# Homework W2 Open Refine

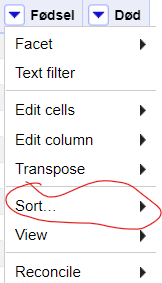
Task 1:

I started by manually taking the data from the given website and writing it into Excel (see attached document). Missing data was replaced by a 0 to indicate that this data was not available to us.

Et billede, der indeholder tekst, skærmbillede, nummer/tal

Automatisk genereret beskrivelseTo sort it by years of birth I started by putting my data into OpenRefine to give me an overlook of my data and to have it able to edit. Considering the fact that a lot of the earlier monarchs and the two interregnum has a value of zero these will appear at the top when sorting by size, but after that, the monarchs appear in the order they should.

Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, display/skærm/fremvisning

Automatisk genereret beskrivelseTo get to this, I used OpenRefined own function of sorting. This allows me to access the function of sorting by numbers and then sort by the smallest number first.

Task 2:

No, OpenRefine does not make changes to your raw data in any way. It takes your raw data and reads it how you told it to read it and then presents it to you in a format that is best/what you specified it to do so you can start doing work on the data. It stores your raw data as you gave it to it and doesn’t change anything in it, any changes being made only happens in the version you are working on in the program

Task 3:

To find what the question is asking for, the two months where the farmer reported the least amount of water, I start by inserting the data in OpenRefine and location the category which is being sought, in this case months\_no\_water and starting doing work on it.

I start by opening a text facet for the category and then I write the following command:

value.replace("[", "").replace("]", "").replace(" ", "").replace("'", "").split(";")

This whole command can be separated into two, with the first part being all the 4 different replace commands. This section is basically there to remove any and all other characters being mentioned in this category so we are only left with the months. This is done by replacing all the characters with nothing.

The second part is the split which basically takes all the ; and replaces them with a new line. The reason for choosing the ; is because it is the last character between the different months. Creating the new lines separates the months into different categories and therefore gives us the count for how many times the farmers have said each specific month.

So the final answer is that the two driest months are September and October.

Et billede, der indeholder tekst, skærmbillede, nummer/tal, software

Automatisk genereret beskrivelse