer to tidy code/documentation standard

Support and policy

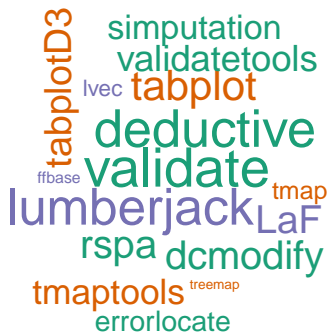
Local user group *kennR!*

- ▶ Beginner's course & advanced workshops
- ▶ User meetings & support
- ▶ Functional management

FOSS Contribution Policy (in short)

- ▶ When relevant to statistics Netherlands, with positive business case.

Packages contributed



- data cleaning
- visualisation
- infrastructure

Current infrastructure

- ▶ R + RStudio on central folder
 - ▶ R-engine usable by non-programmers who just run a script
 - ▶ Selection of R packages pre-installed
 - ▶ Full CRAN repo available internally (there's no direct internet access from most VM's)
- ▶ RDS server (8core, 64G VM's) for heavier work
- ▶ Working on connection to Spark server (Sparklyr)
- ▶ Looking into RStudio/Shiny server but little/no support experience for linux currently exists in SN.

General remarks

Lessons learned

- ▶ Central installation or server solution preferable
- ▶ Training courses are necessary
- ▶ Support is needed when the number of users grow
- ▶ Community is important
- ▶ Internal CRAN mirror for IT security

Collaboration opportunities

- ▶ Packages can be easily shared
- ▶ Interface is unified by R
- ▶ Bottom-up approach much more efficient than defining everything beforehand
- ▶ Survival of the fittest vs. planned standard tools
- ▶ Interesting packages can be found at
 - ▶ Official Statistics Task View (CRAN)
 - ▶ www.awesomeofficialstatistics.org