# Homework W37: OpenRefine

**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**](https://kongehuset.dk/monarkiet-i-danmark/kongerakken)**, but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)**

*Table listing the names of Danish monarch (raw data):*

Et billede, der indeholder tekst, menu, skærmbillede, dokument

Automatisk genereret beskrivelse

*Data dictionary in Danish:*

Et billede, der indeholder tekst, kvittering, skærmbillede, Font/skrifttype

Automatisk genereret beskrivelse

**2.**

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

No – OpenRefine lets the user sort and filter through the data without it affecting the raw data itself.

**3.**

**Fix the**[**interviews dataset**](https://ndownloader.figshare.com/files/11502815) **in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"**

The column containing the data needed to examine which months were reported as the driest is the one named months\_no\_water. However, the data are not in a format eligible by OpenRefine to analyze and therefore need to be transformed.

This is done by clicking on the drop-down menu left of the column name. Under “Edit Cells” there is an option to “Transform…” the data. When clicked on, this will lead to a pop-up window, where the following GREL expression is entered:

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

This action removes any single quotation marks, square brackets, and white spaces. The individual months can then be separated into single points of data that can be read by OpenRefine by clicking on the “Custom text facet…” option under “Facet” in the same drop-down menu as before. Another pop-up window will appear, where the following expression is entered:

value.split(";")

The data in the newly created Text Facet is then sorted by count, whereby it can be concluded that October and September were the two most water-deprived months as reported by the interviewed households.

Et billede, der indeholder tekst, skærmbillede, display/skærm/fremvisning, software

Automatisk genereret beskrivelse