# Systematic Data Validation with R

Mark van der Loo and Edwin de Jonge

Statistics Netherlands Research & Development
@markvdloo @edwindjonge

useR!2021





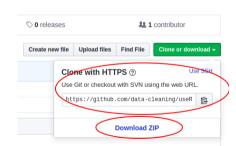
# Materials: github.com/data-cleaning/useR2021\_tutorial

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**



Journal of Statistical Software

Data validation infrastructure for R

Mark P.J. van der Loo Statistics Netherlands

Statistics Netherlands

#### Abstract

Cheving data quality against density have been consequent contributed and the contributed and the contributed and the contributed contributed and contributed contributed

Kryovetic data chedding, data i

#### 1. Introduction

Charles a feature of the market pursuage in lander or formats benefits provide from the contribution of th

JSS 97(10) 1-33.

#### The Data Validation Cookbook

Mark P.J. van der Lao Statistics Nethenlands regi randerloodlichs.nl

2629-22-22 | validate version 1.6.1

#### Preface

This book is about checking data with the validate package for R.

This sension of the lock was rendered with "will little sension 1.8.1. The latest release

inexact generator ("reliable")

The purposes of this book include demonstrating the main tools and workflows of the 'uslidazz'.

#### Prerequisites

Readers of this book are expected to have some knowledge of R. to particular, you should know how to report data into M and some a little about author provided and code before and contract.

#### Citing this work

data-cleaning.github.io/validate

## Theory

#### Wiley StatsRef: Statistics Reference Online

Data Validation

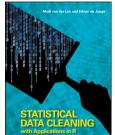
Mark P.J. van der Loo and Edwin de Jonge



Market is a second to search (company of professions seep solds). We consider south a second to search (company of professions seep solds) and second to sec

Secretics Markethrole, The Disput, The Markethrole Sergi Districts Advance Online & 2014 (2021 Adva Miley & Sent, LAI, The Article & XXI (Art Hilly & One, LAI)

Wiley StatsRef Online



Wiley (2018)





# Assignment 1

# code/01\_intro\_validate.R

- 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.



