# Homework W1 Regex

Task 1:

<https://regex101.com/r/J3x5FC/1>

To extract the Dates from the text and put them in YYYY/MM/DD, I do the following:

* To find the days I use (\d+) which looks for all the days marked by having 1 or more digits in them, \d meaning any digit and the + being used for indicating 1 or more digits being in use.
* To find the months I do the exact same with (\d+) cause now it is any digit and considering it is months it will most often be 2 digits being used, henceforth the +
* To find the years I use a slightly different Regex. (\d{4}) this here means again any digits but with {4} symbolizing I am only looking for something with 4 digits in specific.
* In between the days and months I use “.” Which means “any character cause that covers both , . and / in between the different values
* In between months and years I use “. .?” to then again show that it is any character, but this time I have an extra . with a ? to show that there is potentially something else, for example a space in-between.
* All of the three different values are in a “( )” to specify them being separate entities, and then to rearrange them in the asked order I put them as ($3/$1/$2) in the substitution bar

Task 2:

Voyant to R: <https://regex101.com/r/PCq69m/1>

R to Voyant: <https://regex101.com/r/qdgk2U/1>

* To Rearrange to from Voyant to R where it is required to go from Separate lines to the format of “, “ I start by using the Regex (\n) which locates all the instances of where a new line appears and marks them
* Then I substitute all the marked new lines with the format of “, “ which gives the almost wanted result
* Only caveat here is that there is some manual labor to do with the first and last word, cause with the first word there isn’t a “ in front of it, and the last word has an extra ,” at the end. This needs to be removed manually when this Regex solution is used.
* To Rearrange from R to Voyant it needs the exact opposite. Voyant wished for every word to be separated by a new line instead of “, “. So here I start out by finding all instances of “, “ and marking them down as (“, “).
* Then I substitute this with the wished \n to replace all the instances of “, “ with a new line.
* This again leaves some problems at the beginning and the end that needs to resolved manually.

Task 3:

Spreadsheets are good for data overview in the sense that they create a good separation between the values in use and therefore gives a good overview of the data in use. Another good principle to keep in mind is to keep your data sheets clean and where it is easy to look at them and gain a good overview of the situation in use.

Spreadsheets can also be used for visualization and models/graphs but those should be kept separate to avoid a cluttering in the data. The Spreadsheets can easily be used for it and should be, but keep the models separate from the data itself.

The data itself should also be closely monitored as it is entered into the spreadsheet and avoid copy paste as much as possible as that is what can possibly introduce errors cause not all formats and ways of writing the data can be directly translated into spreadsheet programs.

Spreadsheets overall can be used to great results to collect your data and give a good overview of it to have it sorted as you wish so it is easier to process and work with. The spreadsheet functions as a sorting and working program for your data where it is also possible to incorporate other types of visualizations and models of the data.