Reflections

The method of data exploration that I chose to pursue was that of fixing known issues right away (I knew I would need to remove the units from AvgTemp and MeanAnnualPrecip). Then I created rough figures for each variable and identified issues that needed to be resolved. These included misspellings of data values (secimentary instead of sedimentary, or interntaional instead of international).

I then needed to figure out what to do with the variables that had multiple values for some observations. Based on some internet searching, the best course of action I found was to use the tidyr function ‘unnest’ to split the multivalues at the commas that separated them. What this did was create a new observation for each value in the column. Therefore, I needed to create separate dataframes for each multiple value variable I fixed, or else the main czenSites dataframe would have redundant observations. This process was definitely challenging, but rewarding once I figured it out.

Once I had completed these preliminary data cleanings, I returned to the individual figures to identify more subtle issues. Once I had identified those issues, I made the changes back in the data cleaning section.

I am generally happy with the way my figures turned out. I included multiple covariation figures to illustrate noisy data vs the cleaner Latitude x AvgTemp. I wish I had more knowledge to use color to highlight certain aspects of each figure. Additionally, I tried to figure out how to re-order the Stream Order figure (First -> Sixth instead of Alphabetical) but was unable to figure that out.

Overall, I was not surprised by the areas of study that were lacking in representation. Most existing Field sites are in the US and cover a broad range of systems, but there are still many unique systems across the Earth that do not have long-term research organizations established.