**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, but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)*

My data source was found here:

<https://www.kongehuset.dk/monarkiet-i-danmark/kongerakken>

I also doublechecked some of the dates on google when I was unsure about the correct dates.

It is a bit problematic that some of the data about the old kings is missing, but the overall picture is still there.

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

Automatisk genereret beskrivelse**

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

Automatisk genereret beskrivelse**

**Question 2**

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

No, it doesn’t alter anything, it still looks like it did in my excel spreadsheet/table.

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

Automatisk genereret beskrivelse**

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

Automatisk genereret beskrivelse**

**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?"*

I will explain the exact steps to find my result in the following:

First, I uploaded the CSV file to OpenRefine and created a project.

Then I located the row with the data about months without water, named *Months\_no\_water*, and started the cleaning process.

First, I wanted to clean and remove the values[[1]](#footnote-1) surrounding the months such as: [ ], ‘ and the spaces. I did this by clicking *edit cells* and then *transform*, here I wrote the regular expression: value.replace("[", "") and clicked ok. I selected the value in each cell and replaced all the [ with nothing (note that that is way there is nothing between the corotation marks (“”) in the regular expression). I did this to each value and thereby cleaned the data.

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

This is a screenshot of how the table looked, after I cleaned the data.

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

Automatisk genereret beskrivelse**

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

Automatisk genereret beskrivelse**

The first box in this screenshot is how OpenRefine counted the data before I used the split regular expression, and the second is how it sorted the data afterwards. By looking at it, it becomes easy to interpret the driest months (October and September).

A useful link for future use of OpenRefine:

[**https://datacarpentry.org/openrefine-socialsci/02-working-with-openrefine.html**](https://datacarpentry.org/openrefine-socialsci/02-working-with-openrefine.html)

1. Note that I in my explanation uses the term “value”, where I would normally use the term “symbol”, since it is what it is called in the language of regular expressions. [↑](#footnote-ref-1)