Summary

This study examines public opinion on the acceptability of suicide under different circumstances over the years. The data is extracted from the General Social Survey, spanning across several decades. For our examination, we focus on four distinct reasons of suicide: incurable disease (suicide1), bankruptcy (suicide2), family dishonor (suicide3), and weariness of life (suicide4). The research question we have chosen is “how have attitudes towards suicide changed over the years?”. We found this important question to research since mental health is a critical concern among college students and adults. Attaining this insight might aid with improving mental health recovery. The analysis employs data cleaning, trend analysis, and statistical visualization techniques to parse out the social attitudes over the decades. The findings indicate that there is a perceived difference in public opinion for different justifications. Health related reasons were more accepted, while those based on financial or personal dishonor were less so.

Data

The dataset for this analysis is acquired from the General Social Survey (GSS), a sociological survey collected by the National Opinion Research Center. The GSS collects data to explain trends and constants in attitudes and behavior in American Society. The specific dataset for this analysis looks at the justifiability of suicide under various circumstances over the last few decades.

For this particular analysis, key variables were selected based on their relevance to the study: How have attitudes towards suicide changed over the years? For this, we analyzed four key variables to understand if society deems it justifiable to commit suicide if they suffer from an incurable disease (suicide1), if they have gone bankrupt (suicide2), if they have dishonored their family (suicide3), or if they are tired if living (suicide4). These variables are specifically chosen for their potential to gain insight of the societal attitudes towards suicide.

One of the main challenges working with this dataset was managing the size of it. By working with certain smaller data chunks and loading on at a time into memory, this issue was mitigated. When doing this, we had to ensure that we carefully merged the chunks together to ensure consistency and accuracy across the dataset.

Missing data led to challenges of its own, mainly due to missing values. The dataset comprised a significant amount of missing data, which was labeled ‘NaN’. The missing data consist of different labels for non-responses such as “Don’t Know”, “No Answer”, and “Not Applicable”.

The good news is that the GSS dataset already standardize these non-responses to “NaN " for clarity and consistency.

As mentioned above, there is a about more than 50% missing values in total of four suicide variables. In addition, the variables we chose are relatively mode sensitive. Shown in the GSS 2022codebook, 'SUICIDE1' requires further investigation, 'SUICIDE4' is consider as less likely to be mode sensitive, while the other 2 suicide variables have not been accessed by the GSS team at that time. Furthermore, there is no default value for these variables. And the only value for them are 'YES'(1) or 'NO'(0), so we can't impute the missing values with mean, median nor mode. According to the above information, we decide to clean all the NaNs.

Once the key variables were isolated, further data cleaning was necessary. The year variable needs to be converted from a categorical to a numerical format to maintain consistency and facilitate time trend analysis. The suicide related variables, which were initially categorical, were mapped to numerical values, 0 and 1, to simplify the calculation of response frequencies and percentage. This conversion was necessary for analysis as it focused on quantifying the public opinion regarding the justifiability of suicide.

Results

This analysis examines the trends and patterns in the attitudes towards suicide over the period spanning from 1977 to 2022 in America. The examination is based on the responses to four different survey questions labeled 'suicide1', 'suicide2', 'suicide3', and 'suicide4'. This data was converted from a categorical response (‘yes’, ’no’) into a numerical response (1, 0), facilitating statistical examination and visualization.

