Text Mining in R

Want to learn how to get the most out of text data? Today, a lot of data produced contains unstructured text, which can be difficult to transform and analyse without the correct knowledge and tools. In this course you will learn the basics of manipulating and transforming text data as well as how to extract meaning and sentiment in R, using packages such as {stringr} and {tidytext}.

Course Outline

- Appreciating the benefits of text data
- Cleaning and extracting text with {stringr} and regular expressions
- Transforming and mining text with {tidytext}
- Analysing the sentiment of text
- Understanding the content of a text with TF-IDF

Learning Outcomes

Session 1:

By the end of session 1 participants will ...

- be able to clean, manipulate, and transform text data using the {stringr} package.
- use basic regular expressions to extract and remove patterns in text.
- convert unstructured text data into a tidy format suitable for analysis with {tidytext}.
- understand basic text mining concepts, such as tokenization and stop words.

Session 2:

By the end of session 2 participants will ...

- create beautiful plots of text data including word clouds.
- be able to analyse the sentiment of a piece of text and compare sentiment across texts and over time.
- understand how to extract representative words of a text to classify its content.
- be able to understand and perform lemmatization and stemming using {textstem}.

This course does not include:

- A detailed use of the {tidyverse} packages used for data wrangling. For comprehensive coverage of the {tidyverse} see our Data Wrangling in Tidyverse course.

Attendee Feedback

- "Clear explanation of examples."
- "Materials very clear and will be useful in future. Course leader very easy to listen to."