

1:W35: Regular expressions and spreadsheets

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?
 - Marking all numbers:
 - `(\d+)(\d+.*)(\d+)`
 - Substituting numbers
 - `$2$1$3`
2. Write a regular expression to convert the stopwordslist (list of most frequent Danish words) from Voyant in <http://bit.ly/regexexercise3> into a neat stopwords list for R (which comprises "words" separated by commas, such as <http://bit.ly/regexexercise4>). Then take the stopwordslist from R <http://bit.ly/regexexercise4> and convert it into a Voyant list (words on separate line without interpunction)
 - Removing all characters:
 - `(\\, "|\\")`
 - In substitution field add new line:
 - `\\n`
3. In 250 words, answer the following question: "What are the basic principles for using spreadsheets for good data organisation?"

Keep data in a tidy format with variables in columns and observations in rows. Each field should contain only one value. Keep strings and numeric variables separated. Be consistent in naming variables, avoid special characters and spaces except for underscore (_). Name variables in a logical and comprehensible way.