Assignment 4

M. Hill

10/21/2024

Exercise 1

a) This regular expression matches: Strings that contain 'ab' in that order only.

```
strings <- c('ab','acb','cab','bac')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, 'ab') )
```

```
##   string result
## 1     ab   TRUE
## 2    acb  FALSE
## 3    cab   TRUE
## 4    bac  FALSE
```

b) This regular expression matches: Strings that contain 'a' or 'b'.

```
strings <- c('bat','did','goa','less')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '[ab]') )
```

```
##   string result
## 1    bat   TRUE
## 2    did  FALSE
## 3    goa   TRUE
## 4   less  FALSE
```

c) This regular expression matches: Strings that contain 'a' or 'b' at the beginning.

```
strings <- c('boy','goal','apple','labs')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '^[ab]') )
```

```
##   string result
## 1    boy   TRUE
## 2   goal  FALSE
## 3  apple   TRUE
## 4   labs  FALSE
```

d) This regular expression matches: Strings that contains one or more repetition of a digit, then a whitespace, then 'a' or 'A'.

```
strings <- c('525 A', 'aa a', '2 a','2A')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '\\d+\\s[aA]') )
```

```
##   string result
## 1  525 A   TRUE
```

```
## 2   aa a  FALSE
## 3    2 a   TRUE
## 4     2A  FALSE
```

e) This regular expression matches: Strings that contain one or more repetitions of a digit, then zero or more repetitions of a white space, then a or A.

```
strings <- c('HOUSE:3A','HOUSE:3 ',' NO 3 HOUSE:3A','4davros A')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '\\d+\\s*[aA]') )
```

```
##           string result
## 1       HOUSE:3A   TRUE
## 2       HOUSE:3   FALSE
## 3  NO 3 HOUSE:3A   TRUE
## 4      4davros A  FALSE
```

f) This regular expression matches: Strings that contain zero or more repetitions of any character.

```
strings <- c('','superpowerflowerpollen2','45','&&')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '.*') )
```

```
##                    string result
## 1                            TRUE
## 2 superpowerflowerpollen2   TRUE
## 3                      45   TRUE
## 4                      &&   TRUE
```

g) This regular expression matches: Strings that start with two repetitions of any alphanumeric character and 'bar' exactly

```
strings <- c('22bar','Aebar','AAbaer','barYES!')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '^\\w{2}bar') )
```

```
##    string result
## 1   22bar   TRUE
## 2   Aebar   TRUE
## 3  AAbaer  FALSE
## 4 barYES!  FALSE
```

h) This regular expression matches: Strings that contain the group contain 'foo', then a period, followed by 'bar', or the string starts with two repetitions of any alphanumeric character and 'bar' exactly.

```
strings <- c('AAfoo.bar','anaownwko foobar','76barfoo.9','A2foobar')
data.frame( string = strings ) %>%
  mutate( result = str_detect(string, '(foo\\.bar)|(^\\w{2}bar)') )
```

```
##             string result
## 1        AAfoo.bar   TRUE
## 2 anaownwko foobar  FALSE
## 3       76barfoo.9   TRUE
## 4         A2foobar  FALSE
```

Exercise 2

```
file.names <- c( 'S123.P2.C10_20120621_213422.jpg',
                 'S10.P1.C1_20120622_050148.jpg',
                 'S187.P2.C2_20120702_023501.jpg')
temp <- data.frame(site = str_extract(file.names, 'S\\d+'),
    plot = str_extract(file.names, 'P\\d+')
  )

temp
```

```
##   site plot
## 1 S123   P2
## 2  S10   P1
## 3 S187   P2
```