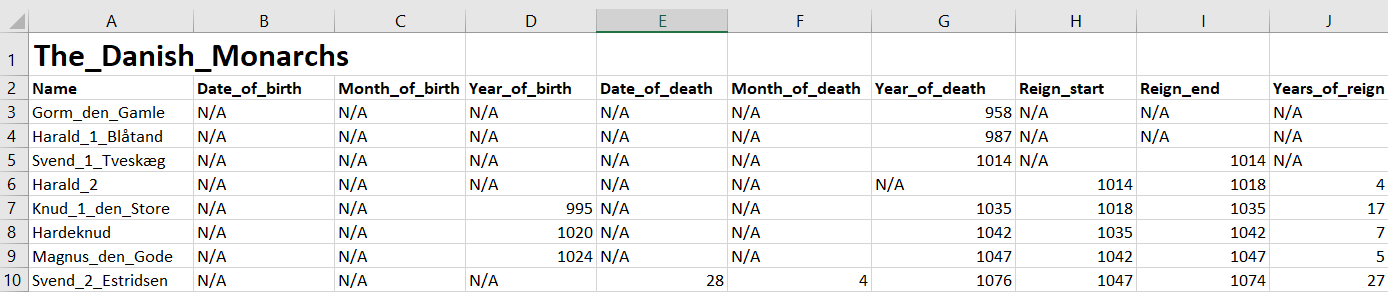
Homework for week 45: Open Refine

Submission on the 16.11.2021

# Create a \*tidy\* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and duration of reign.

– Files are in Brightspace submission (both CVS-version and original version)

This is a screenshot of the document:



**Process:**

* I made sure to only leave one value in each cell, and therefor separated the date, month and year of both birth and death.
* Moreover, I made separated start of reign and end of end of reign. And lastly I calculated the total years of reign
* I did not include periods without a regent. Which were the following years: 1332-1340 and 1533-1536.
* To make sure I had no empty cells I wrote “N/A” everytime I couldn’t find a date, or didn’t have any data to put in.
* I made sure to be consistant in my dataimputs, so all names an values are easier to comprehend. (I might not have done a good job in this aspect tho)
* I made sure to never write any space in the titles.
* In all the royals names I put in \_ to avoid space

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

No, it does not alter the data, but it can alter the layout of the document.

# 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?"

**Process**:

1. First, I reorder the tables so column saying “months\_no\_water” is placed first. That way it’s easier to access. I reorder it by clicking the “all” button and
2. To answer the question, I need to isolate the months that currently are written in the same cell.
3. First I edit the cells 🡪 Transform… 🡪 value.replace(“[“,””).replace(“]”,””).
4. I still have to isolate the months, so I remove the “’” with the code value.replace(“’”,””)
5. Then I split the multivalued cell and get each month in only one row.
6. Since some of the months are showed multiple times in the text facet I recluster the words by merging them into one. So the months are now only showed once in the text facet.
7. Now the count of the months is showed in the text facet and we are able to answer: “Which two months are reported as the most water-deprived/driest by the interviewed farmer households?”

**The answer to the question is September and October.**

See screenshot of the answer below:

