

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 of 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'. Additionally, missing data had different labels for non-responses such as "Don't Know", "No Answer", and "Not Applicable". Additional steps needed to be performed to standardize these non-responses to "NaN" for clarity and consistency.

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. This conversion was necessary for analysis as it focused on quantifying the public opinion regarding the justifiability of suicide.