

LondonR WORKSHOP

Web Scraping and Text Analysis in R



Date: Tuesday 28th March 2017
Time: 3pm – 5.30pm
Venue: UCL Darwin Lecture Theatre (Lower Floor), Darwin Building, Malet Place, London, WC1E 7JG (for directions please see overleaf)
Level: Beginner - with basic level of R
Core Skill: Data Wrangling

This workshop is aimed at users with a basic knowledge of the R language. The workshop is interactive and will include exercises. Participants will require a laptop

Overview:

Web scraping and text analysis can be intimidating topics to approach for the first time. The aims of this workshop are twofold; firstly to introduce participants to web scraping in R using the **rvest** package, to import data easily and quickly into R. Secondly, this workshop will examine some of the fundamentals of text mining, and show how the **tidytext** package can be used to simplify this process. Following this session, attendees should be able to perform basic web scraping, be able to transform the scraped data into a useful format, and perform text mining analyses including exploring relationships between words and sentiment analysis, generating compelling visualisations of this data. This workshop is aimed at users with a basic knowledge of the R language.

Prerequisites:

- Participants should have a working knowledge of the R language having completed the Mango Solutions' Introduction to R or similar.
- Basic knowledge of HTML is beneficial but not necessary.
- Installation information will be sent out separately ahead of the workshop

Outline

Web scraping in R

- Retrieving information from webpages
- Parsing XML

Analysing text data

- Summarising text data
- Sentiment analysis using tidytext
- ngrams and relationships between words

Visualising text data

- Word clouds
- Visualising related words

Registration

Due to the limitations of the venue places on this workshop are limited; to register please email londonR@mango-solutions.com for confirmation of your place.

Once the registration list is full we will operate a waitlist process.

Directions: Malet Place is accessed via the Torrington Place entrance (off Gower Street). Closest tube stations are Euston Square & Russell Square. As you enter Malet Place, Darwin Lecture Hall entrance is on the left between The Makery & The Petrie Museum.