Nikita Cecilie Jensen
19 – 09 – 2023

Teoriorienteret metode

Homework week 37

1. **Create a *tidy* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and start and end of their reign. They should be sortable by year of birth. Suitable source website is for example here, but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)**

SE VEDHÆFTET EXCEL ARK

Link:
https://1drv.ms/x/s!ArmiJDUiWX_7zB7GBdxSe1667hOA

Name	Birth	Death	Start_Reign	End_Reign
Gorm den gamle	NA	958	936	958
Harald Blåtand	NA	986	958	986
Svend Tveskæg	NA	1014	986	1014
Harald 2	NA	1018	1014	1018
Knud den Store	995	1035	1018	1035
Hardeknud	1020	1042	1035	1042
Magnus den Gode	1024	1047	1042	1047
Svend Estridsen	NA	1074	1047	1074
Harald Hen	NA	1080	1074	1080
Knud den Hellige	NA	1086	1080	1086
Oluf Hunger	NA	1095	1086	1095
Erik Ejegod	1056	1103	1095	1103
Niels	NA	1134	1104	1134
Erik Emune	NA	1137	1134	1137
Erik Lam	NA	1146	1137	1146
Svend Knud Valdemar	NA	NA	1146	1157
Valdemar den Store	1131	1182	1157	1182
Knud 4	1163	1202	1182	1202
Valdemar Sejr	1170	1241	1202	1241
Erik Plovpenning	1216	1250	1241	1250
Abel	1218	1252	1250	1252
Christoffer	1219	1259	1252	1259
Erik Klipping	1249	1286	1259	1286
Erik Menved	1274	1319	1286	1319
Christoffer 2	1276	1332	1319	1326
Valdemar 3	1315	1364	1326	1329
Christoffer	1276	1332	1329	1332
Valdemar Atterdag	1320	1375	1340	1375

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

When you use OpenRefine for sorting and filtering data, it does not alter the raw data itself. Instead, it provides a way to view, manipulate, and work with your data in a user-friendly interface without making permanent changes to the original data source.

53 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All	Name	Birth	Death	Start_Reign	End_Reign
1.	Gorm den gamle	NA	958	936	958
2.	Harald Blåtand	NA	986	958	986
3.	Svend Tveskæg	NA	1014	986	1014
4.	Harald 2	NA	1018	1014	1018
5.	Knud den Store	995	1035	1018	1035
6.	Hardeknud	1020	1042	1035	1042
7.	Magnus den Gode	1024	1047	1042	1047
8.	Svend Estridsen	NA	1074	1047	1074
9.	Harald Hen	NA	1080	1074	1080
10.	Knud den Hellige	NA	1086	1080	1086
11.	Oluf Hunger	NA	1095	1086	1095
12.	Erik Ejegod	1056	1103	1095	1103
13.	Niels	NA	1134	1104	1134
14.	Erik Emune	NA	1137	1134	1137
15.	Erik Lam	NA	1146	1137	1146
16.	Svend Knud Valdemar	NA	NA	1146	1157
17.	Valdemar den Store	1131	1182	1157	1182
18.	Knud 4	1163	1202	1182	1202
19.	Valdemar Sejr	1170	1241	1202	1241
20.	Erik Plovpenning	1216	1250	1241	1250
21.	Abel	1218	1252	1250	1252
22.	Christoffer	1219	1259	1252	1259
23.	Erik Klipping	1249	1286	1259	1286
24.	Erik Menved	1274	1319	1286	1319

Data remains intact in its original form when you import it to OpenRefine. You can sort the data based on specific columns or criteria within OpenRefine, and this sorting operation only affects the way data is displayed within the application. It does not change the original order or structure of the data in your source file. You can apply filters to your data to show only specific rows that meet certain criteria.

Sorting and filtering in OpenRefine are non-destructive, and it doesn't permanently remove any data from your source. It only controls what you see in the application's interface. The original data remains unchanged unless you explicitly apply a transformation to modify it.

