


## 2:W35: Open Refine

### DESCRIPTION

Upload your answers to these questions:

1. Create a spreadsheet listing the names of Danish monarchs with their birth- and death-date and start and end year of reign. Make it *\*tidy\**! 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. (Group collaboration is expected and welcome. Remember to attach this spreadsheet to Brightspace submission)

 kings\_data

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

No it does not

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?"
  - a. Started with text transform for the column months\_no\_water
    - i. Text transform on 86 cells in column months\_no\_water: `grel:value.replace("[", "").replace("]", "").replace(";", "")`
  - b. Then saw the months are separated with a “;” so used that to split the cells with just one month in each cell
    - i. Split multi-valued cells in column months\_no\_water
  - c. Then facet by text and use the count function, Nov and Oct show as the two months most water deprived



months_no_water	
21 choices	Sort by: name count
NULL	45
Nov	41
Oct	38
Sept	37
Aug	31
Sept	27
Oct	25
Dec	11
Oct	9
Nov	7
Sept	6

4. Real-Data-Challenge: What are the 10 most frequent occupations (erhverv) among unmarried men and women in [1801 Aarhus](#)? (hint: some expert judgement interpretation)

is necessary, look at the [HISCO classification](#) "Historical International Standard of Classification of Occupations" on [Dataverse](#) if ambitious)

- Cluster editing the erhverv column to get rid of multiple spellings
- Use text facet the civilstand column and pick ugift

×

civilstand

change invert reset

6 choices

Sort by: name count

Cluster

enke 1531

enkemand 694

gift 16617

separeret 5

skilt 6

ugift 25440

(blank) 266

exclude

Facet by choice counts

- Use text facet again for erhverv, sort my count and then can see the top 10

×

erhverv

change

766 choices

Sort by: name count

Cluster

nationalsoldat 207

soldat ved 1. Jyske Inf. Reg. 95

Landsoldat 69

Tienestepige 67

Tienestekarl 54

læredreng 52

Væver 45

Bonde og Gaardbeboer 37

hospitalslem 36

Vanfør 35

Tienestedræng 33