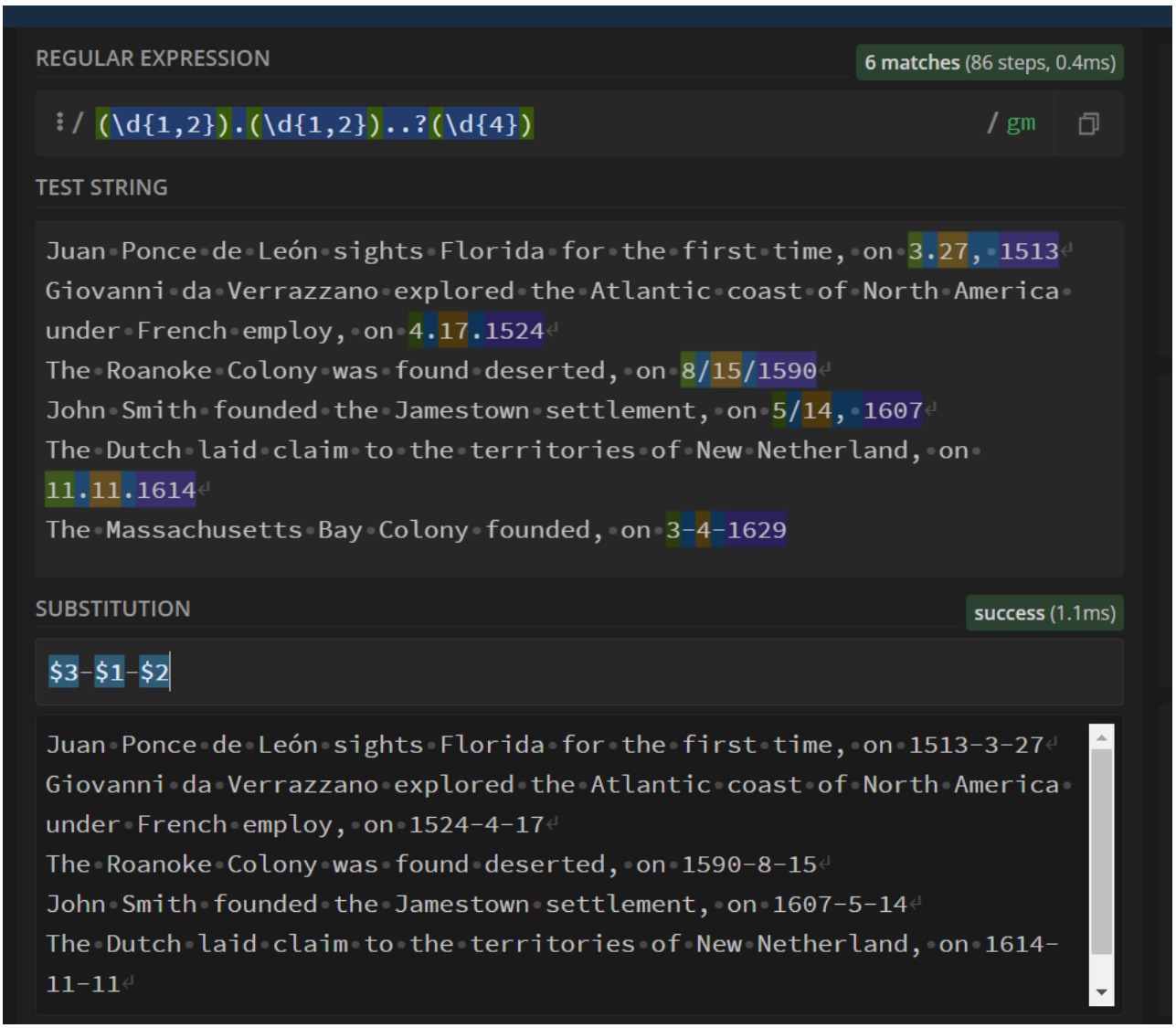
Rikard E. F. Jensen
Weekly homework
Peergrade

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

1: First I would use the regular expression \d{1,2} \d(1,2} +?\d{4}

2: After I would use the substitution toll by isolating the months, days, and years in ().