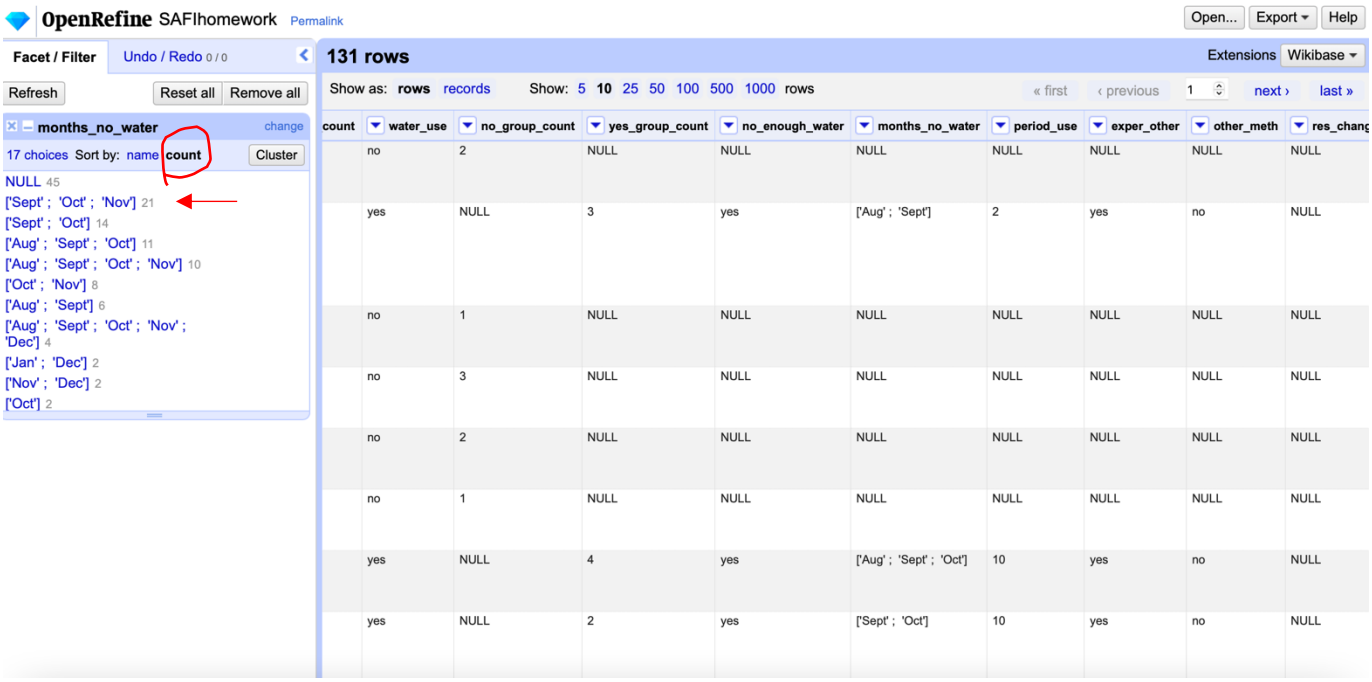
So OpenRefine is a valuable tool for data cleaning and preparation without risking accidental data loss or permanent changes to your source data.

