

# Quantitative Text Analysis using R and quanteda

**Stefan Müller (University College Dublin)  
University of Hamburg, 28<sup>th</sup> May - 2<sup>nd</sup> June 2021.**

*Designed for 15–20 students, this course will be delivered over five days from 28 June until 2 July. The course is a thorough survey of quantitative text analysis with practical implementation using the quanteda R package.*

It starts with a comprehensive overview of text analysis and working with textual data, including importing texts and constructing a corpus; tokenization; creating a document-feature matrix (days 1 and 2). Day 3 covers textual statistics, computation of similarity and distance, and visual methods for comparing texts. Day 4 and 5 will cover more advanced methods such as scaling, classification, and topic modelling. All topics will be accompanied by hands-on exercises in R, done on the students' own computers, but using starter code supplied by the instructors. The course takes place from 10:00am–12am and 1:00pm–2:35pm (Monday–Friday).

No prior experience with quantitative text analysis is needed, although the course presumes some familiarity with the R language.

Students can prepare by making sure they have prepared the following or read the suggested reading below:

- Installed R (at least version 4.0)
- Installed RStudio Desktop
- Familiarised yourselves with cloning a Git repo and using RMarkdown.
- Installed the following R packages: quanteda, quanteda.textmodels, quanteda.textstats, quanteda.textplots, readtext

## Reading List:

Benoit, Kenneth. 2020. "Text as Data: An Overview." In Curini, Luigi and Robert Franzese, eds. Handbook of Research Methods in Political Science and International Relations. Thousand Oaks: Sage. pp. 461-497. ([https://kenbenoit.net/pdfs/CURINI\\_FRANZESE\\_Ch26.pdf](https://kenbenoit.net/pdfs/CURINI_FRANZESE_Ch26.pdf)).

An Introduction to RMarkdown (<https://rmarkdown.rstudio.com/lesson-1.html>).

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." Political Analysis 21(3): 267–297 (<https://doi.org/10.1093/pan/mps028>).