And then putting the year first with the $-function like I did in the picture:



Rikard E. F. Jensen
Weekly homework
Peergrade

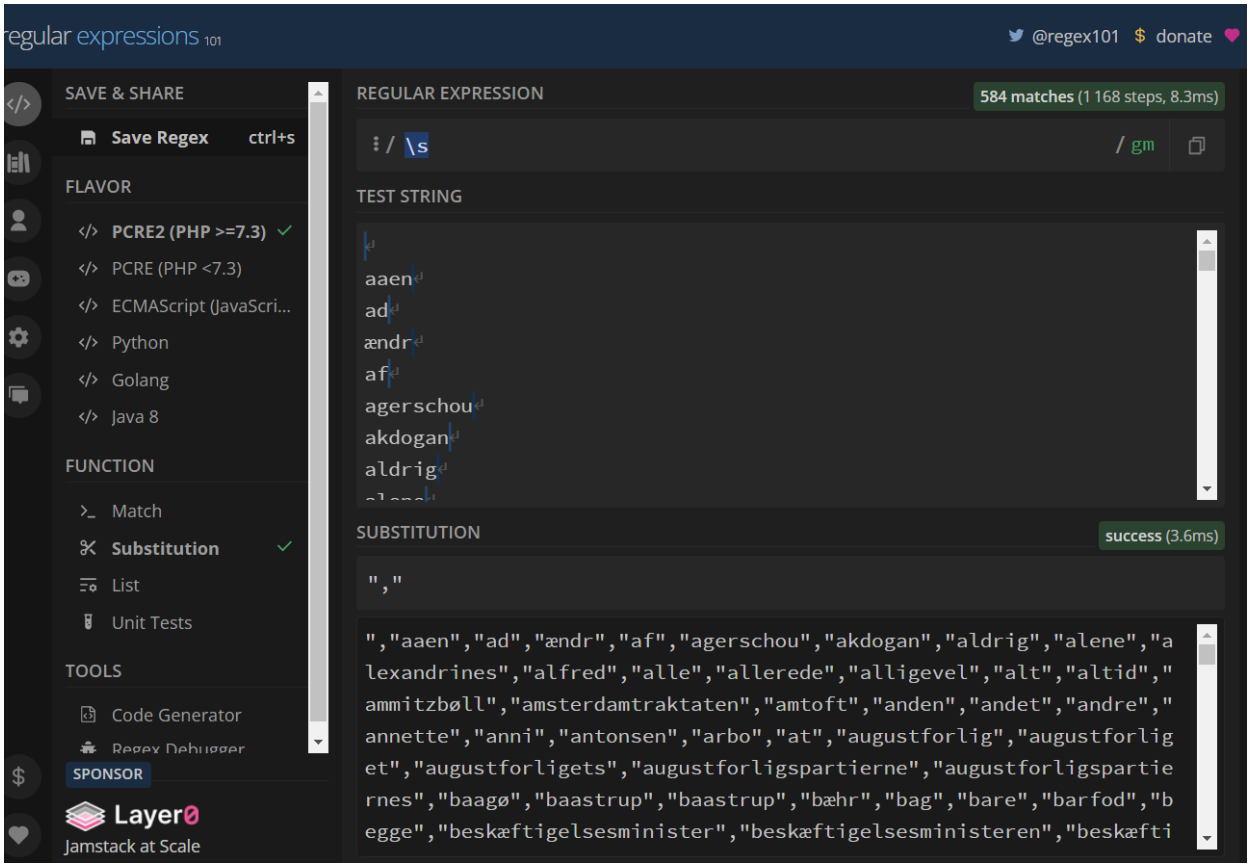
2. **Write a regular expression to convert the stopwordlist (list of most frequent Danish words) from Voyant in http://bit.ly/regexexercise3 into a neat stopword list for R (which comprises "words" separated by commas, such as http://bit.ly/regexexercise4). Then take the stopwordlist from R http://bit.ly/regexexercise4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

