**W45: Open Refine**

***Question 1:***

**Create a tidy spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and duration (start and end of the reign) of reign. They should be sortable by year of birth.**

Spreadsheet is attached.

All the information about the monarchs and their reign is taken from danmarkshistorien.dk.[[1]](#footnote-1)

***Question 2:***

**Does OpenRefine alter the raw data during sorting and filtering?**

No, it does not alter the raw data. It uses a copy of the data, so it never modifies the raw data set.[[2]](#footnote-2) It also keeps a track of all your steps, that you can then undo if wanted.

***Question 3:***

**Fix the interviews dataset in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"**

* First, I find the column with information about water, and that is the one called months\_no\_water.
* I then press the down arrow at months\_no\_water and click “edit cells” and then “transform”. A new window pops up and in this I write the expression value.replace("[","").replace("]","").replace("'","").replace(" ","") to replace unwanted elements.
* I press the down arrow again and choose “facet” and then “custom text facet”. Once again, another window pops up. In this I use GREL and write the expression value.split(“;”) to remove the ; between the months.
* In the left, side all of the months, and how many times they are mentioned, then pop up. I choose to sort them by count so that the one used the most is at the top.
* I can then see that October is the driest month (mentioned 74 times) and September is the second driest month (mentioned 70 times). Below is a screenshot of my results:



1. <https://danmarkshistorien.dk/vis/materiale/kongeraekken/> [↑](#footnote-ref-1)
2. <https://datacarpentry.org/openrefine-socialsci/01-introduction/index.html> [↑](#footnote-ref-2)