REGEX CHEATSHEET ®R

EVERYTHING YOU NEED TO KNOW WHEN WORKING WITH REGULAR EXPRESSIONS

Repetitions of a pattern

*	at least 0 times	
+	at least 1 time	
?	at most 1 time	
{n}	exactly n times	
{n,}	at least n times	
{n, m}	at least n times, at most m times	

Examples in R

abc* "ab" followed by zero or more "c"
abc+ "ab" followed by one or more "c"
abc? "ab" followed by zero or one "c"
abc{2} "ab" followed by 2 "c"
abc{3,} "ab" followed by at least 3 "c"
abc{3,5} "ab" followed by 3 up to 5 "c"

Positions of a pattern

^	the start of the string
$	the end of the string
\b	empty string at edge of a word
\B	empty string not at edge of a word

^starts string starting with "starts"
ends$ string ending with "ends"
\\bab matches "abc" but not "dabc"
\\Bab matches "dabc" but not "abc"

Character classes I

\d	digits
\D	non-digits
\w	word characters (incl. underscore)
\W	non-word characters
\s	space
\S	non-space

3\\d{10} 11-digit strings starting with "3"
a\\D matches "ab" but not "a1"
ab\\w matches "abc" & "ab_" but not "ab."
ab\\W matches "ab." but not "abc"
ab\\sc matches "ab c" but not "abc"
a\\Sc matches "abc" but not "a c"

Character classes II

[:digit:] or [0-9]	digits
[:lower:] or [a-z]	lowercase letters
[:upper:] or [A-Z]	uppercase letters
[:alpha:] or [A-z]	alphabetic characters
[:alnum:]	alphanumeric characters
[:xdigit:]	hexadecimal characters (base 16)
[:blank:]	blank characters (space & tab only)
[:space:]	space characters*
[:punct:]	punctuation characters
[:graph:]	graphical characters
[:print:]	printable characters

ab[[:digit:]]+ matches "ab3" & "ab221"
ab[[:lower:]] matches "abc" but not "abC"
ab[[:upper:]] matches "abC" but not "abc"
ab[[:alpha:]] matches "abc" & "abC"
ab[[:alnum:]] matches "ab3", "abc" & "abC"
ab[[:xdigit:]] matches "abf" but not "abg"
ab[[:blank:]]c matches "ab c" & "ab\tc" but not "abc"
ab[[:space:]]c matches "ab c", "ab\tc" & "ab\nc"
ab[[:punct:]] matches "ab!" and "ab@"
ab[[:graph]:]+ matches "ab!", "abc!" & "ab3@"
ab[[:print:]]+ matches "ab 3", "ab c!" & "abc"

Others

.	any single character
[...]	character list
[^...]	inverted character list
\	escape character
		"or" operator
(...)	grouping
\\N	backreference to Nth group

abc. "abc" followed by any single character
ab[0-9]+ matches "ab3" & "ab221"
ab[^0-9]+ matches "abcd", but not "ab3d"
ab* "ab" followed by "*"
a(b|c) "a" followed by "b" or "c"
a(bcd) "a" followed by sequence "bcd"
(co)\\1 matches "coconut" but not "count"

*Space characters are tab (\t), newline (\n), vertical tab (\v), carriage return (\r), form feed (\f), and space