# Digital Methods: Learning Journal Template

Mette Emily Kendon

Autumn 2020

## 1 09/11/2020

### 1.1 Thoughts / Intentions

**M**y thoughts coming into this lesson about spreadsheets was that it was going to be okay. I had read and understood the Spreadsheet tutorial to Quality Assurance and downloaded OpenRefine properly. At the hands-on-session I found out that cleaning up spreadsheets in OpenRefine was a bit more difficult than I had anticipated. My intention with the following exercises was therefore to get a better understanding of how to clean up messy spreadsheets and get a hold on how OpenRefine can be a great tool for doing this.

### 1.2 Exercises

Exercise no. 1: What are basic principles for using spreadsheets for good data organisation?

- It is important that you do not make multiple tables in the same tab. This can confuse the computer because you are drawing false associations between things for the computer.
- Do not leave empty cells because the computer can misinterpret blank cells. Use other values to indicate an empty cell.
- Be careful with formatting, even though you think it can make the tab look better it can compromise the computer's ability to see associations in the data. It is also important that you place the metadata in a separate file. Metadata includes notes.
- Do not enter more than one piece of information in each cell.
- When choosing the field names, it is important that you keep them short and descriptive. Try not to use spaces, numbers or any special characters.
- Make sure to structure your spreadsheet in a good manner. Is can be done by putting all the variables in columns, each observation in its own row, leave the raw data raw (try not to change the data outcomes too much)

and finally exporting the cleaned data to a text-based format like CSV. This last step is important because this ensures that anyone can use the data.

Exercise no. 2: Does OpenRefine alter the raw data during sorting and filtering?

- No, OpenRefine only keeps a copy of your data and do not modify your original data set.
- The "undo/redo" button in OpenRefine allows you to go back and see your previous steps but also to change them.

Exercise no. 3: Which two months are reported as the most water-deprived/driest by the interviewed farmer households?

- The first step was to open the spreadsheet that we had to use for this exercise in OpenRefine.
- Then I found the column, "months with no water" and moved it to the beginning of the tab to make it more manageable by clicking on the column and pressed the button Edit column and then move to beginning.
- Then I clicked on the column and pushed the button, "facets" and then "text facet". This opened a box on the left side of OpenRefine with an overview of all the months registered in this column. The facet makes it possible to clean up and transform the variables in the column.
- Then I used the drop down arrow on the "months with no water" and pressed edit cells and then transform, which opened a tab that allowed me to custom the text.
- In the box I wrote value.replace ("[","") and OK.
- I pressed edit cells and then transform again and wrote value.replace ("]","") OK.
- I pressed edit cells and then transform again and wrote value.replace (",","") OK.
- I pressed edit cells and then transform one last time and wrote value.replace (" ", "") OK.
- The reason for doing this was to remove [,] and ' from the column and change it into "
- Then because I wanted to split the words I pressed the column again and then "custom facet" and wrote value.Split (",") OK.
- I could now see that the result is that the driest months reported by the farmers households was September by 70 units and October by 74 units.

Exercise no. 4: Finish the tutorial and report what village you think hides behind the number '49'.

- To complete this exercise I followed the tutorial from the hands-on-session named "OpenRefine for Social Science Data" under the heading "Filtering and Sorting with OpenRefine".
- I followed all the steps in the tutorial and based on this I think that Chirodzo is the same as no. 49 because Chirodzo and 49 have the same interview date and similar GPS coordinates.

### 1.3 Final Thoughts

I think that this exercise was very giving because it introduced the different ways you can use OpenRefine for messy spreadsheets. Therefor I could also see myself using OpenRefine in some way for my final assignment.