# Homework uge 1

## Opgave 1:

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

Jeg skriver følgende regular expressions for at få alle ønskede numre markeret, samt diverse mellemrum:

\d+.\d+.\s?\d{4}

<https://regex101.com/r/IaTK79/1>

Herefter tilføjer jeg parenteser for at indføre forskellige kategorier efter måned, dag og år:

(\d+).(\d+).\s?(\d{4})

<https://regex101.com/r/IaTK79/2>

Dette substituerer jeg så rækkefølgen passer med yyyy-mm-dd:

$3-$1-$2

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

## Opgave 2:

**Write a regular expression to convert the stopwordlist (list of most frequent Danish words) from Voyant in**[**http://bit.ly/regexexercise3**](http://bit.ly/regexexercise3)**into a neat stopword list for R (which comprises "words" separated by commas, such as**[**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4)**).**

Jeg skriver følgende regular expressions for at få alle mellemrum markeret:

\n

Dette substituerer jeg til understående, så ordene nu står enkelt i min stopword liste:

", "

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

**Then take the stopwordlist from R**[**http://bit.ly/regexexercise4**](http://bit.ly/regexexercise4)**and convert it into a Voyant list (words on separate line without interpunction)**

Jeg skriver følgende regular expressions til at markere skiftet mellem ordene:

", "

Dette substituerer jeg til understående, så ordene nu står enkelt i min voyant liste:

\n

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

## Opgave 3:

**”What are the basic principles for using spreadsheets for good data organisation?"**

Broman og Woo angiver en liste af overordnet råd til at lave gode spreadshets for dataorganisering. Jeg vil opliste følgende og supplere kort med egne kommentarer.

1. Være konsekvent i sin fremgangsmåde. Det er meget lettere at arbejde med et harmonisk og ens datasæt. Kod derfor dine variable, navne og ”manglende værdier” mm. konsekvent på samme facon.
2. Vælge gode praktiske navne for dine kategorier. Anvend derfor eks. ikke mellemrum og lav ikke navnene længere end højest nødvendigt.
3. Skriv dato på følgende måde: YYYY-MM-DD.
4. Efterlad Ingen tomme celler – dvs. anvend en bestemt kode til at angive manglende værdier for en given variabel.
5. Kun en værdi/information pr. celle. Der er ingen grund til at blande information sammen.

1. Lav rektangulære spreadsheets. Dvs. rækker skal indeholde emner og kolonner bør indeholde variable.

1. Lav en kodeordbog, så du ikke mister overblik over hvad din information betyder, er målet osv.
2. Undlad at lave udregninger i cellerne og dit spreadsheet – de skal kun indeholde ren data.
3. Marker ikke din data med farver eller fed skrift, find i stedet på et alternativ.
4. Anvend datavalidering til at undgå fejl.
5. Gem din data så du altid har en backup og omform også din data til simple tekst-filer (CSV format). De er ofte lette at anvende til senere at kode.

Alle overstående punkter er med til at sikre at ens data er organiseret godt og derfor i større grad anvendeligt for dig eller andre.