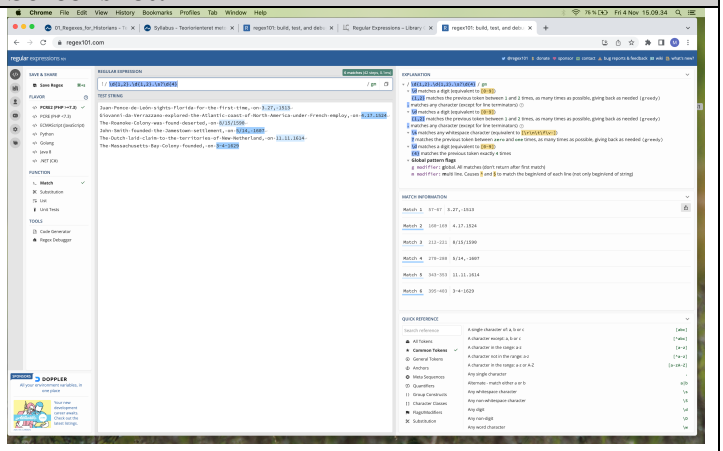
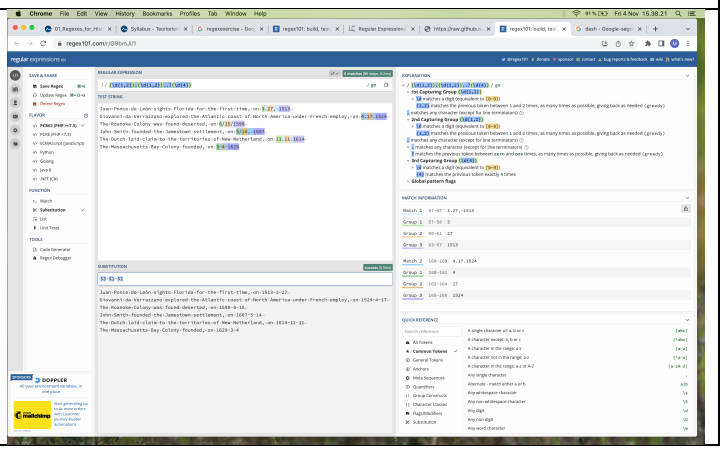
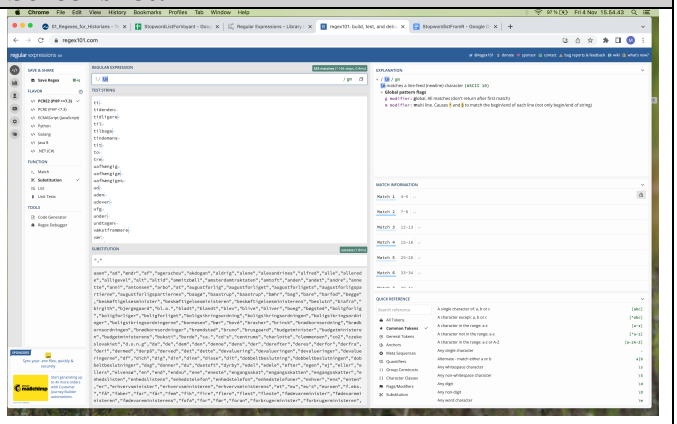
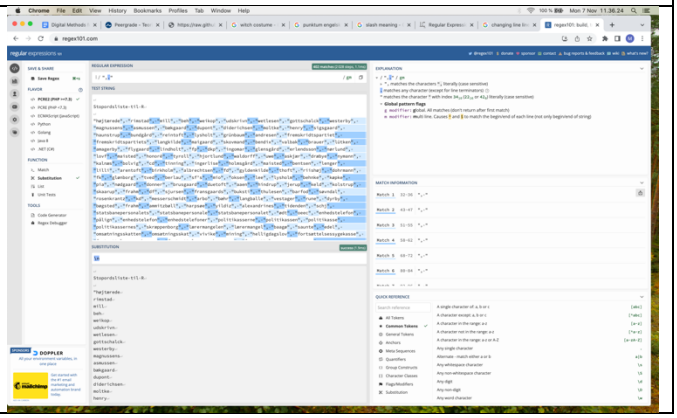
1. **What regular expressions do you use to extract all the dates in this blurb: http://bit.ly/regexexercise2 and to put them into the following format YYYY-MM-DD ?**

Description:	Screenshot:
I used the the following Regex: \d{1,2}.\d{1,2}.\s?\d{4} This searches for the following string: Any digit replicated one or two times which represents the day followed by any character (such as 'dash', 'decimal point' or 'slash') that separates the day from the month. The month too, is represented in the regex as any digit replicated one or two times. Then again it is separated by any character and followed by any digit replicated four times, representing the year. Access via: https://regex101.com/r/G9bmJi/1	
By inserting parenthesis in the regex, I defined the three capturing groups (day, month, year). Thereafter I substituted or rearranged the capturing groups according to the iso 8601 format (yyyy-mm-dd) by typing the code $3-$1-$2. Access via: https://regex101.com/r/UvTJTe/1	

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)**

Description:	Screenshot:
Voyant-style → R-style By using the regex \n, I marked the change to a new line. By substituting this with the character combination "," I changed the Voyant-style Stopword list to an R-style Stopword list. Access via: https://regex101.com/r/QOKSPb/1	
R-style → Voyant-style I used the regex ",," to mark the space between the actual stopwords. Thereafter I substituted this combination with the code \n that means new line. Access via: https://regex101.com/r/hxFyv7/1	

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

The basic principles are twelve practical recommendations for organising spreadsheets from the article *Data Organization in Spreadsheets*. These are important to keep in mind to reduce errors and to ease data analysis. The recommendations are as follows: 1) Be consistent in choosing names for e.g. file names and categorical variables.[1] 2) Choose names that are short but accurate and avoid spaces, use underscore or hyphens instead.[2] 3) Use the iso 8601 date

[1] Karl W. Broman og Kara H. Woo, "The American Statistician", *Data Organization in Spreadsheets* 72, nr. NO. 1, 2-10 (24. april 2018): 2–3, https://doi.org/10.1080/00031305.2017.1375989.
[2] Broman og Woo, 3.

system (yyyy-mm-dd).[3] 4) Leave no cells empty in the spreadsheet, instead fill in e.g. "NA".[4] 5) Write just one thing in each cell and leave units in the column name.[5] 6) The best layout in the spreadsheet is just one rectangle.[6] 7) Use a data dictionary, where you explain variables in a separate file.[7] 8) Do calculations in a separate file or a copy of the file containing the raw data, to avoid the risk of typing errors in the raw data.[8] 9) Despite it being a nice visualiser, do not use highlining or font color as a part of the data, as this I easily lost when converting to other types of data formats.[9] 10) Make backups of your data on a regular basis and use write-protection when not actively using the data file.[10] 11) use data validation to avoid errors – you can for instance use the tool "data validation" in Excel.[11] 12) Save a copy of your data in a plain text format e.g. the CSV format, as this can be accessed without using any special kind of software.[12]

[3] Broman og Woo, 3–4.
[4] Broman og Woo, 4.
[5] Broman og Woo, 4.
[6] Broman og Woo, 4–6.
[7] Broman og Woo, 6–7.
[8] Broman og Woo, 7.
[9] Broman og Woo, 7–8.
[10] Broman og Woo, 8–9.
[11] Broman og Woo, 9.
[12] Broman og Woo, 9.