Digital Methods: Learning Journal Template

Maja Katrine Laier

Autumn 2020

8 09/11/2020

9 Intentions

11:04am: I'm currently looking at the OpenRefine homework from last week to get a quick overview of the information I need to answer question 2. Question 1 was answered in the articles we read as preparation for last monday.

11:30am: I am going away from the homework for a short while to focus on the lecture. Hopefully, it will get recorded, because I am feeling exhausted lately and have a hard time concentrating on all the new words and concepts.

16:30pm: After a break, I am at it again. I work the exercises we were given as homework. I have a pretty good idea as to how to solve question 3, but question 4 is a bit tricky.

21:03pm: Finally, I have finished, after much trial, looking at DataCarpentry´s guide, failing, restarting, and trying again. But I managed to finish and conclude an answer to every question. I am exhausted, but proud.

10 Question 1

What are the basic principles for using spreadsheets for good data organisation? (no more than 250 words)

Data-tidying in spreadsheets is a very important tool for data management. This it the case for many reasons. The most predominant case is that it is easy to view and gather useful information at a glance, another reason is that the dataset will now be malleable for research by the use of digital tools such as **Openrefine** or **RStudios**.

In the article "Tidy Data" by Hadley Wickham, he states three major principles that all spreadsheets should abide by:

1. **Each variable is placed in its own column**
2. **Each observation is placed in its own row**
2. **Each observational unit is placed in its own table**

Another focus worth a brief mention is **The FAIR Principles** of data publication.F.A.I.R. is an acronym for Findability, Accessibility, Interoperability, Reusability and creates common guidelines for how to handle data. This makes automatic machine reading of, for instance, a spreadsheet, easier.

11 Question 2

Does OpenRefine alter the raw data during sorting and filtering?

As far as I can read in the DataCarpentry lesson "Working with OpenRefine", it seems that the system does not alter the raw data, but rather creates a separate copy of the data you upload, so that you, at all times, have the original project and the version you have worked on.

12 Question 3

Fix the interview datasets in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"

At the beginning, I moved the column "months no water" to the beginning of the chart so I could see it more clearly. After this, i started the tidying process.

- Changed Facets to "text facet"

Then I cleaned the input by using the value.replace command in the "Edit Cells" section of the dropdown menu

- Drop down the arrow
- Click "Edit Cells"
- Click "Transform" the code written below

value.replace("[","").replace("]","").replace("""","").replace(" ","") so that you are left behind with a series of months, separated by a ;(semicolon).

Once this is done, I want to erase the semicolon, so that only the names of the months are left.

- Click "Custom text facets"
- Enter code **value.split(";")**

I am now left with a list on the far left of the page, where I click "count" and am left with the answer: **October(74) and September(70) are the months with the least amount of water**. This however is unclear, since there is a NULL count of 45, which might very well warp the data a relative amount. But for the information we have, this is the closest result we will get to.

13 Question 4

Finish the tutorial and report what village *you* think hides behind the number '49'.

This question bugged me for quite a while, I´m not gonna lie, but I do think that I have found a solution that is feasible.

13.1 GPS clue

I move the column "village" down to the GPS longitudes and latitudes in order to have everything close for an easier search. Then I format the Latitude values by the **sorting tool**.

- click the drop-down arrow in the column
- click Sort
- apply the tool to this column alone and click "numbers" to make OpenRefine identify the input as a numeric value instead of default text.

Once this is done, we see that the coordinates match loosely up to the village Chirodzo, but also to the village Ruaca.

13.2 Interview dates

To make sure, I match with the dates to see if there is any direct correspondence between when either village and nr. 49 was visited.

- Move the village names back to the front and place them next to the column "interview dates"
- look for a corresponding date between the village 49, Chirodzo and Ruaca.

We see a clear correspondence. The village 49 was visited on November 16th, the same day as a bulk results from the village named Chirodzo came in.

Answer: The village number 49 is Chirodzo.

13.3 Final Thoughts

I personally feel pretty good about this weeks assignments. They got me yanking my hair out at the beginning, but gradually, I found out where I could find the information I needed and who I could ask for help. I am starting to realise that this course by no means is a one-man-project.

The more I use the Overleaf and OpenRefine tools, the more sure I am that these tools will be used in the final assignment. I quite like the layout of the PDF i get from Overleaf.

What i would like now is some sort of graphics/visualization tool, because all the numbers and coding values are a bit tricky. Perhaps a graph or something might make things more understandable? for lay people?