**Data Cleaning and Manipulation**

Due to the smaller size of the dataset, and for ease of visualization Microsoft Excel was used.

1. Check for duplicates. There were no duplicates
2. Age column had 8 outliers lesser than 18 years and greater than 80. Some values were negative or 3 or more digits. These 8 values were replaced by a median of 31 years.
3. All gender inputs were sorted into Male, Female, or Other for simplification.
4. Using filtering, if Self Employed is entered as "NA", using find and replace, it was replaced with "No".
5. Using filtering, if Treatment is entered as "Yes", and work\_interfere is NA, work\_interfere was changed to "Never"
6. New column created “diagnosed” based on column “work\_interfere”. If work interfere had entries of “Yes” or “No”, diagnosed column is “Yes”. For work interfere “NA”, diagnosed column is “No”.
7. Using filtering, I filtered out all countries that are not the United States and deleted them from the dataset as we are focusing on U.S. for our presentation.
8. Changed state abbreviations to full state names for ease of visualisation
9. Deleted Unnecessary “Timestamp” column.
10. In “no\_employees” column, replaced “-“ with “to”.