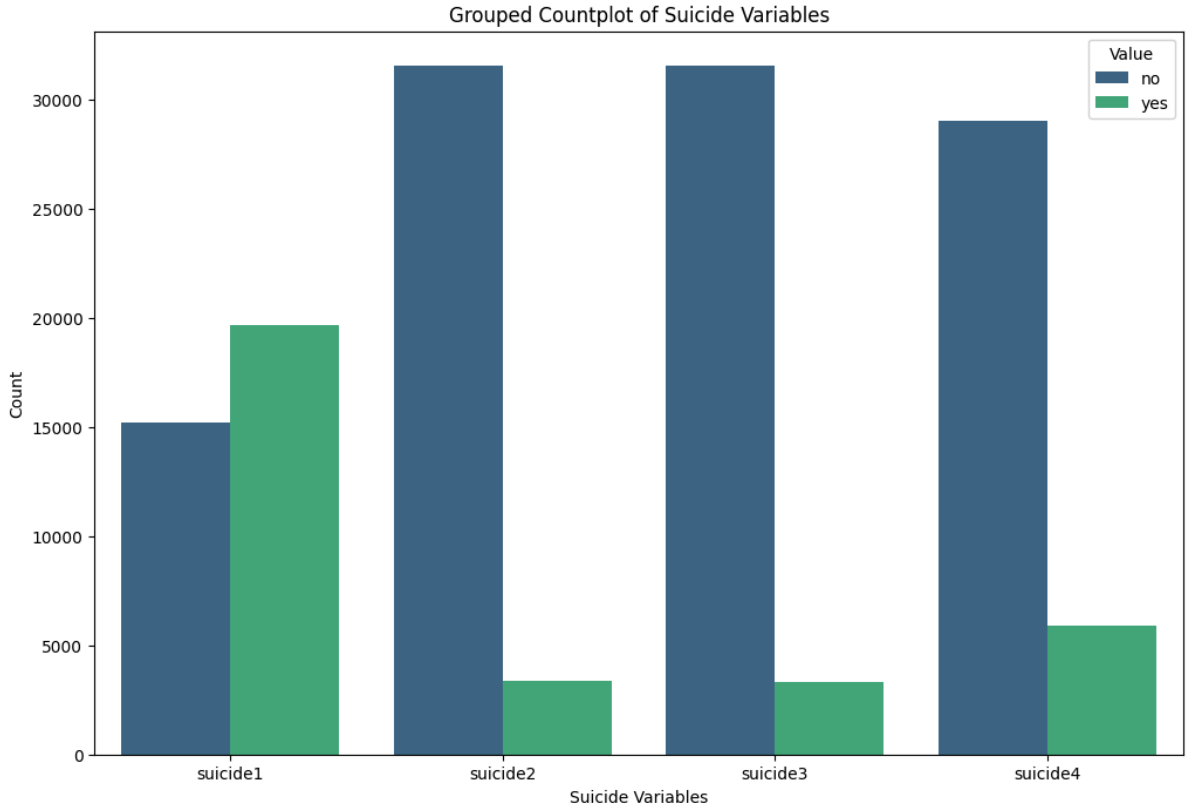

Figure 1

Figure 1 shows the distribution of “yes” and “no" responses for each suicide variable. A distinct pattern can be observed where “suicide1” consistently accumulate a higher “yes” count (around 20,000) compared to theother suicide variables (lower or approximately 5,000). Additionally, it can be observed that for every other variable outside of “suicide1”, there seems to be a higher

“no” count. This suggest that suicide based on health issues is more acceptable and deemed more acceptable than other reasons stated.

	year	suicide1	suicide2	suicide3	suicide4
count	34891.0	34891.000000	34891.000000	34891.000000	34891.000000
mean	1998.449371	0.563756	0.095927	0.095841	0.168553
std	13.332919	0.495926	0.294496	0.294378	0.374362
min	1977.0	0.000000	0.000000	0.000000	0.000000
25%	1986.0	0.000000	0.000000	0.000000	0.000000
50%	1998.0	1.000000	0.000000	0.000000	0.000000
75%	2010.0	1.000000	0.000000	0.000000	0.000000
max	2022.0	1.000000	1.000000	1.000000	1.000000

Table 1

Table 1 includes the statistic for all the variables we have. Since the suicide variables and ‘year’ are categorical variables, so there is no outliers here. The summary statistics for 'year' is not usefull. The only information we need to know is our dataset contain the data from 1977 to 2022.

