section 5

Masaoud AND Thu

March 2 and 3, 2023

Info

- Thu: OH on Tuesday 9-10:30am in SC 316.06. Section on Thursday 1:30-2:30 PM SC 705.
- Masaoud: OH on Tuesday 9-10:30pm in Kirkland Dhall. Section on Friday 12-1pm in Sever 209.

We will not have section not week (the week before spring break)! Good luck on your projects :)

We will now include a very high-level overview of this week's material. Please let us know what questions you have!

- **Factors** are used to impose structure on categorical variables. The **levels** of a factor are the different values that a factor can take on.
  - Some useful functions when working with factors are `fct_reorder()`, `fct_relevel()`, `fct_collapse()`, `fct_drop()`, and `fct_recode()`. There are others in lecture as well!
- We also learned how to manipulate strings with the package `stringr`. For example, you can make all letters in a string uppercase with `str_to_upper()`.
- On the topic of strings, sometimes we are interested in matching a particular pattern in a string. Kelly showed that it is not ideal to "hard code" the patterns we want to match (for example, if we want to match any numbers, we should not specify "0|1|2|...") because it's cumbersome and may fail to catch all the cases we want!
- This brings us to regular expressions, or regexes for short. Regexes are sequences of characters that specifies a pattern in a string, and they can be pretty general.
  - Feel free to use the physical "cheat sheet" from class. An online copy is also available here: https://evoldyn.gitlab.io/evomics-2018/ref-sheets/R_strings.pdf.

Now, we will practice using Regular Expressions! Use `str_view_all()` to identify the pattern with a Regular Expression.

Problem 1

For this problem we are going to use a subset of the character vector `fruit` that lives in `stringr`.

```
set.seed(4433)
fruit_s <- sample(fruit, 20)
```

a. Highlight the pattern "berry".

b. Now highlight all the fruits that contain the pattern "berry". (Don't just highlight the word "berry").

c. Highlight only fruits that contain "berry" and are one word.

d. Highlight any repeated characters (e.g. "oo", "ll").

e. Highlight fruits that start with a vowel.

f. Highlight fruits that don't start with a vowel.

g. Highlight the second to last character of fruits that end in a vowel.

Problem 2

For this problem, let's continue to use these lyrics about numbers:

```
lyrics <- c("But I would walk 500 miles",
            "2000 0 0 party over oops out of time!",
            "1 is the loneliest number that you'll ever do",
            "When I'm 64",
            "Where 2 and 2 always makes a 5")
```

a. Locate all words that start with "w" or "W".

b. Find all the non-numbers.

c. Locate all the words that are at least five letters long.

d. Locate all characters that appear after punctuation.

e. Locate all the places where there are two vowels in a row.

f. Find words that start with vowels.

Hint: Consider using `\\b` to find the start of the word.