230924 DATA 607 Assignment 3

Molly Siebecker

2023-09-24

#1. Using the 173 majors listed in fivethirtyeight.com's College Majors dataset, provide code that identifies the majors that contain either "DATA" or "STATISTICS"

```
majors_url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/college-majors/majors-list
majors_list <- read.csv(majors_url, header = TRUE, sep = ",")
```

```
str_subset(majors_list$Major, "DATA|STATISTICS")
```

```
## [1] "MANAGEMENT INFORMATION SYSTEMS AND STATISTICS"
## [2] "COMPUTER PROGRAMMING AND DATA PROCESSING"
## [3] "STATISTICS AND DECISION SCIENCE"
```

#2.

```
fruit_string <- '"bell pepper" "bilberry" "blackberry" "blood orange" "blueberry" "cantaloupe" "chili p
```

```
fruit_split <- str_split_1(fruit_string, pattern = '^"[:alpha:]+"$|^"[:alpha:]+ [:alpha:]+"$')
```

```
fruit_split_words <- str_remove_all(fruit_split, pattern = '\\\\')
```

```
fruit_df <- data.frame(fruit_split_words)
colnames(fruit_df) <- c("Fruit")
```

```
fruit_df %>%
  separate_longer_delim(Fruit, delim = '" "')
```

```
##           Fruit
## 1  "bell pepper
## 2      bilberry
## 3    blackberry
## 4  blood orange
## 5     blueberry
## 6    cantaloupe
## 7  chili pepper
## 8    cloudberry
## 9    elderberry
## 10         lime
## 11       lychee
## 12     mulberry
## 13        olive
## 14 salal berry"
```

```
fruit_vector <- as.character(fruit_df$Fruit)
print(fruit_vector)
```

```
## [1] "\"bell pepper\" \"bilberry\" \"blackberry\" \"blood orange\" \"blueberry\" \"cantaloupe\" \"chi
```

#3. Describe, in words, what these expressions will match:

(.)\1\1 This regular expression will match any sequence of the same letter three times in a row (ie, "aaa" and "bbb", but not "aaba").

"(.)(.)\2\1" This string that defines a regular expression will match any sequence of four letters in which the third character matches the second, and the fourth character matches the first (ie, "ABBA", "noon", and "illi" in "million"). It is a four-letter palindromic sequence that may or may not be contained within a longer string.

(..)\1 This regular expression will match the same pair of letters twice in a row (ie, "abab", "anan" in "banana", and "emem" in "remember").

"(.).\1.\1" This string that defines a regular expression will match any sequence of five letters in which the first, third, and fifth characters are the same (ie, "eleve" in "eleven").

"(.)(.)(.).*\3\2\1" This string that defines a regular expression will match any sequence of at least six characters in which the last letter matches the first, the penultimate matches the second, and the third-to-last matches the third, with any number of characters in between the third and the third-to-last (ie, "paragrap" in "paragraph").

#4. Construct regular expressions to match words that:

Start and end with the same character: ^(.).*\1$ Example:

```
str_view(words, "^(.).*\\1$")
```

```
##  [36] | <america>
##  [49] | <area>
## [209] | <dad>
## [213] | <dead>
## [223] | <depend>
## [258] | <educate>
## [266] | <else>
## [268] | <encourage>
## [270] | <engine>
## [278] | <europe>
## [283] | <evidence>
## [285] | <example>
## [287] | <excuse>
## [288] | <exercise>
## [291] | <expense>
## [292] | <experience>
## [296] | <eye>
## [386] | <health>
## [394] | <high>
## [450] | <knock>
## ... and 16 more
```

Contain a repeated pair of letters (e.g. "church" contains "ch" repeated twice.): (..).*\1 Example:

```
str_view(words, "(..).*\\1")
```

```
##  [48] | ap<propr>iate
## [152] | <church>
## [181] | c<ondition>
## [217] | <decide>
## [275] | <environmen>t
## [487] | l<ondon>
## [598] | pa<ragra>ph
## [603] | p<articular>
## [617] | <photograph>
## [638] | p<repare>
## [641] | p<ressure>
## [696] | r<emem>ber
## [698] | <repre>sent
## [699] | <require>
## [739] | <sense>
## [858] | the<refore>
## [903] | u<nderstand>
## [946] | w<hethe>r
```

Contain one letter repeated in at least three places (e.g. “eleven” contains three “e”s.): (.).*\1.*\1 Example:

```
str_view(words, "(.).*\\1.*\\1")
```

```
##  [48] | a<pprop>riate
##  [62] | <availa>ble
##  [86] | b<elieve>
##  [90] | b<etwee>n
## [119] | bu<siness>
## [221] | d<egree>
## [229] | diff<erence>
## [233] | di<scuss>
## [265] | <eleve>n
## [275] | e<nvironmen>t
## [283] | <evidence>
## [288] | <exercise>
## [291] | <expense>
## [292] | <experience>
## [423] | <indivi>dual
## [598] | p<aragra>ph
## [684] | r<eceive>
## [696] | r<emembe>r
## [698] | r<eprese>nt
## [845] | t<elephone>
## ... and 2 more
```