# 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 ?

Step 1:   
\d{1,2}.?\d{1,2}..?\d{1,4}

\d matches any single digit

{1,2} makes it so that it will match with a number with either 2 or 1 digits

. is used to make sure ., / and - will fit. The dot means any character

? i used to make sure that the cases with a extra space between the preceding character will be matches as well

{1,4} makes it so that it will fit with a number with either 1,2,3 or 4 digits to get all the years

Step 2:   
(\d{1,2}).(\d{1,2})(..?\d{1,4})

by containing every part of the code in () it can be change around i substitution

Step 3:   
Indsæt i substitution: $3-$1-$2

this code change the former contained part so that it will be YYYY-MM-DD

# 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)

Exercise 3:

([A-zøæåüé0-9\.]+)\n

Explanation: ([A-zøæåüé0-9\.]+)\n

/

gm

**1st Capturing Group** ([A-zøæåüé0-9\.]+)

**Match a single character present in the list below** [A-zøæåüé0-9\.]

+ matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)

A-z matches a single character in the range between A (index 65) and z (index 122) (case sensitive)

øæåüé matches a single character in the list øæåüé (case sensitive)

0-9 matches a single character in the range between 0 (index 48) and 9 (index 57) (case sensitive)

\. matches the character . with index 4610 (2E16 or 568) literally (case sensitive)

\n matches a line-feed (newline) character (ASCII 10)

<https://regex101.com/r/GooeJP/1>

exercise 4

<https://docs.google.com/document/d/1XB4S7GGR3UekW3pc1hma6tupxBKGI6B7FbkhuU4J3Q4/edit?usp=sharing>  
  
<https://regex101.com/r/W664rR/1> regex til ovenstående liste

Explanation:

/

"|(["])|([,])|([A-Za-zøæåüé0234]+)

/

gm

**1st Alternative** "

" matches the character " with index 3410 (2216 or 428) literally (case sensitive)

**2nd Alternative** (["])

**1st Capturing Group** (["])

**Match a single character present in the list below** ["]

" matches the character " with index 3410 (2216 or 428) literally (case sensitive)

**3rd Alternative** ([,])

**2nd Capturing Group** ([,])

**Match a single character present in the list below** [,]

, matches the character , with index 4410 (2C16 or 548) literally (case sensitive)

**4th Alternative** ([A-Za-zøæåüé0234]+)

**3rd Capturing Group** ([A-Za-zøæåüé0234]+)

**Match a single character present in the list below** [A-Za-zøæåüé0234]

+ matches the previous token between one and unlimited times, as many times as possible, giving back as needed (greedy)

A-Z matches a single character in the range between A (index 65) and Z (index 90) (case sensitive)

a-z matches a single character in the range between a (index 97) and z (index 122) (case sensitive)

øæåüé0234 matches a single character in the list øæåüé0234 (case sensitive)

step 1

"|(["])|([,])|([A-Za-zøæåüé0234]+)

step 2

\n$3

step 3

copy it all into a stopword list

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

It is important to name variables and files properly, this has a dual purpose, 1. avoiding special characters except for underscores and hyphens, as these often have another meaning in programming language 2. avoid using spaces to lessen workload. You should focus on short meaningful names consisting of only letters and numbers.

When working with dates and numbers not all spreadsheets work the same, so it’s good to understand how your specific spreadsheet handles numbers and dates. However, it’s usually good to work with YYYY-MM-DD for dates.

Only put one piece of data in each cell and don't merge cells.

Organize your data into rectangles, with rows corresponding to subject and columns to variables, in case the dataset cannot fit into 1 rectangle, make a set of rectangles

When you make a spreadsheet, it is good to have a “dictionary document” as this allows you to know what everything means. This usually goes hand in hand with metadata and something like a readme.txt if one such thing is needed.

When working with data management in a spreadsheet, thy musnt conduct in calculations, as it is by the academic community of pentagon hackermen (and women) considered to be a severe case of tomfoolery and will result in the penalty of 2 days of tickle torture.

# Challenge (OPTIONAL)!Can you find all the instances of 'Dis Manibus' invocation in the EDH inscriptions in <https://bit.ly/regexexercise5>? Beware of the six possible canonical versions of the Dis Manibus formula!