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?"**

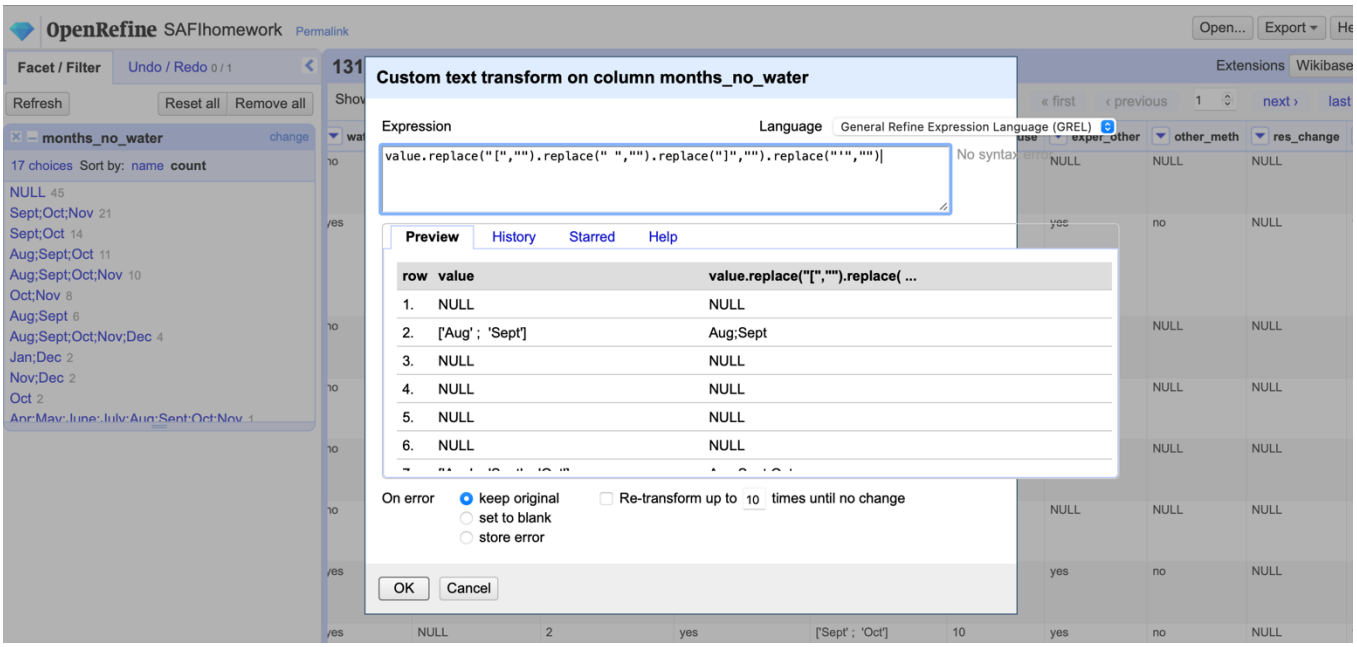
October and September. For my result see the steps + screenshots below.

Steps:

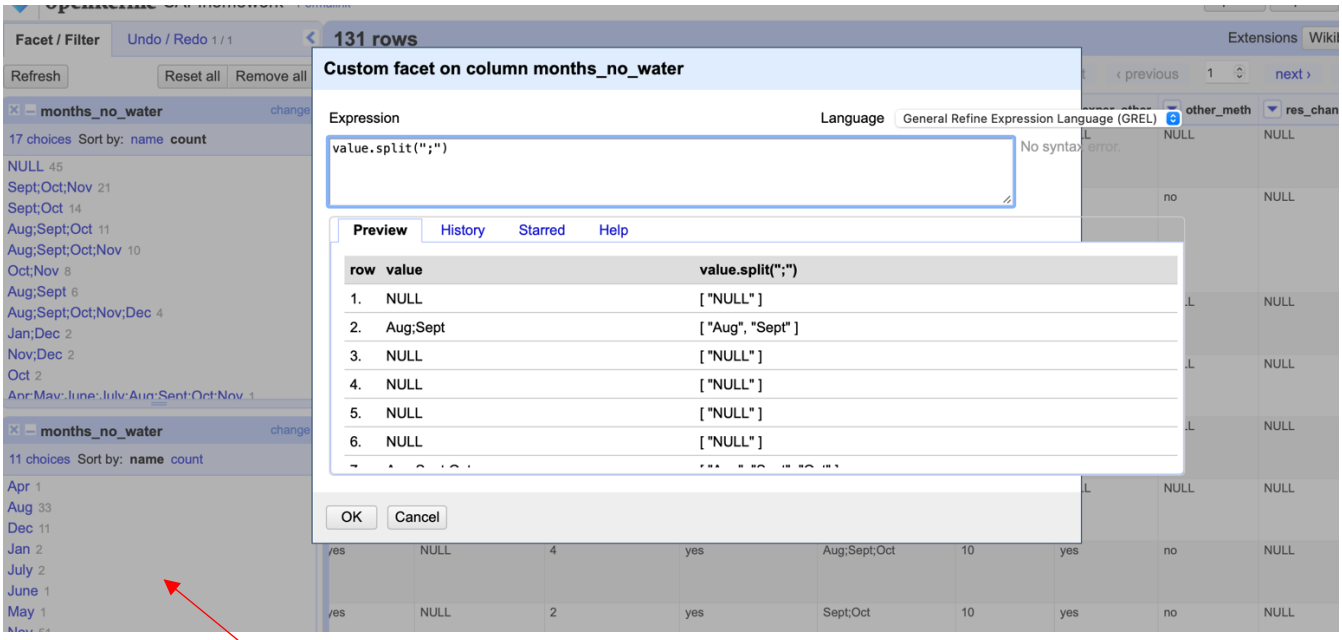
1. I Imported the data to OpenRefine.
2. I found the variable "month_no_water".
3. I clicked on the name and then clicked on "Facet".
4. I clicked on "Text facet".
5. So, I got the Facet/Filter box on the left.
6. I clicked on "count" on the facet box (se screenshot – red circle).
7. The box tells me that "Sept, Oct, Nov" is the most water-deprived months.