2.1 From Voyant to R

1: Used the regular expression \s

2: Used the substitution function and put a "," in it

Like in the photo down below:



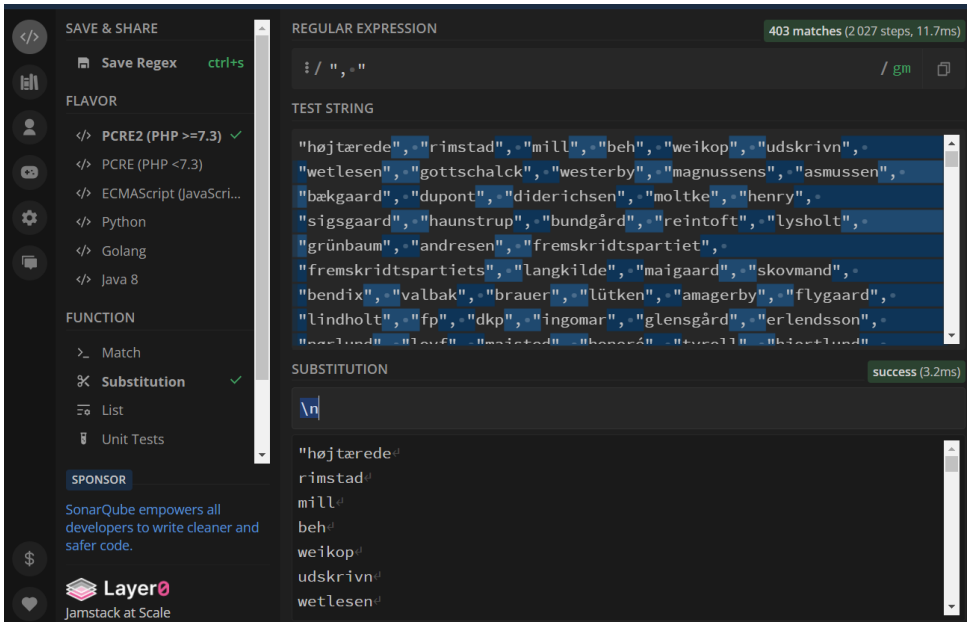
Rikard E. F. Jensen
Weekly homework
Peergrade

2.2: From R to voyant

First I wrote the regular expression “, “ to isolate the “,” that I didn’t want

Then I use the command \n to make the new line

Like in the photo below:



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

In the text “Data Organization in Spreadsheets” by Karl W. Broman & Kara H. Woo, can we see some good rules to remember when using spreadsheet. These rules help us stay organized and active when working in spreadsheets. I will here list the rules here:

1: Be Consistent - Entering and organizing your data in a consistent way from the start will prevent you and your collaborators from having to spend time harmonizing the data later.

2: Choose Good Names for Things - do not use spaces, either in variable names or file names and avoid special characters, keep it short, but meaningful

3: Put Just One Thing in a Cell - The cells in your spreadsheet should each contain one piece of data. Do not put more than one thing in a cell.

Rikard E. F. Jensen
Weekly homework
Peergrade

4: Make it a Rectangle - The best layout for data within a spreadsheet is as a single big rectangle with rows corresponding to subjects and columns corresponding to variables. The first row should contain variable names, and do not use more than one row for the variable names

5: Create a Data Dictionary - It is helpful to have a separate file that explains what all the variables are. It is helpful if this is laid out in rectangular form, so that the data analyst can make use of it in analyses.

6: No Calculations in the Raw Data Files - primary data file should contain just the data and nothing else: no calculations, no graphs.

7: make backups and save the data in plain text files