

Automated report production with RStudio

R is a powerful tool for data analysis and, when coupled with RStudio, the reproducibility feature of coding is further enhanced. Using RStudio and literate programming it is possible to automate the production of reports, replacing tedious and irreproducible practices of copying and pasting.

In this 1-day course (ca. 6 hours), based on a case-study, we combine data from different sources to produce four reports on HIV prevalence, one for each of the four continents, on a given year. We learn how the reports can be produced automatically when data for a new year arrives. All of this without leaving the RStudio environment. As a bonus we learn how to use maps in R.

In this course you will

- further your knowledge of R markdown: You will further understand the YAML header of a R Markdown document and how to specify parameters that the document will depend on;
- use functional programming: you will learn how to use the function(al)s `map()` and `walk()` from the package `purrr` to apply a function to the entries of a vector or a list. This will be useful when we read the data in and also to generate all the reports, one per continent, at once;
- further your coding skills: use conditional `if`-statements to add information to the report depending on the current parameters. `ifelse` statements are also used;
- build on your `ggplot2` skills by learning how to produce maps with statistical information in them;
- get started with writing functions in R: you will be able to write a simple function;
- learn how to present a table.

Pre-requisites

The following are pre-requisites:

- Know how to import data in Excel format into R (package `readxl`).
- Understand the concept of tidy data and be confident with the use of the packages `tidyr` and `dplyr` for data wrangling.
- Know how to use the package `ggplot2` for data visualisation.
- Be familiar and at ease with literate programming in RStudio.

All pre-requisites are taught at the required level in “Data Science with R” <https://github.com/datasciencecampus/DSWR>

Intended audience

This course is of intermediate level and it is intended for practitioners working in the UK government who have an interest in applying the taught material to their day-to-day work.