

# Data Science with R

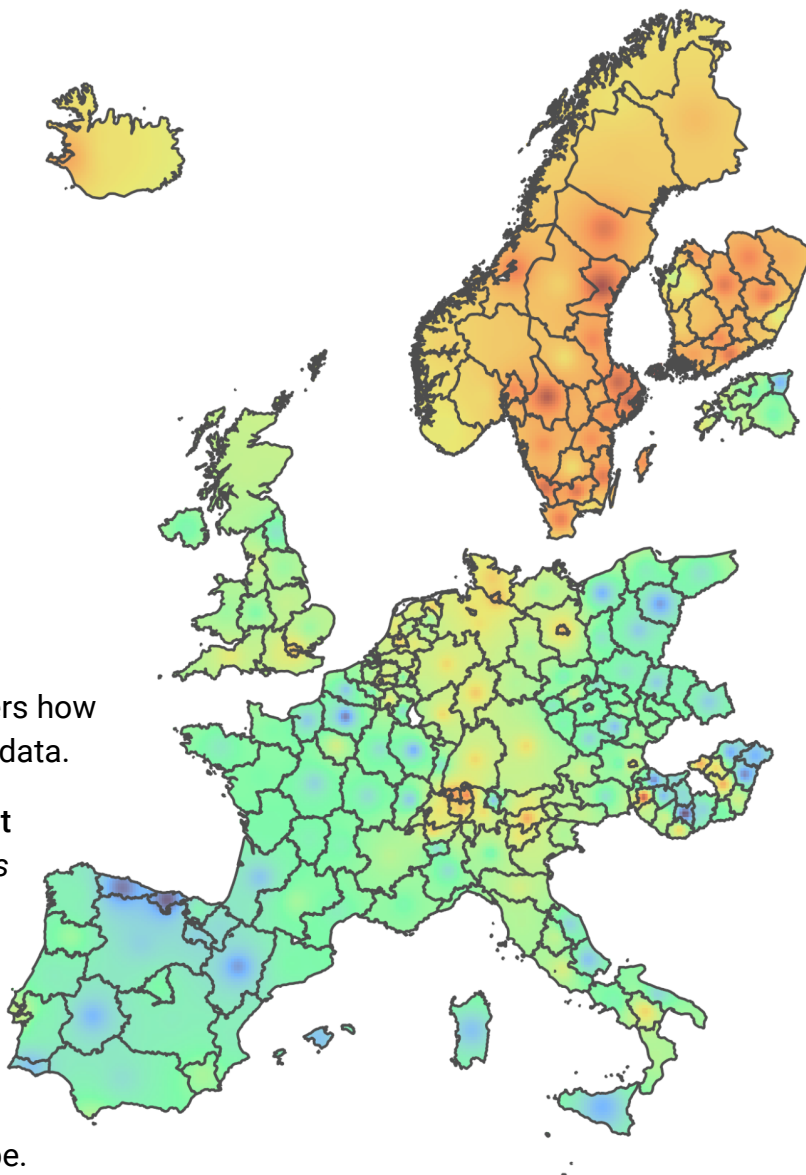
François Briatte

Spring 2023

This workshop is about analysing social science data with [R](#) and [RStudio](#). It covers how to access, wrangle, visualise and model data.

This workshop does not require any past experience with data science, but it *does* require some prior exposure to basic, introductory statistics, as well as an ability to use a laptop for proper work.

You will be surveyed on your current skills when the class starts, and will be pointed to readings to catch up if need be.



## System requirements

This workshop requires a laptop, running a recent operating system on which you have full admin privileges. Tablets will *not* do. 'Thin clients' like Chromebooks *might*.

Every class requires connecting to the Wi-Fi, using Google Drive to download files, and possibly installing software. You will also need a fully charged laptop battery, as well as enough disk space and live memory to host and run the course material.

For the duration of the course, I recommend working from your Desktop and outside of 'cloud' services like Apple iCloud.

## Teaching material

All course material will reach you through Google Drive and through weekly emails. Additional material for this course are available at [github.com/briatte/dsr](https://github.com/briatte/dsr) and on the course wiki at [github.com/briatte/dsr/wiki](https://github.com/briatte/dsr/wiki).

## Course sessions

The outline below is subject to change if we need to slow down to spend more time on correcting exercises and discussing statistical points.

1. **Software**                      Setting up R and RStudio
2. **Workflows**                    Working with files, folders and objects
3. **Data**                            Accessing data in various formats
4. **Visualisation**                Building plots for exploratory data analysis
5. **Description**                    Descriptive statistics and distributions
6. **Association**                    Statistical tests for means and proportions
7. **Correlation**                    Exploring linear (and nonlinear) relationships
8. **Regression**                    Ordinary Least Squares, a.k.a. linear regression
9. **Nonlinearity**                 Logistic regression and generalised linear models
10. **Surveys**                      Working with survey data and survey weights
11. **Classification**                An introduction to clustering and partitioning
12. **Extensions**                 Working with text, maps, and additional tools

## Course workload

This syllabus does not go into any specifics, in order for me to be able to adapt how the course will unfold for your class. Expect, however, the following to happen.

- You will be **reading from R-focused and statistics-focused handbooks** every week. I will also provide **video** material to watch, and additional resources that will help you achieve common data science tasks with R.
- You will be asked to **form groups** and to **work on weekly group exercises**, some of which will be graded. Deadlines will be set in class, and instructions will be discussed at the beginning of every session.
- You will be required to **regularly provide feedback** on how the course is working for you as we go, and to watch for what quantitative and statistical skills are asked from you in the internship offers that you consider applying to.

The course does not end with a final paper. Instead, you will be asked to complete a final exercise, and to deliver a one-page report on your experience with learning data science using R and the course material.

## Course handbooks

Gerring and Christenson. 2017. [Applied Social Science Methodology](#).

Hanck et al. 2023. [Introduction to Econometrics with R](#). **Free online**

Healy. 2019. [Data Visualization](#). **Free online**

Irizarry. 2022. [Introduction to Data Science](#). **Free online**

Ismay and Kim. 2023. [Statistical Inference via Data Science](#). **Free online**

Imai. 2018. [Quantitative Social Science](#).

Llaudet and Imai. 2022. [Data Analysis for Social Science](#).

Li, Q. 2021. [Using R for Data Analysis in Social Sciences](#).

Rodrigues, [Modern R with the tidyverse](#). **Free online**

Sanchez and Marzban. 2020. [All Models Are Wrong](#). **Free online**

Wickham et al. 2022. [R for Data Science](#). 2nd ed. **Free online**

## Additional resources

Below are some of the resources that I will recommend on top of the readings. More are listed in the course slides and at [github.com/briatte/dsr/wiki](https://github.com/briatte/dsr/wiki).

Bail, [SICSS Boot Camp](#). **Videos**

Briatte. [Quantitative Social Science Data](#).

CRAN. [Task Views](#) · see e.g. [econometrics](#), [spatial analysis](#), and [Web technologies](#).

Pew Research Center. [Survey Methods 101](#). **Videos**

Roodjuin et al. [Basic and Inferential Statistics](#). **Videos**

Stack Overflow. [Questions tagged with 'R'](#).

RStudio (now Posit). [Cheatsheets](#) · see e.g. [data wrangling](#).

## Credits

Thanks go to [Kim Antunez](#), [Timothée Gidoin](#), [Joël Gombin](#), [Emiliano Grossman](#), [Ivaylo D. Petev](#) and [Jan Rovny](#), for the reasons listed [elsewhere](#).