Pull the Right Strings with stringr: Exercises

By providing a set of wrappers to existing functions, the stringr package allows for simple, consistent and efficient manipulations of strings in R. Even though there are some more basic packages that offer strings-related functions, you might find yourself in need for a more complete and straightforward solution for handling strings in R.

With a simple and consistent syntax, stringr provides some very convenient functions around pattern matching, characters manipulation, whitespace handling and more. The full reference of the package can be found here.

Please find below a set of exercises that will help you practice a variety of stringr functions. The focus is on practical operations that data analysts are required to perform on a daily basis. Answers to the exercises are available here. And, don’t forget to check out our other exercise sets on the stringr package by following the stringr tag.

For the following exercises we will use this data:

```
addresses <- c("14 Pine Street, Los Angeles", "152 Redwood Street, Seattle", "8 Washington Boulevard, New York")
```

```
products <- c(“TV “, ” laptop”, “portable charger”, “Wireless Keybord”, ” HeadPhones “)
```

```
long_sentences <- stringr::sentences[1:10]
```

```
field_names <- c(“order_number”, “order_date”, “customer_email”, “product_title”, “amount”)
```

```
employee_skills <- c(“John Bale (Beginner)”, “Rita Murphy
```

(Pro)”, “Chris White (Pro)”, “Sarah Reid (Medium)”)

Exercise 1
Normalize the addresses vector by replacing capitalized letters with lower-case ones.

Exercise 2
Pull only the numeric part of the addresses vector.

Exercise 3
Split the addresses vector into two parts: address and city. The result should be a matrix.

Exercise 4
Now try to split the addresses vector into three parts: house number, street and city. The result should be a matrix.
Hint: use a regex lookbehind assertion

Exercise 5
In the long_sentences vector, for sentences that start with the letter “T” or end with the letter “s”, show the first or last word respectively. If the sentence both starts with a “T” and ends with an “s”, show both the first and the last words. Remember that the actual last character of a sentence is usually a period.

Learn more about string manipulation with stringr in the online course Learn R by Intensive Practice.

Exercise 6
Show only the first 20 characters of all sentences in the long_sentences vector. To indicate that you removed some characters, use two consecutive periods at the end of each sentence.

Exercise 7
Normalize the products vector by removing all unnecessary whitespaces (both from the start, the end and the middle), and by capitalizing all letters.

Exercise 8

Prepare the field_names for display, by replacing all of the underscore symbols with spaces, and by converting it to the title-case.

Exercise 9

Align all of the field_names to be with equal length, by adding whitespaces to the beginning of the relevant strings.

Exercise 10

In the employee_skills vector, look for employees that are defined as “Pro” or “Medium”. Your output should be a matrix that have the employee name in the first column, and the skill level (without parenthesis) in the second column. Employees that are not qualified should get missing values in both columns.