

W45: Open Refine

DESCRIPTION

Upload your answers to these questions:

1. Create a **tidy** spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and duration of reign. They should be sortable by year of birth. Suitable source websites are [here](#) and [here](#), but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)

Spreadsheets are attached

2. Does OpenRefine alter the raw data during sorting and filtering?

No. And that is good, since we use OpenRefine to process data and not create it. Another good thing is that it saves your progresses, mistakes, and work in general. So, if a mistake happens you and your work-partners can go back and easily find and undo the mistake.

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?"

I first have to get all the unwanted signs and characters out of the line. I do that by use the replace function and replace it with nothing. Like I did in the picture below. After I split the words that are separated by the a ";". I do that with the split function in the end of the expression like below.

Edit Facet's Expression based on Column months_no_water

Expression Language General Refine Expression Language (GREL) ▼

`value.replace("[\", \"'\"]\", \"\").replace(\";\", \"\").replace(\"[\", \"\").replace(\"'\", \"\").replace(\"\\\", \"\\\").split(\";\")` No syntax error.

Preview History Starred Help

row	value	value.replace("[\", \"'\"]\", \"\").replace(...
1.	NULL	["NULL"]
2.	['Aug' ; 'Sept']	["Aug", "Sept"]
3.	NULL	["NULL"]
4.	NULL	["NULL"]
5.	NULL	["NULL"]

OK Cancel

After that I can read that **October** with its 74 households and **September** with its 70 households where the 2 months where that was a most households were without water. I could read that from the box below after I sorted by count instead of name.



The screenshot shows a Peergrade interface with a table titled "months_no_water". The table is sorted by count, with "Oct" having the highest count at 74, followed by "Sept" at 70. The table has 11 choices in total. The interface includes a "change" button and a "Sort by" dropdown menu currently set to "name".

Month	Count
Oct	74
Sept	70
Nov	51
NULL	45
Aug	33
Dec	11
Jan	2
July	2
Apr	1
June	1