

Homework W37

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

Knud den Store (995-1035) is considered here as the second Knud. The first Knud was Knud (1.) Hardegon (died before 935), also called Hardeknud Svendsen, father of Gorm den Gamle(?-958).

Gorm den Gamle's year of birth is estimated to year 908 and year of birth, 958
Harald Blåtand's year of death is estimated to year 987

We made the spreadsheet tidy by making sure that we used the different principles when working with data in spreadsheets. For instance we made sure to only use one variable in each column, and we only used one variable when describing data that could not be found in the dataset.

Furthermore, when dates and other data was not available to us we wrote NA (not applicable) to make it clear to the reader that we know that the cell doesn't contain data.

Link to spreadsheet:

https://docs.google.com/spreadsheets/d/1IQ9IExIS9EuHBU_Edc5_f8uuoa3WfEsj_X8DEJCHqs8/edit#gid=0

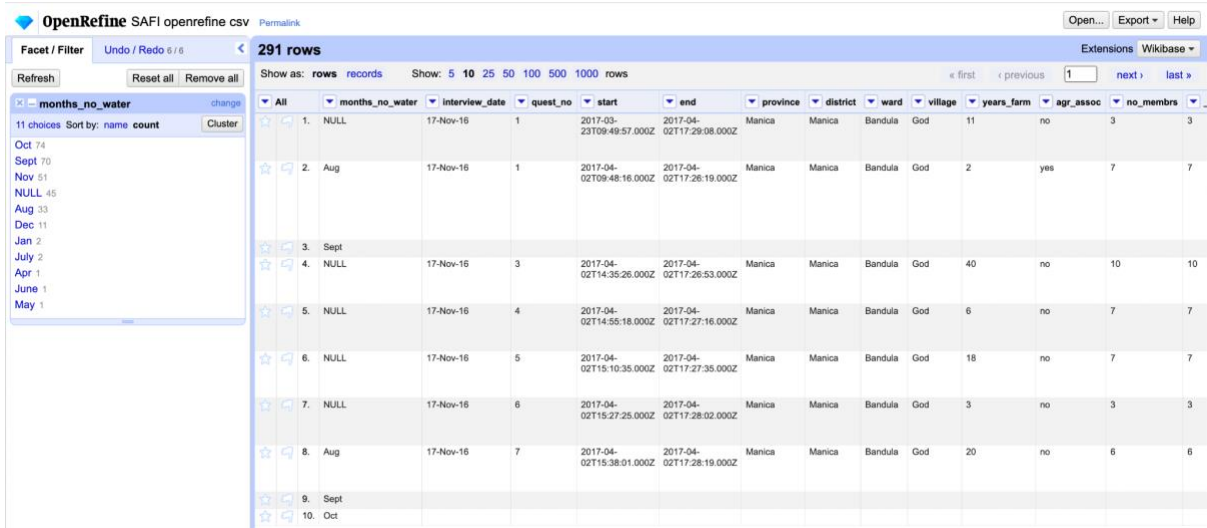
2. Does OpenRefine alter the raw data during sorting and filtering?

Openrefine does not alter the raw data. Openrefine generates a code in which you are able to see your modifications. You can always find the different steps and alterations under the function called "undo/redo"

It can be useful to extract to save with the data in order to reproduce the "experiment"

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?"

Made by Lasse, Kristian, Gustav and Clara



0. Create project

1. Move column months_no_water to position 0

2. Text transform on 0 cells in column months_no_water: `grel:value.split(",")`

3. Text transform on 86 cells in column months_no_water: `grel:value.replace("[', ''].replace('\"', '\"').replace('\"', '\"').replace('\"', '\"')`

4. Text transform on 0 cells in column months_no_water: `grel:value.split(",")`

5. Text transform on 0 cells in column months_no_water: `grel:value.split(",")`

6. Split multi-valued cells in column months_no_water

https://docs.google.com/document/d/1txvsn3FxAOL72md-NxWO38BoTXf-CtgNBji6ky9uXY/edit?addon_store

The two months which are reported to be the most water-deprived/driest by the interviewed farmer households are October (80) and September (70).

We fixed the dataset and found the answer by:

- Moving the column called “month_no_water” to the beginning of the spreadsheet in OpenRefine.
- Then we transformed the text by using GREL, where we removed the square brackets, space and apostrophe. The text in the cells were now separated by semicolons instead.
- Afterwards, we used the function “split multi-valued cells” under “edit cells”. This made it possible to split the months into separate values.