Systematic Data Validation with R

Mark van der Loo and Edwin de Jonge

Statistics Netherlands Research & Development
@markvdloo @edwindjonge

useR!2021
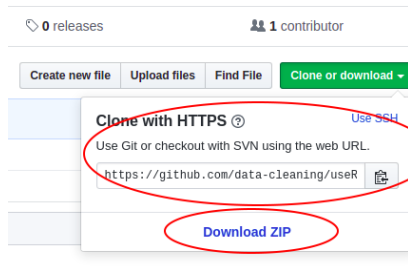
Materials: `github.com/data-cleaning/useR2021_tutorial`

Clone:

```
git clone https://github.com/data-cleaning/useR2021_tutorial
```

Or download:

Please join `#tut_validate` **on slack**

This tutorial

Topics
- Intro to `validate`
- Expressing data checks
- Following data quality in scripts: `lumberjack`
- Managing data validation rules: `validatetools`

Form for each Topic
1. You start with hands-on scripts
2. Presentation: background
3. Extra assignment (if there is time)

Materials and references

Software

JSS **97**(10) 1–33.

data-cleaning.github.io/validate

Theory

Wiley StatsRef Online

Wiley (2018)

Assignment 1

<div align="center">

code/01_intro_validate.R

</div>

- You will be assigned to breakout rooms
- Execute the code line by line.
  - **Discuss what happens with your breakout partner(s)**
  - **Write down any questions you might have**
- After the session, type your questions in the chat.