Homework W36 Regular Expressions

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 ?
   * <https://regex101.com/r/ildmLl/1>
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)
   * <https://regex101.com/r/skVsaG/1> (from Voyant to R)
   * <https://regex101.com/r/6tI1SQ/1> (from R to Voyant)
3. In 250 words, answer the following question: "What are the basic principles for using spreadsheets for good data organisation?"

Spreadsheets are widely used for lots of different purposes within different kind of researching areas. It is ideal for data research, storage, organizing and analyzing because of its rectangular design which offers and easy way to make calculations and visualizations of data. When working in spreadsheets there are some basic principles which should be followed to make it easier to navigate in the analysis and avoid or reduce data errors. First it is important to be consistent in everything you do when working with data in spreadsheets especially when naming, categorizing, or identifying variables or data. Consistency helps you navigate and not make your work interfere between data and variables or across datasets. Consistency is important when writing dates as well. Dates must be written as YYYY-MM-DD because if technical variations and formats can vary when working in different spreadsheets. Furthermore, be aware to leave no empty cells but instead write NA when data is missing or unknown. Also only write one piece of data in each cell. To keep the data raw and analyzable there should be no calculations in the raw data set. Calculations and graphs can be made in and additional spreadsheet. And keep the dataset visual simple and clean in a rectangular format and without highlighting to make it easier for yourself, collaborators, and the computer to read and process the data. Finally, it may be helpful to create a Data Dictionary in a separate spreadsheet, which explains and clarifies the data and variables. A Data Dictionary can help prevent and reduce the misconception of the data and the variables.

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!.
   * <https://regex101.com/r/uqHjf9/1>

Et billede, der indeholder tekst, Webside, Website, nummer/tal

Automatisk genereret beskrivelse

I couldn’t save a link for the whole regular expression with all the text, but here is a screenshot with 23791 matches ☺