- Change histograms to use merged data frame ('df')
- ANALYZE (Data section of write up)
    - Mention the scale of graphs in the write up → 1 is the least happy, not the happiest
    - Merging the dataframes, combining for the kernel to show
    - Mention the different categories for 'null' responses, explain that for our research question, the difference was unimportant

- RESULTS
    - Paste the graphs
    - Because the kernel density plot is not including nulls, the fact that a different number of people responded in 2018 and 2022 is not a factor→ the graph shows the percentage out of the actual responses
    - Histograms → the number of responses for 1, 2, and 3 are shown in the bins

ADD: MERGED DATAFRAME HISTOGRAM
EXPLAIN: why histogram was less effective at communicating actual data than kernel density plot

Charlie: Results Section
Tori: Summary
Tu-Yen: Data
- Two to five pages providing visualizations, statistics, and a discussion of your findings. If you have a lot of plots or tables, that's OK, but try to focus on a few key pieces of evidence rather than doing every single pairwise comparison of some set of variables
The data used in this report was derived from the General Social Survey, or GSS, which

conducts surveys almost annually on social, economic, and political values of people in the

United States. To achieve our analysis, two datasets were obtained from this survey: the data

from 2018 (before the pandemic) and the data from 2022 (after the pandemic). Each dataset was

read into its respective dataframes to be cleaned and visualized in the code: df_2018, and

df_2022.

After reading the data in, it needed to be cleaned. Due to the number of survey questions

the GSS has, we updated these data frames to only include the variable needed for analysis:

'HAPPY'. This variable looked at people's answers to the question: "Taken all together, how

would you say things are these days--would you say that you are very happy, pretty happy, or not

too happy?". The data is given in numeric values of 1, 2, or 3 based on their answers, where 1 represents very happy, 2 represents pretty happy, and 3 represents not too happy. After the necessary data was extracted, the columns were renamed based on the year; 'HAPPY' in df_2018 became 'HAPPY_2018', and 'HAPPY' in df_2022 became 'HAPPY_2022'. This was done for better clarification and because later on, both datasets will need to be merged and it's important to differentiate between each column to avoid confusion. Then, the values were ensured to be numeric using the pd.to_numeric() function. Finally, null and outlier values were cleaned. The data had no outliers when checked (every value was either 1, 2, or 3), so nothing had to be handled in that case. Regarding null values, null counter variables were created to see the sum of null values in each data set (null_2018, null_2022); the 2018 data only had 4, and the 2022 data only had 24. One of the challenges faced was determining what to do with these null values, but this was solved after looking at the GSS documentation. According to the code book, the null values are caused by three different occurrences: don't know, no answer, or skipped on the web survey. Other than that, there was nothing to go off of to determine the reasoning for each of the null values. So, it was decided that the null values would not provide us much information, nor would it skew the results drastically, so we left them as is.

The final step to preparing our data for analysis was merging the two data frames into one. As of now, the two data frames have remained separate (df_2018 and df_2022). It was important to combine the two in order to maintain clean, consistent code, and allow for easier data manipulation. In the end, we had a data frame where the rows were each individual observation/person surveyed, and the variables (columns) were the HAPPY values mentioned before, and the year that survey was conducted (2018 or 2022). The final data frame created was "df_merged".

MG: Conclusion
- For future research section at the conclusion, could use this point to expand on (suggesting additional work)