Overview of Database Structure

The TERA database is comprised of 35 tables that contain different types of records, and shorter list of views which combine important data from those tables into the format that will ultimately be used with Tableau. The following views are the ones that are ultimately exported into a CSV and imported into Tableau:

- historyFinalView
- moistureFinalView
- onchangeFinalView
- procFinalView
- temperatureFinalView

You can update and export these views per Darci and Scott's documentation, but you can now use the new Export Blueprint to automatically create your respective where clauses based on the tags selected in the Tag Library sheet. This means that the new workflow for importing additional tag is as follows:

1. Adjust the Tag Library selections as desired.
2. Use generated where clauses to update each view.
3. Export all affected views.
4. Merge csv files in Tableau.

Export Size Issue

One issue I ran into while trying to export additional data that had not been previously documented was the apparent ballooning of size seen in the exported data when comparing it to the packaged Tableau workbook we inherited.

In an attempt to figure out why this was occurring I took several approaches to analyzing the data, looking for discrepancies between my new export and the existing one that might clue me in to why my export is so much larger.

I created a table in a new Tableau worksheet that displays all the tags currently exported and which parent table they belong to. I then wrote a query that does the same thing but for the data that is ideally exported. I built out a series of tables in Google Sheets that compares these two snapshots and identifies where there is extra and missing data. Surprisingly, however, I

found that there was actually more extra data in the previous export than missing data, so I hit a dead end.

What I ended up discovering is that despite the shortcoming of Tableau's proprietary file types, files with their .hyper extension, which is used for data extracts, somehow only take up around 16.7% of the storage space of an equivalent CSV file.

Exporting Data

An ideal approach for exporting additional tags (i.e. categories of data values denoted by a combination of a particular measurement and chamber) would be simply writing a query for that specific data and merging it with the existing extract. However, the problem with this is that Tableau doesn't allow you to perform unions with .hyper files (the existing extract), unless you have access to Tableau Prep, which is a complimentary piece of software to Tableau which we do not have a license for. So instead you have to have CSV files for each of the five aforementioned views, union all of them in Tableau, and then finally extract the result into a .hyper file.

Single Column Data vs. Pivot

One of the primary factors that made working on this project difficult is the way the tables in the database are structured. The three primary fields in the TERA database are the Chamber, Value, and Date. However, rather than each distinct measurement recorded during the experiment occupying its own column, they are all stacked into the same column, known as 'Val', which depending on the row might contain data based on temperature, moisture, humidity, etc. In order to identify which values correspond to which measurement, an additional column is needed, called 'Measurement'. It's a categorical column which describes which type of measurement an adjacent value corresponds to.

Using these two columns in tandem, as opposed to a wider matrix approach, has significant disk space and performance drawbacks and proved to be a bit of headache during development. In hindsight it might have been worthwhile to pivot the data when we initially took on the project, but at the time we decided that the Tableau workbook we were using was too far along development for a restart to be a viable option.