**Homework Week 45**

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 ?

I have used the regular expressions : [\d]{1,2}.[\d]{1,2}.\s[\d]{1,4}|[\d]{1,2}.[\d]{1,2}.[\d]{1,4}

I have used the ”.\s” to find the whitespace and I only use one, because i know where it is.

I have also used the ”|” because not all of the dates have a whitespace – In the class we did in a easier way, but i dont remember how.

I have found the right regular expression: (\d{1,2}).(\d{1,2})..?(\d{4})

I use ”..?” instead of ”\s”

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)

I have used the regular expressions: ("..?\s")|(")

I highlight everything there is not a word and I use a pipe to catch the first ” and the last ”

Then I make a new line i the substitution without using any of the groups that a i highligted

I have used the regular expressions: ^(.+)$\n\*

The ^ express the start of a line and the (.+) express the any character with one or unlimited times and the $ express the end of a line and \n express a new line and the \* express zero or unlimited

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

For using spreadsheets for good data organisation the basic principles that you must be consistent. I means that you have to do the same thing every time. The next principles is write dates like YYYY-MM-DD, is it so much easier when there is only one way to write dates. You can not leave any cells empty beacuse it can be a problem to remeber if it is on purpose or if it is just a mistake. After that you must always put just one thing in a cell, it will save you for mixing up different things and it is a better way to keep it organised. As i write about organize, you should organize the data as a single rectangle (with subjects as rows and variale as columns, and with a single header row). Again it make easier to get the system, not only for you but also for everyone else who will see the data. To help other people to understand your data, create a data dictionary. Do not include calculations in the raw data files, it will confuse you, so make a file for that. Do not use front color or highlighting as data it can give problem with the look. Choose good names for things so you can remeber what it is about. Make backups, it is really bad if loose your data. Use data validation to avoid data entry errors. Save the data in plain text files.

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!