8. I wanted a more specific answer to the question, so I customized the facet to split the month groups, so the months became separated. I clicked “edit cells”, and then “transform”. I used regular expression to replace first. See screenshots below →



9. Then I split the months, by clicking “Facet” → “Custom text facet” →

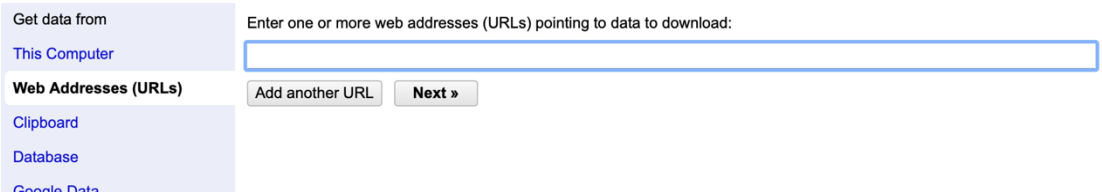


10. A new text facet appeared on the left, where I could find my result by count.

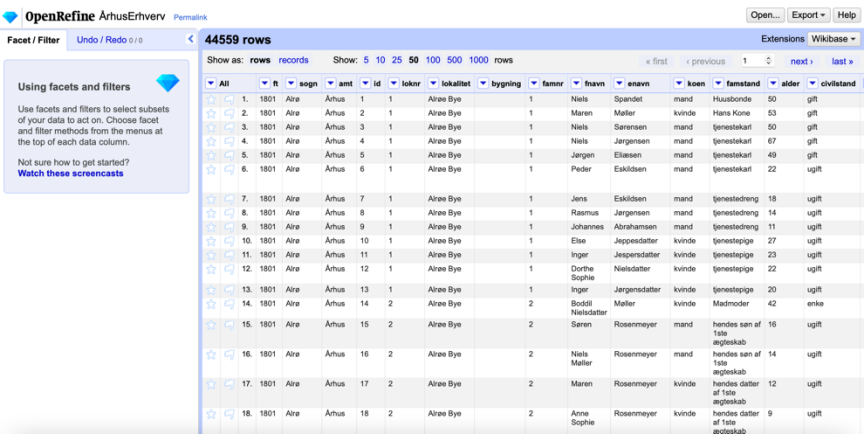
11. Result. October (74) and September (70) are the most water-deprived month. (see screenshot below)

4. **OPTIONAL 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. As an inspiration, [check out this chapter Making a living outside marriage from the] Swedish Gender and Work project of Maria Agren)**

