

Regex

Objective: Provide a brief intro to Regex

Slides on Github

What is Regex?

- Aka Regular Expressions
- Helps define patterns for matching, extracting, and/or transforming data
- Available in many major programming languages (Python, R, Java, Javascript, etc.)

Basic Concepts

- Character set [...]
 - Matches any single character
- Optional element ...?
 - 0 or 1 occurrences
- Range [0-9]
 - Matches any single character in range
- (Kleene) Star/Asterisk ...*
 - Matches 0 or more occurrences
- Plus sign...+
 - Matches 1 or more occurrences

Basic Concepts

- Dot .
 - Matches any character (except line breaks)
- Escape Character \...
 - Take next character literally (no special meaning)
- Alternation ... | ...
 - Boolean “or”
- Exact quantifiers {n} {a, b} {a, }
- Capturing group (...)
 - Group multiple tokens; capture group for backreference

Character Classes

- Character Classes
- `.` matches any character except newline
- `\w\d\s` word, digit, whitespace character
- `\W\D\S` non-word, non-digit, non-whitespace character
- `[abc]` a b or c
- `[^abc]` any character other than a b or c
- `[a-g]` any char between a and g (inclusive)

Anchors

- `^abc` start of string
- `abc$` end of string
- `Xyz\b` word boundary
- `xyz\B` not word boundary

Go and try Regex!

- Interactive learning:
 - <https://regexone.com/>
 - <https://librarycarpentry.org/lc-data-intro/>
- Read a book if you need to:
 - https://eloquentjavascript.net/09_regex.html
- Practice and use it with your programming language:
 - <https://regexcrossword.com/>