Max Weber Multidisciplinary Research Workshop – 19th May 2016

"Introduction to text analysis"

Speaker: Dr Pablo Barberà (New York University)

Description

The popularity of text as data is increasing rapidly within the social sciences. "Scholars have long recognized this, but the massive costs of analyzing even moderately sized collections of texts have hindered their use in political science research" (Grimmer and Stewart 2013) and elsewhere in the social sciences. This situation has changed with increasing computing power and more capable computing tools. In the coming years, the relevance of text data will further increase as more and more human communication is recorded online.

This workshop provides an introduction to text analysis using R. We will cover methods to conduct quantitative analysis of textual and web data, with an emphasis on social media data, applied to the study of social science questions. The workshop is made up of three "modules", each consisting of an introduction to a topic followed by examples and applications using R. The first module will cover how to format and input source texts, how to prepare the data for analysis, and how to extract descriptive statistics. The second module will discuss automated classification of text sources into categories using dictionary methods and supervised learning. Finally, the third module will discuss unsupervised classification of text into categories using topic modeling.

This workshop will be led by Pablo Barberá. He is currently a Moore-Sloan Fellow at New York University and will join the University of Southern California as Assistant Professor in July. His primary research interests include social media and politics, quantitative political methodology, electoral behavior, and political representation. Pablo Barberá is the author of several R packages to collect and analyze social media data using R, regularly analyzes text for his research, and has taught a number of courses in this area.

Program

10:00 am - 12:00 pm – Morning session

12:00 am - 13:30 pm – Lunch

13:30 am - 15:30 pm – Afternoon session I

15:30 am - 16:00 pm – Coffee break

16:00 am - 18:00 pm – Afternoon session II