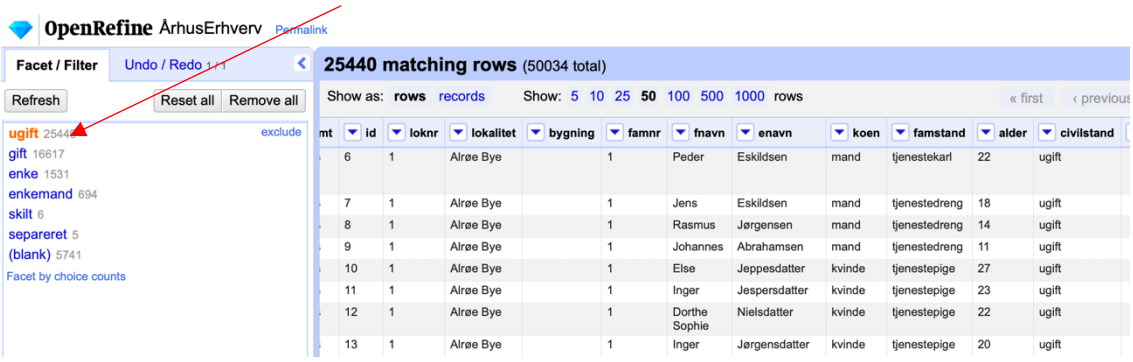
I used OpenRefine – and copied the URLs to open the data.



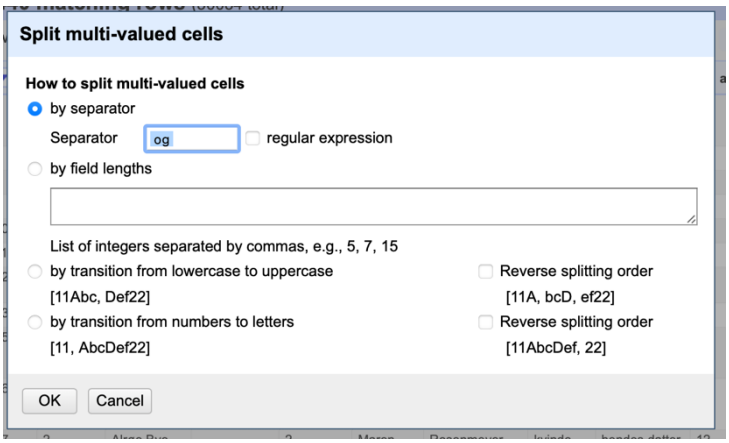
The data looked like this:



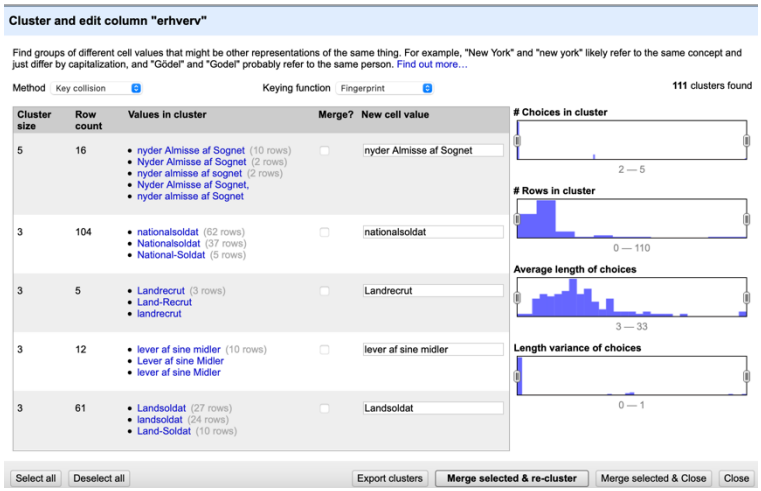
I clicked on ”include” in the text facet on the ”civilstande” I wanted. In this case ”ugift”.



Then I open a new text facet for “erhverv” and get a lot of mess, which needed to be cleaned. First, I separated people with several different “erhverv” by using: “edit cells” → “split multi-valued cells”.



And then I just keept cleaning… with multiple tools. Ex. “Cluster and edit”.



After more time cleaning, I got all the possible “erhverv” down to 435 choices. So, this is what my result was after some time… I didn’t get further with cleaning.