

# WU #7 - Strings

Monday, September 23, 2024

DS 002R - Jo Hardin

Name: \_\_\_\_\_

Names of people you worked with: \_\_\_\_\_

Briefly describe to the group your favorite aspect of the neighborhood where you grew up.

**Task:** Consider the following made up data on enrollments. Try as many of the following tasks as possible in a tidy pipeline. Use `str_*`() functions.

```
classes <- data.frame(
  sem      = c("SP2023", "FA2023", "SP2024"),
  area     = c("History", "Math", "Anthro"),
  enroll   = c("30 - people", "20 - people", "25 - people"),
  instructor = c("Ernesto Capello", "Lori Ziegelmeier", "Arjun Guneratne")
)

classes
```

	sem	area	enroll	instructor
1	SP2023	History	30 - people	Ernesto Capello
2	FA2023	Math	20 - people	Lori Ziegelmeier
3	SP2024	Anthro	25 - people	Arjun Guneratne

1. Define a new variable `num` that adds up the number of characters in the `area` column.
2. Change the areas to `history`, `math`, `anthro` instead of `History`, `Math`, `Anthro`. (Hint: there is an easier way to do it than `str_replace()`.)
3. Create a variable that identifies which courses were taught in spring.
4. Change the semester labels to “fall2023”, “spring2024”, “spring2023”.
5. In the `enroll` variable, change all e’s to 3’s (just because?)
6. Use `sem` to create 2 new variables: 1 with the semester (SP/FA) and 1 with the year.

## Solution:

```
# Define a new variable "num" that adds up the number of characters in the area label
classes |>
  mutate(num = str_length(area))
```

	sem	area	enroll	instructor	num
1	SP2023	History 30 - people	Ernesto Capello	7	
2	FA2023	Math 20 - people	Lori Ziegelmeier	4	
3	SP2024	Anthro 25 - people	Arjun Guneratne	6	

```
# Change the areas to "history", "math", "anthro"
classes |>
  mutate(area = str_to_lower(area))
```

	sem	area	enroll	instructor
1	SP2023	history 30 - people	Ernesto Capello	
2	FA2023	math 20 - people	Lori Ziegelmeier	
3	SP2024	anthro 25 - people	Arjun Guneratne	

```
# Create a variable that id's which courses were taught in spring
classes |>
  mutate(spring = str_detect(sem, "SP"))
```

	sem	area	enroll	instructor	spring
1	SP2023	History 30 - people	Ernesto Capello	TRUE	
2	FA2023	Math 20 - people	Lori Ziegelmeier	FALSE	
3	SP2024	Anthro 25 - people	Arjun Guneratne	TRUE	

```
# Change the semester labels to "fall2023", "spring2024", "spring2023"
classes |>
  mutate(sem = str_replace(sem, "SP", "spring")) |>
  mutate(sem = str_replace(sem, "FA", "fall"))
```

	sem	area	enroll	instructor
1	spring2023	History 30 - people	Ernesto Capello	
2	fall2023	Math 20 - people	Lori Ziegelmeier	
3	spring2024	Anthro 25 - people	Arjun Guneratne	

```
# In the enroll variable, change all e's to 3's (just because?)
classes |>
  mutate(enroll = str_replace_all(enroll, "e", "3"))
```

	sem	area	enroll	instructor
1	SP2023	History 30 - p3opl3	Ernesto Capello	
2	FA2023	Math 20 - p3opl3	Lori Ziegelmeier	
3	SP2024	Anthro 25 - p3opl3	Arjun Guneratne	

```
# Use sem to create 2 new variables, one with only the semester (SP/FA) and 1 with the year
classes |>
  mutate(semester = str_sub(sem, 1, 2),
         year = str_sub(sem, 3, 6))
```

	sem	area	enroll	instructor	semester	year
1	SP2023	History 30 - people	Ernesto Capello	SP	2023	
2	FA2023	Math 20 - people	Lori Ziegelmeier	FA	2023	
3	SP2024	Anthro 25 - people	Arjun Guneratne	SP	2024	