Reflection 4

As alluded to in my previous reflection, I decided to spend the next fifty hours focusing on Google Refine. I believe that while I'm not by any means an expert, I've gotten a pretty good grasp of everything needed for my internship.

I learned this project hasn't been updated over the past ten years. All aspects of Google Refine were transferred over to a new program known as OpenRefine. Due to the relative niche of this program, many of the tutorials are somewhat outdated with the latest version being over thirteen years old. As such, most of my learning was done through these videos and tinkering with tools inside the program with fake data. To better simulate this, I decided to copy the structure of how the data in my internship was organized, by using generic information. I then modified that over and over to see different results. The handiness of this program allowed me to go back to previous versions of my data with new values and filters without much hassle.

As a result of this, I learned a few varied methods of data structure and sorting. While most of these structures wouldn't fit my internship model, I believe it was still an informative experience that I could potentially use in the future. The overall nature of OpenRefine also allowed for easier sorting and organization methods than Excel and Google Sheets. The only drawback was this program requires a download and a browser used as a server to actively work with the program. While I do not believe this had any performance issues, it may throw a few people off. OpenRefine also has an issue when sorting data from multiple sheets of data. I fixed this problem by having separate projects with each different sheet of data, which were then easily integrated into one document. Overall, I would recommend OpenRefine as a potential solution to any data sorting woes.