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 ?

Et billede, der indeholder tekst, elektronik, skærmbillede, software

Automatisk genereret beskrivelse

<https://regex101.com/r/hHFXp1/2>

1. 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)

Et billede, der indeholder tekst, skærmbillede, software, nummer/tal

Automatisk genereret beskrivelse

<https://regex101.com/r/hHFXp1/3>

Et billede, der indeholder tekst, skærmbillede, software, Webside

Automatisk genereret beskrivelse

<https://regex101.com/r/hHFXp1/4>

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

For good spreadsheet use to organize data, one should be consistent and not use the spreadsheet as you would normal lab notes. Therefore it is important to set up well formatted tables beforehand.

*Layout:* The best layout for data is a big rectangle with rows corresponding to subjects and columns responding to variables. The first row should be for variable names, but only one row, should be used for that.

*Consistency*: it is important to stay consistence while working with data. If you are working with a partner you won’t have big problems harmonizing the data later.

*Naming*: When naming variables in coding stay consistence, otherwise, you might miss date while analyzing it. When choosing a data language, stay consistence with that as well, so that when others look at it, they have an easier time reading it.

*Cells:* Always fill in all cells, if we are missing data, rather fill it “NA” instead of leaving it blank

*Data in the cells:* Only let a cell contain one piece of data. Rather make new cells with new information. Never merge cells as that would violate the “more data in one cell” rule, or leave cells without data.

*Dates:* Be aware of dates in excel, while it likes to treat texts like dates as well. Therefore write dates in the YYYY-MM-DD.

*Data dictionary:* Create a data dictionary, that explains your variable. In that way, if others or yourself are in doubt about the datas content. This could also help you make your variables names shorter.

1. 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!.

Et billede, der indeholder tekst, software, Webside, Website

Automatisk genereret beskrivelse

The file would not save, probably because of saving limit.