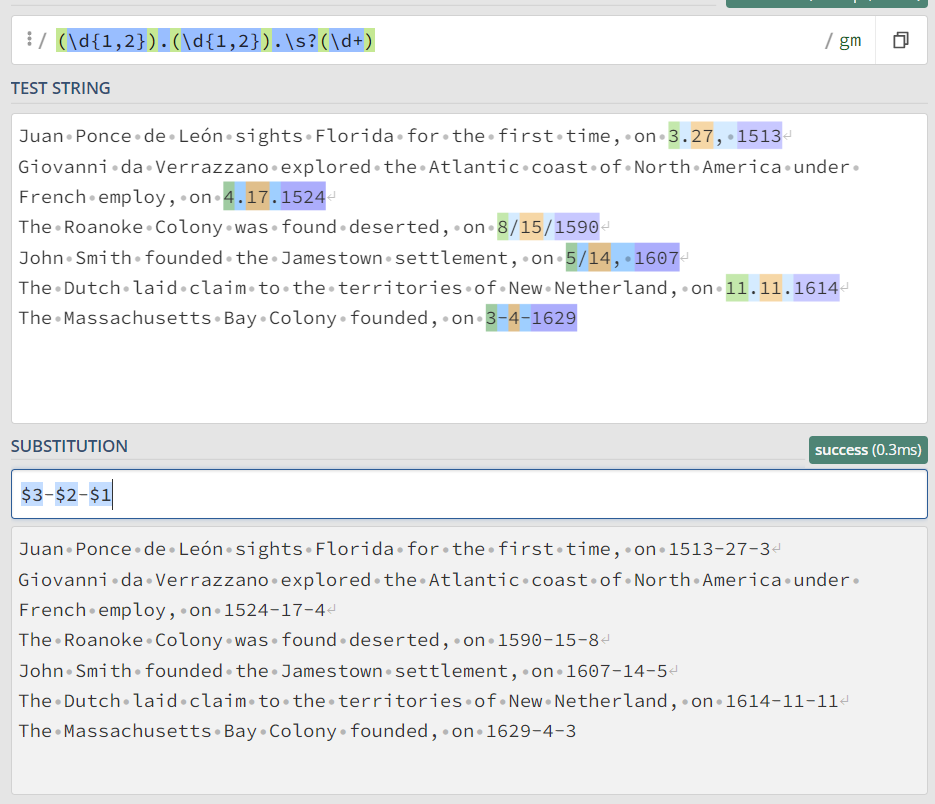
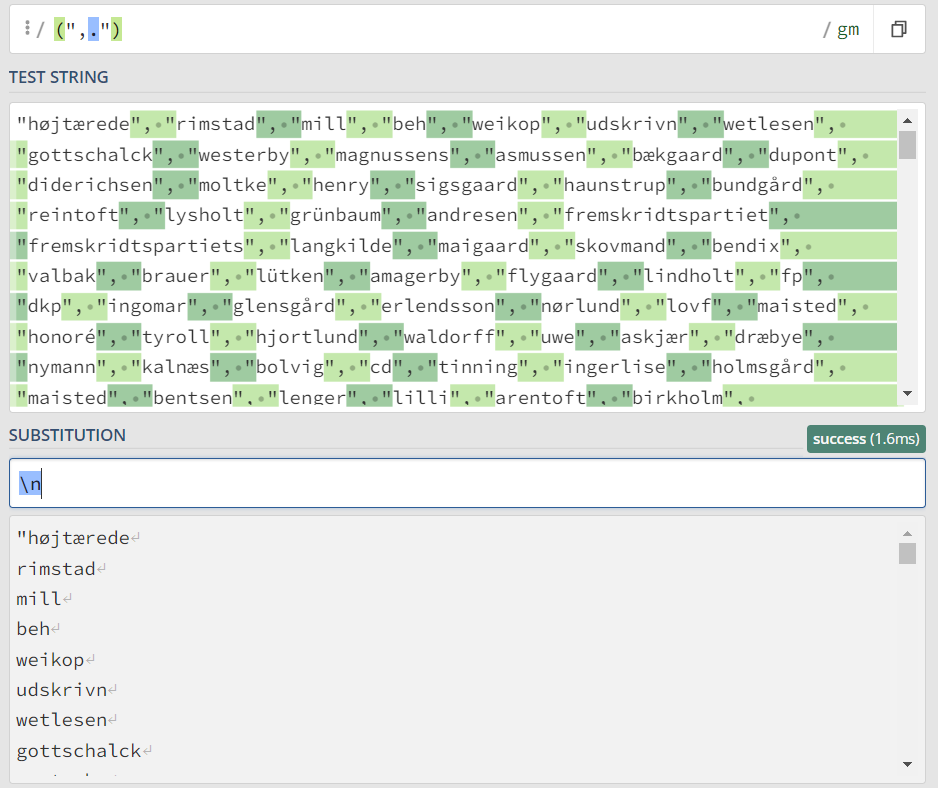
### **1: What regular expressions do you use to extract all the dates in this blurb:** [**http://bit.ly/regexexercise2**](http://bit.ly/regexexercise2) **and to put them into the following format YYYY-MM-DD ?**