For suicide variables, their counts exactly the same since we cleaned all the NaNs. The mean of variable 'suicide1' is strongly greater than the other three, nearly 50% (0.5) more. And it is the only one with the mean value above 50% (0.5), means more than half of the respondents think a person has the right to end his or her own life if this person if he/she has an incurable disease. This matches the result shown in the Figure 1. Min, 25%, 50%, 75% and max statistic are also meaningless for suicide variables. Again, becuase they are categorical varibales, the only value they have are 1 for yes, and 0 for no.

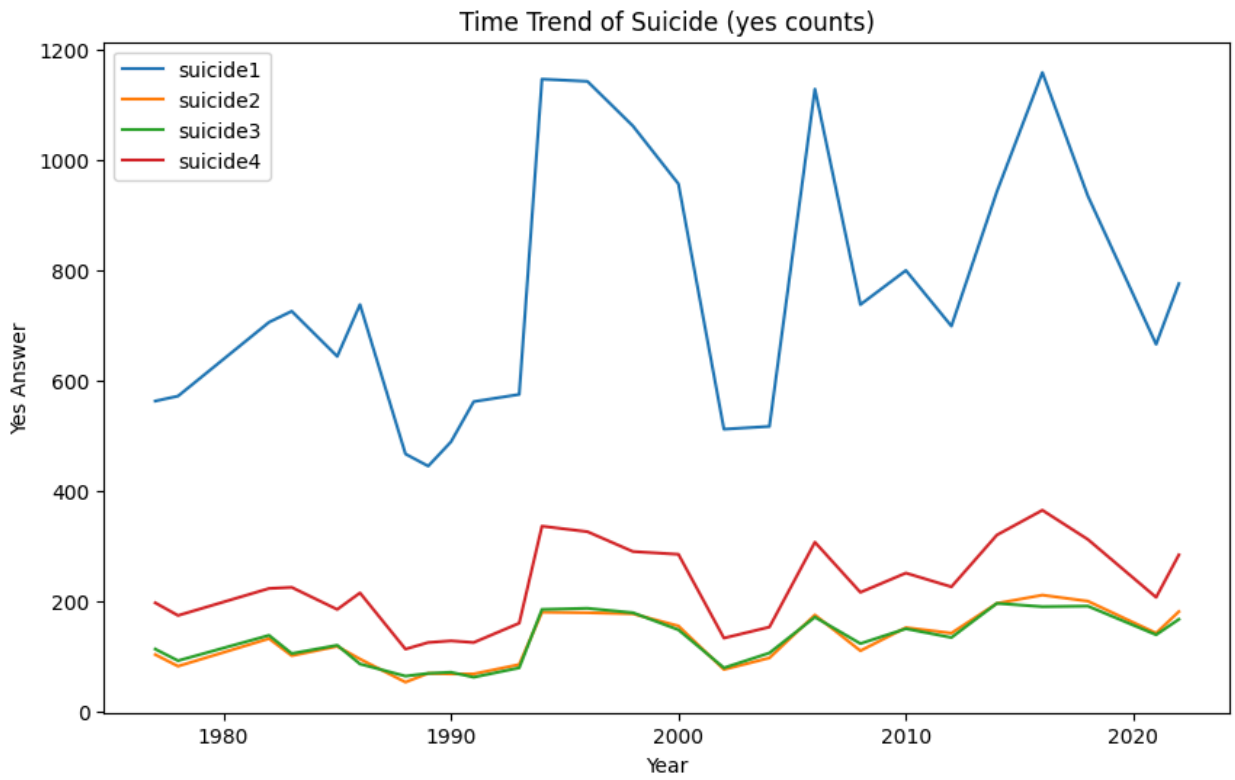

Figure 2

Figure 2 presents a trend of "yes" counts across all four variables over the years. It is worth noting the ascending trend of "suicide1" response, indicating a shift towards a more affirmative stance. In contrast, the other variables, "suicide2", "suicide3", and "suicide4", demonstrate relative stability in "yes" counts over the years.

In the Figure 3, the percentage of "yes" responses for each year was calculated, revealing an ascending trend for "suicide1", with the percentage of "yes" responses surpassing 70% in most recent years. The other variables maintained a lower and constant "yes" percentage, barely surpassing the 25% mark.

