A UNIFIED APPROACH FOR WRITING AUTOMATIC REPORTS
Parameterization and Generalization of R-Markdown
20 june 2020 9:30 - 12:30

Cristina Muschitiello - Niccolò Stamboglis

AUDIENCE

This workshop is intended for R users who want to learn how to develop their own functions for the creation of customized, but reproducible, R-Markdown reports to improve their projects and/or package development abilities.

OVERVIEW

In this 3 hours workshop, we show a 2-level parameterized (or “super-parameterized”) programming approach for report development consisting in:

- a parameterized R markdown document to run customized analysis
- A parameterized render function for Rmd parameters definition.

The analysis will be applied to a practical case analysing different facets of a real-life application.

Indeed, a parametrized R-markdown report ensures reproducibility, while guaranteeing customisation needs. We show how to define Rmd parameters outside the Rmd code by making use of a render function and how to use this approach in the context of a faceted study or a package definition. This approach offers enhanced generalizability to the writing of markdown documentation, while allowing for customization requirements.

AGENDA

1. R-Markdown basics:
    - Authoring Framework for Data Science
    - Rmd components:
        - Metadata
        - Text
        - Code
        - …
    - Output formats
2. Parameterization of an R-Markdown document
    - Document parameters for format control
    - Output Parameters for content control
3. Generalized R-Markdown documents for reproducibility
    - Render function
    - Parameterized render function

- Generalized render function: where reproducibility meets customization needs.

4. Package example
   - Example of inclusion of a render function within an R package

PREREQUISITES

Basic skills in R and R-Markdown are strongly recommended. Introductory examples are provided, however the core material might be challenging for beginners.

WORKSHOP MATERIAL AND ENVIRONMENT SETTING

The workshop will be conducted on RStudio Cloud. Participants can follow through in two ways:

1. **RStudio Cloud**: the use of RStudio Cloud implies NO need for package installation. Participants will find a set environment complete with all workshop materials. Participants using RStudio cloud need to:
   a. Go to the **this link**
   b. Use one of the followings browsers: Apple Safari or Google Chrome
   c. Create an RStudio Cloud account (free).

2. **On your computer**: Participants who do not want or can't use RStudio cloud can download the workshop material from Github at **this link**. In this case the requirements are:
   a. R Version >= R-3.6.1
   b. RStudio Version >= 1.2.5033
   c. Previous installation of the PrescRiptions Package (included in the workshop material)
   d. Run of the *workshop_packages_to_install.R* script provided with the material.

PRESENTERS

Niccolò Samboglis works as a data scientist at InfoCamere, the IT Consortium of the Italian Chambers of Commerce. He is currently completing his PhD in Health Economics at City, University of London. Niccolò has experience in the field of data science applied to the healthcare sector. Niccolò is a co-organizer of Data Beers Padua.

Cristina Muschitiello works as Data Scientist at InfoCamere, the IT Consortium of the Italian Chambers of Commerce and is, also, researcher on agro-environmental topics in affiliation with the Council for Agricultural Research and Agricultural Economy Analysis (CREA), Research Unit for Cropping Systems in Dry Environments (CRA-SCA). Phd in Methodological Statistics at the University of Bari, Cristina was previously R-developer at the Statistics Division (ESS) of FAO (Food and Agriculture Organization of the United Nations) and author of the algorithm for the development of World's Food Balance Sheet. In less recent years she was lecturer in Statistics at the University of Bari.