Homework 2

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

It changes only the raw data you retrieves in the program, not in the file save on your desktop. You can save a new edition, with the refines.

2. 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?"

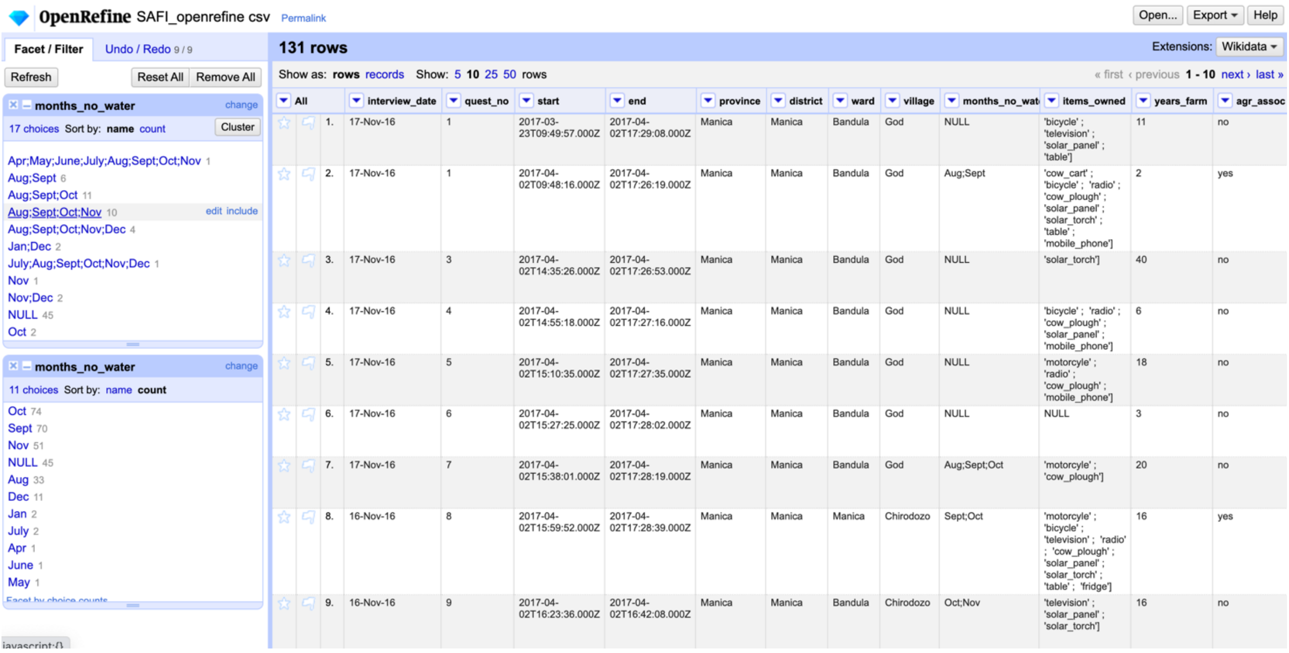
Step 1; Find a column with the right information 'Months_no_water' (I belive).

Step 2; Was to use regex, to replace the squarebrackets, citationssigns, and spaces, with nothing

value.replace("[","") value.replace("[","") value.replace("[",",","")

Step 3; After this, we entered custom text facet, and made the list split on semicolon

Value.Split(";")



months_no_water

17 choices Sort by: name count

Apr;May;June;July;Aug;Sept;Oct;Nov 1
Aug;Sept 6
Aug;Sept;Oct 11
Aug;Sept;Oct;Nov 10
Aug;Sept;Oct;Nov;Dec 4
Jan;Dec 2
July;Aug;Sept;Oct;Nov;Dec 1
Nov 1
Nov;Dec 2
NULL 45
Oct 2

months_no_water

11 choices Sort by: name count

Oct 74
Sept 70
Nov 51
NULL 45
Aug 33
Dec 11
Jan 2
July 2
Apr 1
June 1
May 1

Facet by choice counts

131 rows

Show as: rows records Show: 5 10 25 50 rows

« first ‹ previous 1 - 10 next › last »