The significant increase in "yes" response for "suicide1" post 2010 could warrant further investigation to understand the underlying causes. It may be beneficial to look at policy

changes, social movements, or public health interventions that could have influenced public opinion.

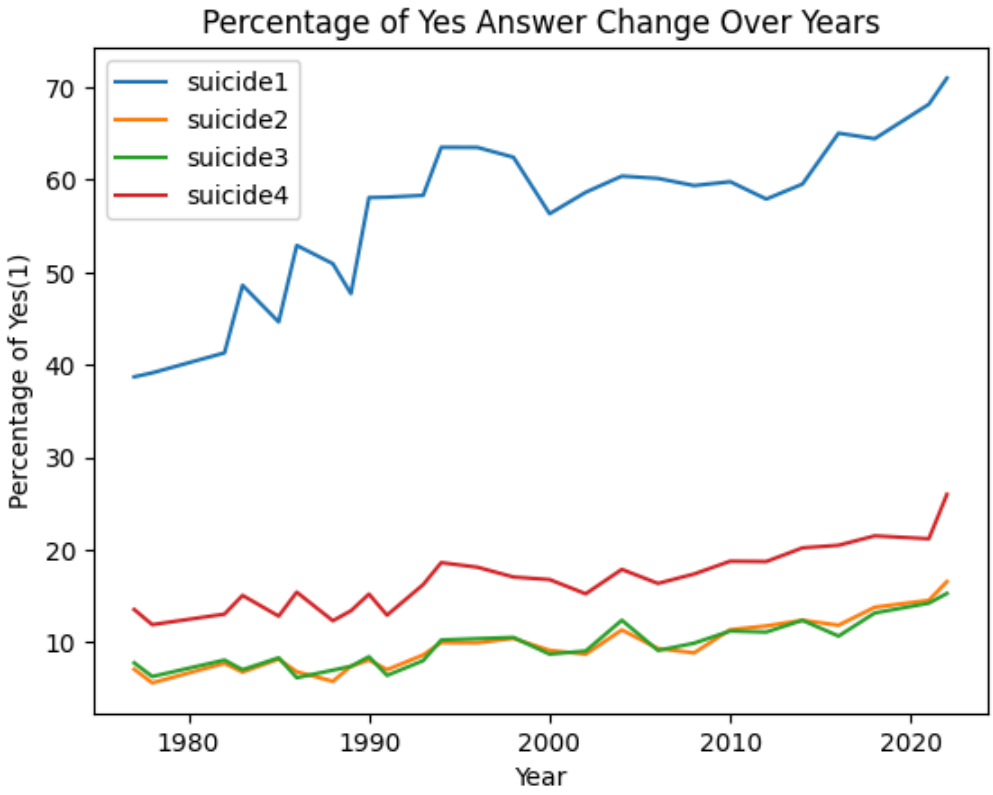

Figure 3

Conclusion

This project was an extensive analysis of the societal attides towards suicide and justifications in various scenarios related to suicide over the 45 year span. Some criticism may arise in our analysis regarding the binary nature of a “yes” or “no” response towards suicide. On one hand, this simplifies the complex issue of suicide, but this binary approach allows for clear, digestible insight into trends over a certain period of time, servering as a valuable starting point for understanding societal change. This survey encompasses a large demographic, which stands firm against any critiques of sample size and representativeness. While the cope of the project focuses on a numerical analysis, more areas of study is recommended to expand understanding of Americans attitude towards suicide. First can be understanding how and if public policy influences public opinion of these sensitive issues. This could be looking at healthcare practices or medial portrayal of suicide. Another way to get a better understanding of the issue is doing a qualitative analysis. This could be an in depth interview or focus groups that could provide better context behind these numerical trends observes, especially regarding the increase in affirmative responses to “suicide1”. Lastly, in the digital age, it is pivotal to understand the corelation between social media and the attitudes towards suicide. A dedicated study of online discourse and shifts in survey responses could be enlightening.