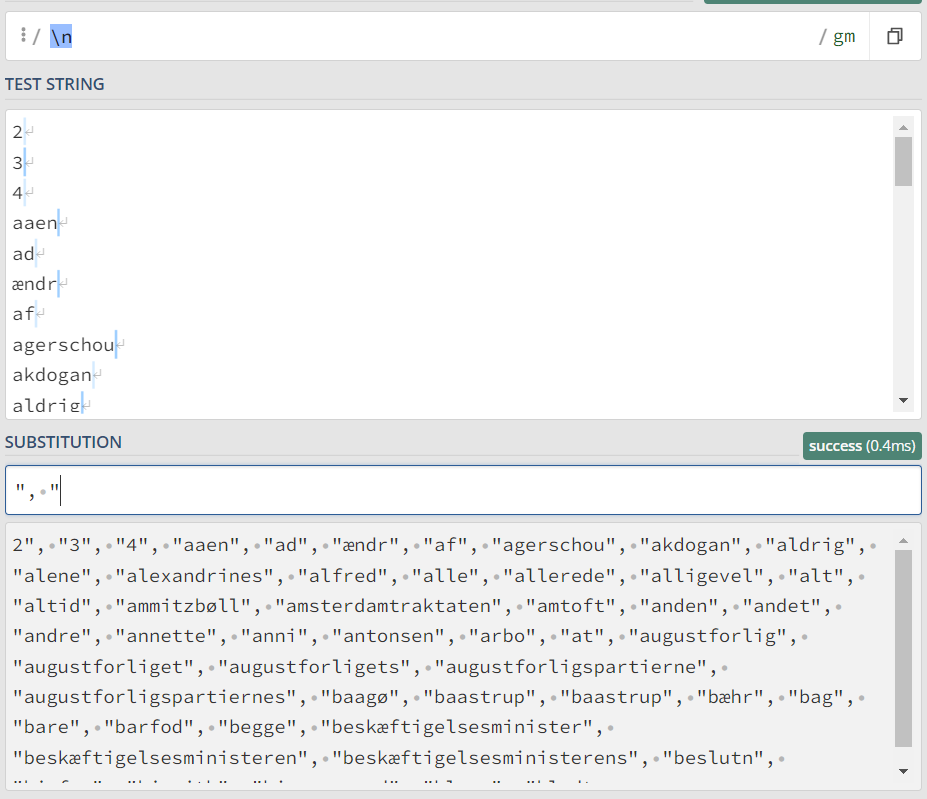
### 

### **2: Write a regular expression to convert the stopwordlist (list of most frequent Danish words) from Voyant in** [**http://bit.ly/regexexercise3**](http://bit.ly/regexexercise3) **into a neat stopword list for R (which comprises "words" separated by commas, such as** [**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4)**). Then take the stopwordlist from R** [**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4) **and convert it into a Voyant list (words on separate line without interpunction)**

The first quotation would have to be removed manually



The first quotation would have to be added manually



### **3: In 250 words, answer the following question: "What are the basic principles for using spreadsheets for good data organisation?"**

Data organisation is the foundation of a research project. It is important to keep track of the analysis and the original data, and that's why you should always create a new file or tab with clean or analysed data, while preserving the original dataset.

A researcher should keep track of the steps in the clean up or analysis. That is why it is important to take notes either in another text file or a new tab in the spreadsheet solely for note taking.[[1]](#footnote-1)

If a researcher plans on doing calculations or making data visualisation within spreadsheets, he/she should do so in a separate spreadsheet/file in order to keep their primary data tidy[[2]](#footnote-2).

The basic principles for spreadsheets are:

* Don't modify your raw data. Make copies before making any changes
* Keep track of the steps during the clean up or analysis.
* being consistent
* write dates like YYYY-MM-DD
* do not leave any cells empty
* put just one thing in a cell
* organise the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row)
* create a data dictionary
* do not include calculations in the raw data files
* do not use font colour or highlighting as data
* choose good names for things
* make backups
* use data validation to avoid data entry errors
* save the data in plain text files.

1. Formatting Data Tables in Spreadsheets: <https://datacarpentry.org/spreadsheets-socialsci/01-format-data/> [↑](#footnote-ref-1)
2. Woo, K., & Broman, K., 2017: Data Organization in Spreadsheets [↑](#footnote-ref-2)