# **Digital Methods**

Week 36: Homework

**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?**

Et billede, der indeholder tekst, Font/skrifttype, software, nummer/tal

Automatisk genereret beskrivelse

Link: <https://regex101.com/r/IMVnHh/1>

**2.** **Write a regular expression to convert the stopword list (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)**). 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)**

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

Automatisk genereret beskrivelse**

Voyant coverted into R: <https://regex101.com/r/PKqoAE/1>

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

Automatisk genereret beskrivelse

R converted into Voyant: <https://regex101.com/r/XKZek0/1>

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

Using spreadsheets for good data organization is essential for effective data organization and collaboration.

A basic principle to follow is to be consistent. Being consistent with your data simplifies further work. It's important to use the same units of measurement or naming conventions throughout the spreadsheet. Another basic principle to follow is choosing good names for things. It's important to think carefully when choosing names. Special characters or spaces in a name can complicate the process. Again, it's important to be consistent. It's also important to write dates as YYYYY-MM-DD, as this is a global standard.

A third basic principle is to remember not to leave any cells in a spreadsheet blank. It's better to write 'NA' instead of leaving it untouched. This shows that it is intentional that the cell should not contain any data. Also, each cell should only contain one type of data and you should not merge cells.

A principle for the layout of a spreadsheet is to put your data in a rectangular shape. The first column should always be the names of the variables. Another principle is to create a data dictionary, as it can clarify and explain what the variables are.

A final principle to remember in good data organization is to always back up your data and use data validation to avoid errors – this way you can avoid incorrect data entry. It's always a good idea to have backups to avoid loss.

Here are several key principles for effective data organization, including some that I consider to be particularly vital.

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

**Et billede, der indeholder tekst, skærmbillede, Webside, Website

Automatisk genereret beskrivelse**

Link: <https://regex101.com/r/8PoDeH/1>

(The file could not save with all the text string included – see screenshot)