Master Class: Working with unconventional data using R

Working with text and a little bit of maps

Valeria Rueda
School of Economics, University of Nottingham
valeria.rueda@nottingham.ac.uk

Learning goals

At the end of the session, you should…
- Have jumped in the R pool
- Familiarized yourself with regular expressions and their use
- Created one map with `ggplot`
- Getting started with text pre-processing for text analysis: stemming, cleaning
- Getting started with TF-IDF Representations

Structure (will be updated)

1. Regular Expressions
2. Working with text
    a. Processing and tokenizing
    b. Representing text as data: TF-IDF representations

Prerequisites
Materials

Laptops with a working version of R installed. To work with R, RStudio or Jupyter are recommended. RStudio may be the easiest to sort out.

Make sure you have tested your version of R and basic commands, such as read.csv(), are running before the lecture. This video presents a tutorial on how to do load a dataset in R.

You should also preinstall the following packages: `tidyverse, ggmap`.

To install a package, run the command: `install.package("package".` For instance: `install.package("tidyverse").`

Prior knowledge

This is a hands-on class: it is structured around activities to be done by students during the session. We will discuss all challenges faced.

NO prior knowledge of R is required. This class can be taken as an opportunity to try R for the first time.