Lea Skriver Hansen
Digithal methods for historians
17/11/2021

Weekly homework 2: Openrefine

Task 1: In collaboration with Andreas Emil Mikkelsen

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. Suitable source websites are here and here, but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission):

Link to spreadsheet:
https://docs.google.com/spreadsheets/d/1WQwQh4STPjD4DAv8gFnZ53KjtbPCfDwriULrgK448vw/edit?usp=sharing

- This spreadsheet is using the danish royal website as the source
- When writing the names of the kings and the queen we put _ in instead of a space, so we did not confuse the spreadsheet and mess with the data
- When we discovered there was two periods where there was no one in the reign we decided to call the second period interregnum_2, so we did not confuse the spreadsheet
- We could not find a digital method that worked to transfer the dates directy, but we succeeded in doing it manually
- When we should figure out how long their reign lasted we used the formular =I()-H() since the I column was the one with the end of the periode the person reigned and H the column with the beginning of the reign

Task 2:

Does OpenRefine alter the raw data during sorting and filtering?

- No, it does not because the program function is to make the data more manageable and easier to work with

Task 3:

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?"

- I started by moving the column named no_water to the front so it was easier to access. Then I clicked on edit cell and then on transform, in transform I entered the following code: value.replace("[","").replace("]","").replace("""",""). After I had entered the code and removed all the [] and , and ‘, and spaces. After that I clicked on edit cell → split muliti- valued cells to get the months to stand individualy by using; as seperater. Then I clicked on cluster in the

facet/filter and instructed openrefine to merge all the months. After that step I sorted the months by count and figured out that October and September were the driest months.

The code for this task:

```
[
  {
    "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/text-transform",

"engineConfig": {

"facets": [],

"mode": "row-based"

},

"columnName": "months_no_water",

"expression": "grel:value",

"onError": "keep-original",

"repeat": false,

"repeatCount": 10,

"description": "Text transform on cells in column months_no_water using expression grel:value"

},

{

"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 "

],

"fromBlank": false,

"fromError": false,

"to": " Oct "

},

{

"from": [

" Aug ",

" Aug "

],

"fromBlank": false,

"fromError": false,

"to": " Aug "

},

{

"from": [

" Sept ",

" Sept "

],

"fromBlank": false,

"fromError": false,

"to": " Sept "

},

{

"from": [

" Nov ",

" Nov "

],

"fromBlank": false,

"fromError": false,

"to": " Nov "

},

{

"from": [

```
          "  July ",

          " July "

        ],

        "fromBlank": false,

        "fromError": false,

        "to": "  July "

      }

    ],

    "description": "Mass edit cells in column months_no_water"

  }

]
```