**OpenRefine and Spreadsheets**

1. Create a \*tidy\* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and start and end of reign. They should be sortable by year of birth. Suitable source websites 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)

Attached to Brightspace.

Some of the birth dates of the monarchs are indistinguishable and I have therefore chosen to mark them NA (no answer). Alternatively, I could’ve decided upon one birth date.

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

No. OpenRefine does not alter the raw data. Instead, OpenRefine scripts the changes made during sorting and filtering. This allows you to always track the changes you did and access the raw data when I doubt.

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

The problem when wanting to answer the question above, is that several months appear in the same cell. This makes it difficult for the computer to differentiate the months and thus does not give a clear answer.

Et billede, der indeholder tekst

Automatisk genereret beskrivelse

In order to split up the months and get a cumulated answer I did the following:

|  |  |  |
| --- | --- | --- |
| Step | Description | Screenshot |
| 1. | Upload dataset in csv file format into OpenRefine. |  |
| 2. | Recognize which columns refer to water. The column “months\_no\_water” shows the months which were considered the driest by the interviewed farmer households. |  |
| 3. | I then moved the column to the beginning to better manage the data. |  |
| 4. | I then pressed the arrow next to “months\_no\_water”, went to “edit cells” and then to “transform.” |  |
| 5. | I then inserted the regular expression into the custom text transform window:  value.replace(“[“,””).replace(“]”,””).replace(“’”,””).replace(“ “,””)  This eliminates unwanted symbols. |  |
| 6. | I then go to “facet” and “custom text facet.” |  |
| 7. | I inserted the regular expression into the custom text facet window:  value.split(“;”)  This splits up the months. |  |
| 8. | I then end up with this text facet. |  |

This text facet allows me to answer the question: Which two months are reported as the most water-deprived/driest by the interviewed farmer households?

October and September are the months which are reported as the most water-deprived/driest by the interviewed farmer households. 74 farmers reported October as a month without water and 70 farmer reported September as a month without water.

The script is inserted below:

[

{

"op": "core/column-move",

"columnName": "months\_no\_water",

"index": 0,

"description": "Move column months\_no\_water to position 0"

},

{

"op": "core/text-transform",

"engineConfig": {

"facets": [],

"mode": "row-based"

},

"columnName": "months\_no\_water",

"expression": "grel:value.replace(\"[\",\"\").replace(\"]\",\"\").replace(\"'\",\"\").replace(\" \",\"\")",

"onError": "keep-original",

"repeat": false,

"repeatCount": 10,

"description": "Text transform on cells in column months\_no\_water using expression grel:value.replace(\"[\",\"\").replace(\"]\",\"\").replace(\"'\",\"\").replace(\" \",\"\")"

}

]