Opgave 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 ?

regular expresseion: (\d+).(\d+).\s?(\d{4})

substitution: $3-$1-$2

opgave 2: 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)

regular expression: ", "

substitution: “\n

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

Den første regel er at være konsekvent i sin dataorganisering. Dette gør at man slipper for oprydning i koden senere. Når man så skal navngive noget i sin kode, er det en god ide at undgå mellemrum, kort og præcist er bedst i forhold til navne, og hvis man har brug for et mellemrum, kan man istedet benytte \_. Det er vigtigt at man udfylder alle celler med data, men man skal dog ikke have mere end en ting i hver celle af gangen. For det bedste layout i et excel ark er det god ide at opstille sin data i en rektangel. Den første række bør indeholde variabelnavne. For bedre at kunne finde rundt i sin egen data, kan man lave en separat fil til forklaring af variablerne. I en sådan fil kan man eksempelvis skrive ting som; det præcise data navn som det fremgår i data filen, forklaringer af varibalernes betydning, forventede minimum og maximum værdier, måleenheder, og versioner af variabelnavne som passer i datavisualiseringer. Pas på med at lave udregninger i dit primære dataset, gør man dette løber man en risiko for at skrive fejl, som måske først senere opdages ind i sit datasæt, sørg derfor for, at den primære datafil kun indeholder data, og kopier indholdet over i en anden fil hvis der skal analyseres og gøres ved.