

Regular expressions

STRING MANIPULATION WITH STRINGR IN R



Charlotte Wickham

Assistant Professor at Oregon State
University

Regular expressions

- A language for describing patterns

`^\d+`

- "the start of the string, followed by any single character, followed by one or more digits"

Regular expressions as a pattern argument

```
str_detect(c("R2-D2", "C-3P0"), pattern = "^\\.\\d+")
```

```
TRUE FALSE
```

```
START %R%  
  ANY_CHAR %R%  
  one_or_more(DGT)
```

```
<regex> ^\\.\\d+
```

rebus

```
START %R%  
  ANY_CHAR %R%  
  one_or_more(DGT)
```

Regular expression

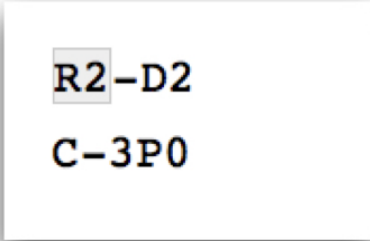
```
^\\.\\d+
```

Regular expressions as a pattern argument

```
str_detect(c("R2-D2", "C-3P0"),  
  pattern = START %R%  
            ANY_CHAR %R%  
            one_or_more(DGT))
```

TRUE FALSE

```
str_view(c("R2-D2", "C-3P0"),  
  pattern = START %R%  
            ANY_CHAR %R%  
            one_or_more(DGT))
```



R2-D2
C-3P0

[In HTML viewer](#)

Let's practice!

STRING MANIPULATION WITH STRINGR IN R

More regular expressions

STRING MANIPULATION WITH STRINGR IN R



Charlotte Wickham

Assistant Professor at Oregon State University

Regular expression review

Pattern	Regular Expression	rebus
Start of string	^	START
End of string	\$	END
Any single character	.	ANY_CHAR

Regular expression review

Pattern	Regular Expression	rebus
Start of string	^	START
End of string	\$	END
Any single character	.	ANY_CHAR

Regular expression review

Pattern	Regular Expression	rebus
Start of string	^	START
End of string	\$	END
Any single character	.	ANY_CHAR

Regular expression review

Pattern	Regular Expression	rebus
Start of string	^	START
End of string	\$	END
Any single character	.	ANY_CHAR

Regular expression review

Pattern	Regular Expression	rebus
Start of string	^	START
End of string	\$	END
Any single character	.	ANY_CHAR

Regular expression review

Pattern	Regular Expression	rebus
Start of string	<code>^</code>	START
End of string	<code>\$</code>	END
Any single character	<code>.</code>	ANY_CHAR
Literal dot, carat or dollar sign	<code>\. \^ \\$</code>	DOT, CARAT, DOLLAR

Alternation

```
(dog|cat)
```

```
or("dog", "cat")
```

```
<regex> (?:dog|cat)`
```

```
str_view(c("kittycat", "doggone"),  
         pattern = or("dog", "cat"))
```

```
kittycat  
doggone
```

Character classes

```
char_class("Aa")
```

```
<regex> [Aa]
```

```
str_view(c("apple", "Aaron"),  
         pattern = char_class("Aa"))
```

```
negated_char_class("Aa")
```

```
<regex> [^Aa]
```

```
str_view(c("apple", "Aaron"),  
         pattern = negated_char_class("Aa"))
```

aapple

Aaron

apple

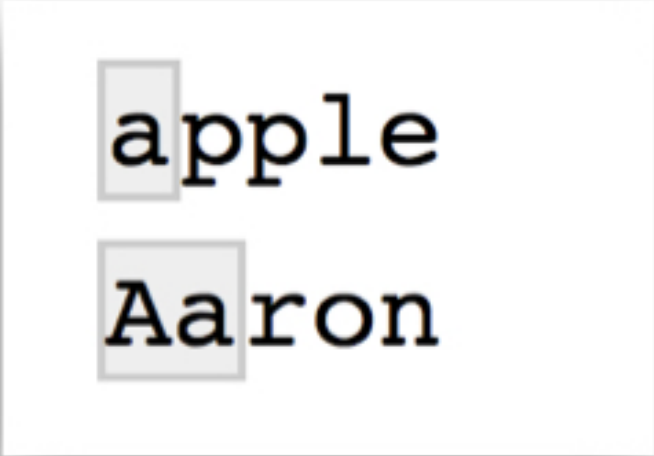
Aaron

Repetition

Pattern	Regular Expression	rebus
Optional	?	optional()
Zero or more	*	zero_or_more()
One or more	+	one_or_more()
Between n and m times	{n}{m}	repeated()

Repetition

```
str_view(c("apple", "Aaron"),  
         pattern = one_or_more("Aa"))
```



apple
Aaron

Let's practice!

STRING MANIPULATION WITH STRINGR IN R

Shortcuts

STRING MANIPULATION WITH STRINGR IN R



Charlotte Wickham

Assistant Professor at Oregon State
University

Ranges in character classes

```
DOLLAR %R% char_class("0123456789")
```

```
<regex> \\$[0123456789]
```

A digit

```
char_class("0-9")
```

```
<regex> [0-9]
```

A lower case letter

```
char_class("a-z")
```

```
<regex> [a-z]
```

An upper case letter

```
char_class("A-Z")
```

```
<regex> [A-Z]
```

Shortcuts

DGT # A digit -->

```
char_class("0-9")
```

```
<regex> \d
```

```
<regex> [0-9]
```

WRD # A word character -->

```
char_class("a-zA-z0-9_")
```

```
<regex> \w
```

```
<regex> [a-zA-z0-9_]
```

SPC # A whitespace character

```
<regex> \s
```

National Electronic Injury Surveillance System (NEISS)

- neiss package <https://github.com/hadley/neiss>
- Injuries reported in ER of random sample of hospitals

19YOM-SHOULDER STRAIN-WAS TACKLED WHILE
PLAYING FOOTBALL W/ FRIENDS

National Electronic Injury Surveillance System (NEISS)

- neiss package <https://github.com/hadley/neiss>
- Injuries reported in ER of random sample of hospitals

19YOM-SHOULDER STRAIN-WAS TACKLED WHILE
PLAYING FOOTBALL W/ FRIENDS

19 year old male

Let's practice!

STRING MANIPULATION WITH STRINGR IN R