

2:W35: Open Refine

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)
 - a. Link to spreadsheet:
https://docs.google.com/spreadsheets/d/18bc1U0Xa_XkwxHNv9C_OYTzzFO1oGQqC6_3kZ7mYYeA/edit?usp=sharing
2. Does OpenRefine alter the raw data during sorting and filtering?
 - a. No, it only happens at the UI in OpenRefine.
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. Go into custom facet on column and
 - i. `value.replace("[", "").replace("]", "").replace("","").replace(" ", "")`
 - ii. And then split by: `value.split(";")`
 - iii. `Value.split(";")`
 - b. By then I get the result: November 41
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)
 - a. I could do this in OpenRefine by loading the data by the URL, apply text facet, and sort by count. The top 10 occupations are the listed as below. I am aware that we have a problem with naming, namely several of the variables are called by different names although the occupation is the same. This could be easily fixed in R with the `replace()` function or something similar.

b. I achieve this with the following command in R:

i. `data.frame(sort(table(df$erhverv), decreasing = TRUE)[1:10])`

Variable	Frequency overall
Bonde og Gaardbeboer	1780
Huusmand med Jord	660
bonde og gaardbeboer	231
bonde og gÅrdbbeboer	223
soldat ved 1. Jyske Inf. Reg.	136
Bonde og Gaard Beboer	102
National Soldat	99
Bonde og Gaardmand	98
Gaardbeboer	97
nyder Ophold af Gaarden	90

Variable	Frequency for unmarried women
Tienestepige	61
hospitalslem	21
Væverske	18
Lever af at spinde	17
Inderste	14
Tjener faderen	13
Lever af almisse	12
Spinder	12
Vanfør	10
almisselem	9

Variable	Frequency for unmarried men
National Soldat	96
Soldat ved 1. Jyske Inf. Reg.	94
nationalsoldat	61
Tienestesoldat	47

læredreng	42
Nationalsoldat	36
Bonde og Gaardbeboer	32
indluddeitTienestedræng	32
Væver	32
gårdskarl	30