W45: Open Refine

Create a \*tidy\* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and duration of reign. They should be sortable by year of birth.

Spreadsheet with the Danish monarchs

<https://docs.google.com/spreadsheets/d/1WQwQh4STPjD4DAv8gFnZ53KjtbPCfDwriULrgK448vw/edit?usp=sharing>

Docs that explain the spreadsheet

<https://docs.google.com/document/d/1nwjCTqA_eFeN3nXqS9n2k1oBjvwj6TyNOqomnuKEfL4/edit?usp=sharing>

Does OpenRefine alter the raw data during sorting and filtering?

No, the point of the program is to put your data in order. To hide unimportant data and focus on the important part. Therefore the raw data itself is not changed but the display of it is.

Fix the interviews dataset in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"

Firstly, I moved the months\_no\_water column to the beginning so it was easier to keep track of

Secondly, I transformed the cell with the code: value.replace("[","").replace("]","").replace("'","") to get rid of [ and ] and ‘

Thirdly, I split the multi-valued cells in column months\_no\_water by ; so that the months stood separately

Fourthly, I merged all the clusters and rearranged them by count in the text facet

This was the result: October was the driest month followed by September

This would be the code for the excise

[

{

"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(\"'\",\"\")",

"onError": "keep-original",

"repeat": false,

"repeatCount": 10,

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

},

{

"op": "core/multivalued-cell-split",

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

"keyColumnName": "months\_no\_water",

"mode": "separator",

"separator": ";",

"regex": false,

"description": "Split multi-valued cells in column months\_no\_water"

},

{

"op": "core/mass-edit",

"engineConfig": {

"facets": [],

"mode": "row-based"

},

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

"expression": "value",

"edits": [

{

"from": [

" Oct ",

" Oct",

"Oct ",

"Oct"

],

"fromBlank": false,

"fromError": false,

"to": " Oct "

},

{

"from": [

" Nov",

" Nov ",

"Nov ",

"Nov"

],

"fromBlank": false,

"fromError": false,

"to": " Nov"

},

{

"from": [

"Sept ",

" Sept ",

" Sept"

],

"fromBlank": false,

"fromError": false,

"to": "Sept "

},

{

"from": [

"Aug ",

" Aug "

],

"fromBlank": false,

"fromError": false,

"to": "Aug "

},

{

"from": [

" July ",

"July "

],

"fromBlank": false,

"fromError": false,

"to": " July "

}

],

"description": "Mass edit cells in column months\_no\_water"

}