**Homework by Christian Kjeldsen:**

**Part 1.**

Date formatting

<https://regex101.com/r/XN2I7H/1>

**explanation:**

I achieved this by using the following command (\d{1,2}).(\d{1,2})..?(\d{4}) which then showed the 6 times a date appeared in the text. The command uses the data numeration which is

YYYY-MM-DD but here the data is the opposing way in the text mm-dd-yyyy. That we needed to transform to the YYYY-MM-DD form.

To start with we need to sample alle the dates in the original text and here is how:

Because there is only 30 or 31 days in a month the first two digits can either be 1,2 and the same goes for the month which is also only available in two digits. That is why there is the same command there and then you have the last \d where there is 4 digits because of the numeration of the years have got 4 digits. And here the \d stands for any digit and then the 1,2 stands for the 1 og number 2 digit in that section. The last bit is the ? This reads both if the number includes a dot like expected and if it does not and because of that it gets all the data and not just the ones ending with a dot after the to first \d’s

And then you end that limitation and separate that again via a dot . that separate the two paragraphs in the formal. And then you repeat the same column for the month and separate that and then for the last part you again start off with a \d for any digit and now you have specified it for 4 digits in the paragraph because this is what registers the years. Now we have the new YYYY-MM-DD system and only need to transform the text to it the way we do this is to give the section for years the ( \d{4}) the mark $3 and the two ( \d{1,2}) the maks $2 $1

Then in the substitution you write $3$2$1 so it becomes YYYY-MM-DD in the substation text.

**Part 2.**

**Stop Word list from Voyant to R**

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

**Explanation**

In Voyant you have a certain list of word that are commonly used this

You type in to a stop word list for Vouyant to know, now this appears as one single column and that’s not working In R so we need to transform it in to R format and we do this by using the following command (\w+)(\b).

The first \w is for looking for any letter from a-z and then the + acts as a quantifier two the previous command to it repeats the letter command \w regardless of the length of any given word. Then you just write $1 and $2 in the substitution.

Stop Word from R to Voyant

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

In this list we reverse the transition from R to Voyant format. The way we do this is by marking the horizontal listen in the R form and make it vertical like the Voyant form. here we use the following formular “,\* ” (here the Asterix should be the dot representing the a space between two words) and then in the other text span we type in \n so the listing is vertical like in Voyant.

**part. 3.**

**what is good data management In a spreadsheet?**

A good spreadsheet is firstly a well-organized one, it is clear and consequently managed in order to keep the work in it as efficient as can be. Now how do one achieve this? According to the text red for Fridays lecture by Broman K.M and Kara H. Woo[[1]](#footnote-1)

It is largely to do with assigning the rows specifically to one type of data. Example a date a year a month. Then it is about making sure to keep track of your work progress and to mark any outstanding data not by a digit but by a word that clearly standout.[[2]](#footnote-2) And make sure to use the same date system for the project preferably YYYY-MM-DD according to Woo and Broman. [[3]](#footnote-3) Another thing they mention is to keep different parts of a project on different pages so two different parts doesn’t collide in the spreadsheet. [[4]](#footnote-4) Then do not fit more than one data point into a cell.[[5]](#footnote-5) Then they mention another thing that is to keep a dictionary of the terms that you use in the spreadsheet. This so that you are able to keep track of your shorted terms that you might have been using in your spreadsheet. [[6]](#footnote-6)The next important thing to remember is the layout of the text which needs to be formed into a rectangular shape and be so consisted because if this Is it makes it much easier to organize different sets of data and graphs.[[7]](#footnote-7)

1. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-1)
2. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-2)
3. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-3)
4. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-4)
5. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-5)
6. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-6)
7. Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1,2-10, DOI: [10.1080/00031305.2017.1375989](https://doi.org/10.1080/00031305.2017.1375989) [↑](#footnote-ref-7)