Class Activity 11

Your name here

April 16 2024

Problem 1

Let's learn about combining strings with different separators first.

```
place <- "Central Park"
activity <- "jogging"
activities <- c("jogging", "picnicking", "boating")
my_sentence <- str_c(place, " is great for ", activity, ".", sep = "")
```

a. What happens when a `str_c` entry is a vector?

Answer:

```
my_sentences <- str_c(place, " is great for ", activities, ".", sep = "")
```

b. How do you combine strings with `str_glue`?

Answer:

c. What does `str_flatten` do?

Answer:

d. What will using a `\n` separator do in the command below?

Answer:

e. Does `str_length` count spaces and special characters??

Answer:

f. How do you count the number of `e`'s in a string?

Answer:

```
text <- "The quick brown fox jumps over the lazy dog."
pattern <- "e"
vowel_count <-
Error: <text>:4:0: unexpected end of input
2: pattern <- "e"
3: vowel_count <-
  ^
```

g. What happens with negative positions?

Answer:

h. How do you extract a `substring` with positive and negative positions?

Answer:

```
my_sentence <- "Central Park is great for jogging."
```

i. With a vector of positions?

Answer:

j. How do you extract multiple `substrings` using a vector of positions?

Answer:

Problem 2

a. Use the string parsing functions that you learned today to do tasks described in the comments below:

```
s1 <- "12%"  # remove %
s2 <- "New Jersey_*"  # remove _*
s3 <- "2,150"    # remove comma(,)
s4 <- "Learning #datascience is fun!"   # extract #datascience
s5 <- "123 Main St, Springfield, MA, 01101"   # separate info

# Cleaning steps
s1_clean <-
s2_clean <-
s3_clean <-
s4_clean <-
s5_clean <-

# Print cleaned strings
s1_clean
Error in eval(expr, envir, enclos): object 's1_clean' not found
s2_clean
Error in eval(expr, envir, enclos): object 's2_clean' not found
s3_clean
Error in eval(expr, envir, enclos): object 's3_clean' not found
s4_clean
Error in eval(expr, envir, enclos): object 's4_clean' not found
s5_clean
Error in eval(expr, envir, enclos): object 's5_clean' not found
```

Problem 3

a. Let's look at the following dataset containing information about movies and their release years. We'll extract the release year from the movie title, create a new column with decades, and count the number of movies in each decade.

```
# Sample dataset
movies <- tibble(
  title = c(
    "The Godfather (1972)", "Pulp Fiction (1994)", "The Dark Knight (2008)",
```

```
    "Forrest Gump (1994)", "The Shawshank Redemption (1994)", "The Matrix (1999)",
    "Inception (2010)", "Interstellar (2014)", "Parasite (2019)", "Fight Club (1999)"
  )
)
movies
# A tibble: 10 x 1
   title
   <chr>
 1 The Godfather (1972)
 2 Pulp Fiction (1994)
 3 The Dark Knight (2008)
 4 Forrest Gump (1994)
 5 The Shawshank Redemption (1994)
 6 The Matrix (1999)
 7 Inception (2010)
 8 Interstellar (2014)
 9 Parasite (2019)
10 Fight Club (1999)
```

```
# Processing the dataset
movies_processed <- movies %>%
  mutate(
    release_year = ),
    decade =
  ) %>%
  count()

# Print the processed dataset
movies_processed
Error: <text>:4:21: unexpected ','
3:   mutate(
4:     release_year = ),
                       ^
```