Steps:

1. Create a Google Doc (1-2 pages) describing the data wrangling steps you took to clean the dataset. Include answers to these questions in your submission:
    1. What kind of cleaning steps did you perform?
    2. How did you deal with missing values, if any?
    3. Were there outliers, and how did you handle them?
2. Submit a link to the document.
3. Discuss it with your mentor at the next call.
4. Revise and resubmit if needed.
5. Convert the final document to a .pdf and add it to your GitHub repository for this project. This document will eventually become part of your milestone report.

Procedures:

alzheimer_by_age.csv

- I took the original dataset, I sorted by LocationID [ID of US states] which I placed as the first column in the dataset and I also generated a list of columns. Based on this list of columns, I dropped the ones which do not give me any useful information or they give duplicate information from other columns. The goal here is to group the dataset by an informative column and remove any columns that doesn't make sense - effectively condensing the data set.

alzheimer_by_age_2.csv

- After I performed the above steps I then sorted the dataset by Question [questions asked] as that is a far more informative column than LocationID. This file is then saved to alzheimer_by_age_2.csv

alzheimer_by_age_3.csv

- Here once again, I dropped 3 columns that did not generate any useful information [that I missed the first time around].
- With the columns sorted out, I first replaced all the empty cells with N/A, and then I dropped the N/A cells in the columns that had them. Turns out there were only 3 columns that had empty cells in them at this point: Data_Value, Low_Confidence_Limit and High_Confidence_Limit.
- Lastly I found what unique questions were asked to better get an idea of what filter out for questions. This will be done on a Word document (in alzheimer_by_age_3.csv, it will still have all of the questions, just that not all questions will be analyzed).