**Homework 2**

**Question 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)

We used the provided link to get the data about the Danish monarchs and doublechecked on google when we were unsure about dates. We manually transcribed all dates into a spreadsheet (Excel) in the following manner : Name, fødsel\_år, død\_år, start\_regeringstid og slut\_regeringstid. At some of the older monarchs there is some insecurity about the exact year for either their birth, death or start and beginning of their reign, therefore we have written NA to indicate a missing value. There were also cases where there are some insecurities about years, but where we have chosen to include as many as we could. Also, during the overlapping period where the three kings Svend, Knud and Valdemar reigned at the same time, we decided to include all of them.

Here is a screenshot of the dataset :

Et billede, der indeholder tekst, skærmbillede, menu, nummer/tal

Automatisk genereret beskrivelse

Et billede, der indeholder tekst, nummer/tal, menu, Font/skrifttype

Automatisk genereret beskrivelse

I also checked it in OpenRefine to see how it looked :

Et billede, der indeholder tekst, skærmbillede, nummer/tal, menu

Automatisk genereret beskrivelse

**Et billede, der indeholder tekst, skærmbillede, nummer/tal, menu

Automatisk genereret beskrivelse**

**Question 2 -** Does OpenRefine alter the raw data during sorting and filtering?

First of all, when you upload raw data, it shouldn’t normally alter any of the raw data given that the program accepts a wide range of file types. Here after, it is possible to alter the data to any given purpose. OpenRefine saves all the changes, and it is possible to go back to each change, to either see what we did, to delete the change or to start over totally / go back to the raw data.

Are we talking about our specific case with kongerækken, then no, the program did not alter any of our data after uploading it.

**Question 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?"

We will explain the exact steps to find our result in the following :

First of all, we uploaded CSV file to OpenRefine and created a project.

The we located the column with the data about months without water, named months\_no\_water, and started the cleaning process.

First, we wanted to clean and remove the values surrounding the months like [ ] ‘ and the spaces. We did this by clicking *edit cells* and then *transform*, here we wrote the regular expression : value.replace("[", "") and clicked ok. We selected the value in each cell and replaced all of the [ with nothing. We did this for each value and cleaned them, so it looked like the following (this is only a small part of it to give an example).

Et billede, der indeholder tekst, skærmbillede, Font/skrifttype, nummer/tal

Automatisk genereret beskrivelse

Et billede, der indeholder tekst, elektronik, skærmbillede, software

Automatisk genereret beskrivelse

Our data for this column now only separated by ;

Afterwards we created a custom text facet. We clicked *facet* and then *custom text facet*. Here we wrote the regular expression value.split(";"). We did this to make the program count the values surrounded by ; instead of having it count the values by boxes, to look at and count the months separately. This gave us a box to the left which we sorted by *count* and here we could see that the two driest months was October with 74 and September with 70.

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

Automatisk genereret beskrivelse