Data Journal — SC NID Revisions:

This dataset appears to be fairly clean at a first glance. Very few variances in spelling, formatting, limited amount of “unknown, none, NA” in cells.

1. First, I formatted the dam names for consistency in spelling in a new column, “formatted\_dam\_names,” by scrolling through the filter and identifying discrepancies. Once discrepancies were found, used the find & replace tool to adjust to the correct name (after independent research)
   * + I did not worry about adjusting case levels with this column, since each dam would be unique and still searchable regardless of case match
2. Then, formatted County column in a new column titled “formatted\_County”:
   * + Started by correcting any spelling errors/typos
     + Some dams had two counties listed.
       - Using text to columns, sectioned out those with two counties listed. The second counties were placed in a column titled “formatted\_County2” using text to columns with a delimiter of a backslash, then with a delimiter of a comma.
     + Using find and replace, standardized capitalization of county names (since uniformity is preferred and this was a data load easy to manually adjust using find & replace)
3. Largest work was with formatting the inspection dates:
   * I began by formatting the column to MM/DD/YYYY
   * Using text to columns with the backslash as a delimiter, separated out month, day, and year into separate columns.
     + This allows for better search and graphing capabilities, as some only had month and year.
     + With the year separated out, that also allowed me to scan through for any errors, including removing specific times of inspection that had been included in the cells. This was done manually after scanning the filter list.
4. Finally, using text to columns, I broke out the “other names” column for the dams to account for those with multiple values in their cells. Cells initially had up to three values, so these were broken into “formatted\_other\_name1,” formatted\_other\_name2” and so forth.
   * + In this field, there were no “unknowns” or “nones” but there was the value “0” in several of the columns.
       - Using the filter table, I created an empty cell that previously read as a “0”