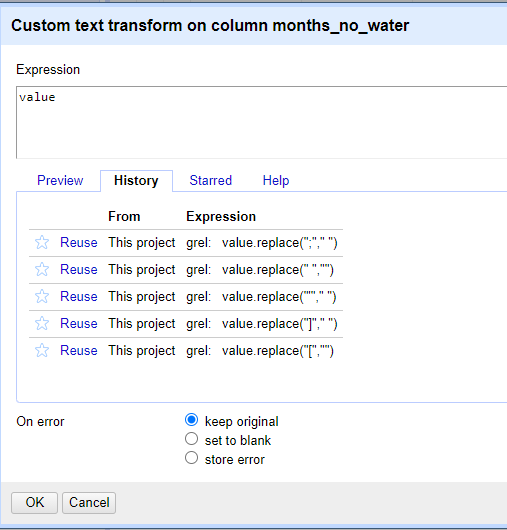
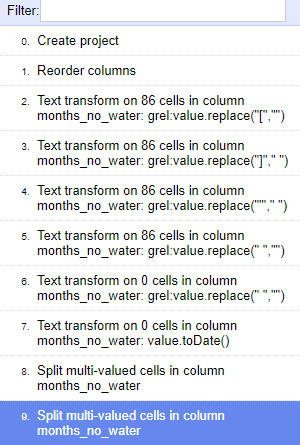
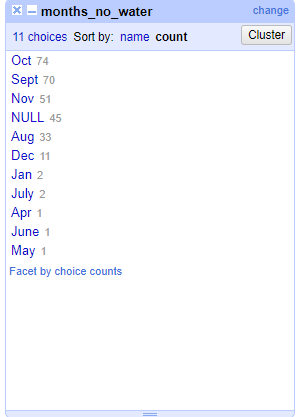
Week 45: Open refine

1. Put all your variables in columns - the thing you’re measuring, like ‘weight’ or ‘temperature’. Put each observation in its own row. Don’t combine multiple pieces of information in one cell. Sometimes it just seems like one thing but think if that’s the only way you’ll want to be able to use or sort that data. Leave the raw data raw - don’t change it! Export the cleaned data to a text-based format like CSV (comma-separated values) format. This ensures that anyone can use the data, and is required by most data repositories.
2. I assume that Openrefine doesn’t not alter the raw data when your sort or filter it. You are able to go back and forth in your steps.
3. I assumed that the question could be answer by looking at the column “months\_no\_water”. I used the following expressions to transform the data and make it readable.







1. To find the village that hides behind the number 49 I first sorted on gps\_Latitude as a number with the smallest first. I did the same with gps\_longitude. I then used to drop down arrow on village and selected edit column 🡪 move column to end. The allowed me to compare village names with GPS coordinates. I found the village 49 correspondent with the village of Chirodzo.