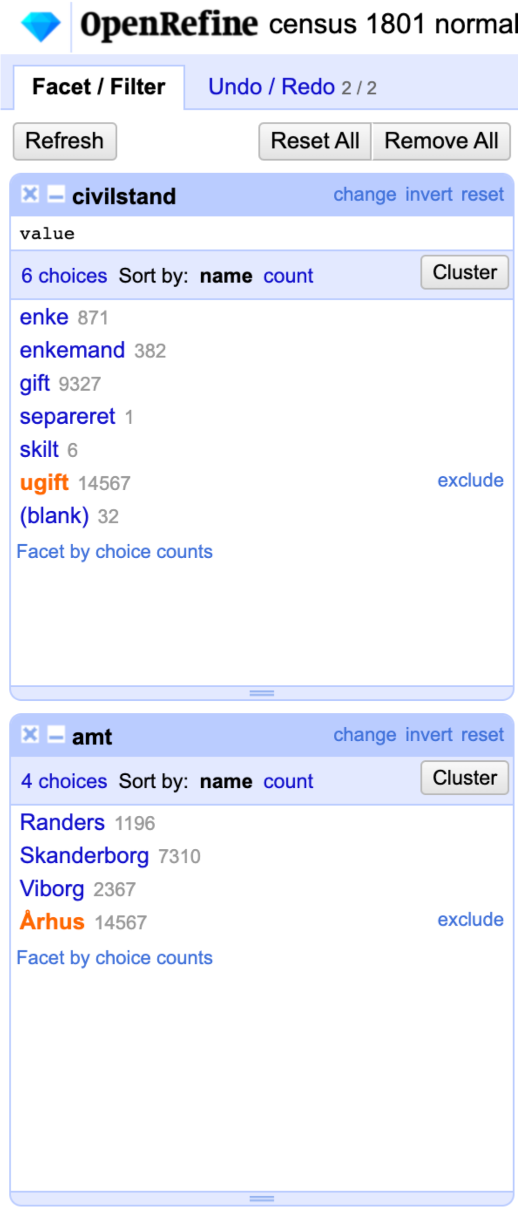
All	interview_date	quest_no	start	end	province	district	ward	village	months_no_wat	items_owned	years_farm	agr_assoc
1.	17-Nov-16	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manica	Bandula	God	NULL	'bicycle' ; 'television' ; 'solar_panel' ; 'table']	11	no
2.	17-Nov-16	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manica	Bandula	God	Aug;Sept	'cow_cart' ; 'bicycle' ; 'radio' ; 'cow_plough' ; 'solar_panel' ; 'solar_torch' ; 'table' ; 'mobile_phone']	2	yes
3.	17-Nov-16	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manica	Bandula	God	NULL	'solar_torch']	40	no
4.	17-Nov-16	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manica	Bandula	God	NULL	'bicycle' ; 'radio' ; 'cow_plough' ; 'solar_panel' ; 'mobile_phone']	6	no
5.	17-Nov-16	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manica	Bandula	God	NULL	'motorcyle' ; 'radio' ; 'cow_plough' ; 'mobile_phone']	18	no
6.	17-Nov-16	6	2017-04-02T15:27:25.000Z	2017-04-02T17:28:02.000Z	Manica	Manica	Bandula	God	NULL	NULL	3	no
7.	17-Nov-16	7	2017-04-02T15:38:01.000Z	2017-04-02T17:28:19.000Z	Manica	Manica	Bandula	God	Aug;Sept;Oct	'motorcyle' ; 'cow_plough']	20	no
8.	16-Nov-16	8	2017-04-02T15:59:52.000Z	2017-04-02T17:28:39.000Z	Manica	Manica	Manica	Chirodozo	Sept;Oct	'motorcyle' ; 'bicycle' ; 'television' ; 'radio' ; 'cow_plough' ; 'solar_panel' ; 'solar_torch' ; 'table' ; 'fridge']	16	yes
9.	16-Nov-16	9	2017-04-02T16:23:36.000Z	2017-04-02T16:42:08.000Z	Manica	Manica	Bandula	Chirodozo	Oct;Nov	'television' ; 'solar_panel' ; 'solar_torch']	16	no

October and September is the driest months.

3. What were the 10 most frequent occupations (erhverv) among unmarried men and women in 1801 Aarhus?

Step 1; So I simply just toke 'cevilstand', made a text facet, and chose "ugift"

Step 2; toke 'amt', made a text facet, and chose "Århus"



OpenRefine census 1801 normal

Facet / Filter | Undo / Redo 2 / 2

Refresh | Reset All | Remove All

civilstand change invert reset

value

6 choices Sort by: name count | Cluster

enke 871
enkemand 382
gift 9327
separeret 1
skilt 6
ugift 14567 exclude
(blank) 32

Facet by choice counts

amt change invert reset

4 choices Sort by: name count | Cluster

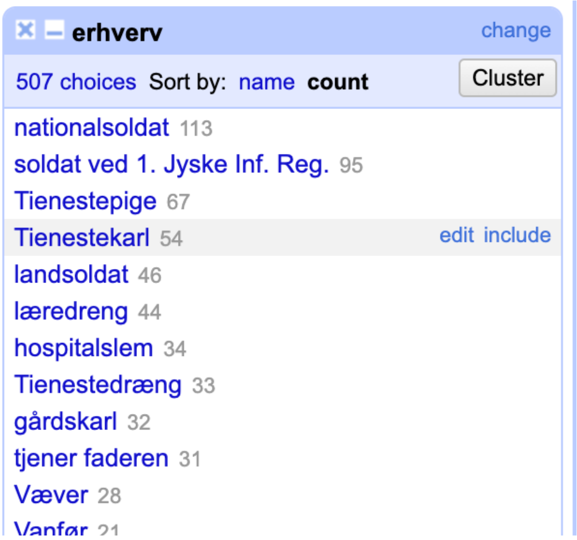
Randers 1196
Skanderborg 7310
Viborg 2367
Århus 14567 exclude

Facet by choice counts

Step 3; Then I clustered as many as I could, with my knowledge, and the ones that seemed to be alike.

Step 4; Then I chose count, to find the top 10

Homework 2



erhverv

507 choices Sort by: name **count**

nationalsoldat	113
soldat ved 1. Jyske Inf. Reg.	95
Tienestepige	67
Tienestekarl	54
landsoldat	46
læredreng	44
hospitalslem	34
Tienestedræng	33
gårdskarl	32
tjener faderen	31
Væver	28
Vanfør	21

1. Nationalsoldat
2. Soldat ved 1. Jyske inf reg.
3. Tienestepige
4. Tienestekarl
5. landsoldat
6. læredreng
7. hospitalslem
8. tienestedræng
9. gårdskarl
10. tjener faderen

4; maybe if you clustered more occupations, the list would have different. However this would be a matter of judgement.