

Introduction to Regular Expressions

Are you trying to search for or clean up certain segments of text in your documents or data files, but it's taking hours to search and replace, because you have to type many slightly different versions of the keywords you're searching for? Perhaps you need to replace month names with just the first three letters, but you don't want to do this 12 times (once for each month), for lowercase, capitalized first letter, and all-caps versions? Or perhaps you only want to do so when the month name is part of a full date expression, e.g. "March 2, 2001", but not in the phrase "they're planning a march on Wall Street".

Introducing regular expressions, special text strings that allow you to define text patterns in flexible and complex ways. Regular expressions are very useful in many applications. They can be used to search for and replace text in a simple text editor, or in your datasets in Python or R. In this workshop, we'll learn how to build regular expressions using special characters, sets or quantities of characters, the absence of certain characters, this-or-that alternatives, and other options. Then we'll briefly demonstrate how to use these expressions in Python and R, to get you familiar with the package you'll use in your own programming.

Technology requirements: Laptop required. Most of the workshop will be done using a web browser, so you will need an up-to-date browser. We recommend Google Chrome. In addition, you will need the Sublime text editor. Follow [this link \(https://www.sublimetext.com/3\)](https://www.sublimetext.com/3) to download and install the version for your operating system.

Optional: If you are already programming in Python or R and want to try out regular expressions in that language, we'll have a brief demonstration of each language's regex package at the end of the workshop. So bring your laptop with your preferred Python or R environment installed. This part is optional; we will not discuss how to begin programming in those languages if you have not already done so.