

Worksheet 8 - Regular Expressions I

Monday, September 29, 2025

DS 002R - Jo Hardin

Your Name: _____

Names of people you worked with: _____

What is the most challenging thing you have to do in the next 10 days?

Task:

The Jewish Festival of Lights is written as **חנוכה** in Hebrew and translated to many different spellings in English. The correct English spelling of the holiday can be given in any of the following thirteen variants, according to <https://www.holidays.net/chanukah/spelling.htm>.

```
lights <- c("Chanuka", "Chanukah", "Chanukkah", "Channukah", "Hanukah",  
           "Hannukah", "Hanukkah", "Hanuka", "Hanukka", "Hanaka", "Haneka",  
           "Hanika", "Khanukkah")
```

```
lights
```

```
[1] "Chanuka" "Chanukah" "Chanukkah" "Channukah" "Hanukah"  
[6] "Hannukah" "Hanukkah" "Hanuka" "Hanukka" "Hanaka"  
[11] "Haneka" "Hanika" "Khanukkah"
```

Provide a regular expression which, when used in `str_extract()`, will output all thirteen variants but not the string “Christmas”.

Solution:

`[CHK]h?ann?[aeiu]kk?ah?`

Here are few different variants:

```
str_extract(c(lights, "Christmas"), "[CHK]h?ann?[aeiu]kk?ah?")
```

```
[1] "Chanuka"    "Chanukah"  "Chanukkah" "Channukah" "Hanukah"
[6] "Hannukah"  "Hanukkah"  "Hanuka"    "Hanukka"   "Hanaka"
[11] "Haneka"    "Hanika"    "Khanukkah" NA
```

```
str_extract(c(lights, "Christmas"), "[CK]?[Hh]ann?[aeiu]kk?ah?")
```

```
[1] "Chanuka"    "Chanukah"  "Chanukkah" "Channukah" "Hanukah"
[6] "Hannukah"  "Hanukkah"  "Hanuka"    "Hanukka"   "Hanaka"
[11] "Haneka"    "Hanika"    "Khanukkah" NA
```

```
str_extract(c(lights, "Christmas"), "[CHK]h?an{1,2}[aeiu]k{1,2}ah?")
```

```
[1] "Chanuka"    "Chanukah"  "Chanukkah" "Channukah" "Hanukah"
[6] "Hannukah"  "Hanukkah"  "Hanuka"    "Hanukka"   "Hanaka"
[11] "Haneka"    "Hanika"    "Khanukkah" NA
```

Note that if you don't use the entire matching regular expression, you won't get the full variants in return.

```
str_extract(c(lights, "Christmas"), "[CHK]h?an{1,2}")
```

```
[1] "Chan"  "Chan"  "Chan"  "Chann" "Han"   "Hann"  "Han"
[8] "Han"   "Han"   "Han"   "Han"   "Han"   "Khan"  NA
```