

General Happiness in the United States Before and After the COVID-19 Pandemic: an Analysis  
of the 2018 and 2022 General Social Survey Responses

Contributors: Tu-Yen Dang, MaryGrace Gozzi, Victoria DaRosa, Carlos Perez

## I. SUMMARY

The purpose of this paper was to analyze the data provided by the General Social Survey that gauged a respondents' level of happiness. The survey asked participants "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?". Respondents then rated their happiness on a three-point scale, where 1 represents "Very Happy", 2 represents "Pretty Happy", and 3 represents "Not Too Happy". Specifically, this paper examined responses to the question in two distinct time periods: 2018, preceding the COVID-19 pandemic, and 2022, following the pandemic. In order to analyze the difference in answers, data was extracted from two separate datasets, each originating from the General Social Survey from 2018 and 2022. From these two data sets, only the variable "HAPPY" was considered for analysis. This variable contained a respondent's answer (1 through 3) or null if the question wasn't answered or was answered with "Don't Know" for various reasons. The collected data was put into combined data sets to create multiple graphs. These visuals included histograms and kernel density plots for each separate year, as well as for the aggregate data spanning both years.

From the graphs that were created, it was found that the survey from 2022 had more respondents. Additionally, on average there were less “Very Happy” responses and more “Somewhat Happy” and “Not Happy” responses in the 2022 survey. In 2018, out of 2,344 analyzed responses, 29.91% answered “Very Happy”, 55.76% answered “Pretty Happy”, and 14.33% answered “Not Too Happy”. In 2022, out of 3,520 analyzed responses, 22.13% answered “Very Happy”, 55.17% answered “Pretty Happy”, and 22.70% answered “Not Too Happy”. While these findings suggest a significant shift in self-reported happiness levels, it's important to recognize that various external and internal factors may contribute to these changes. Other socio-economic, political, and personal circumstances not captured in this analysis can also play a role in how individuals rate their happiness, but it is important to acknowledge COVID-19's effect on people. COVID-19 has had a substantial impact on people's lives, which likely influenced the decline in self-reported happiness levels.

## II. DATA

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.

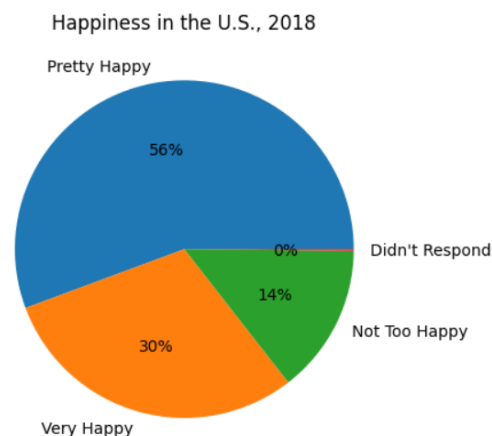
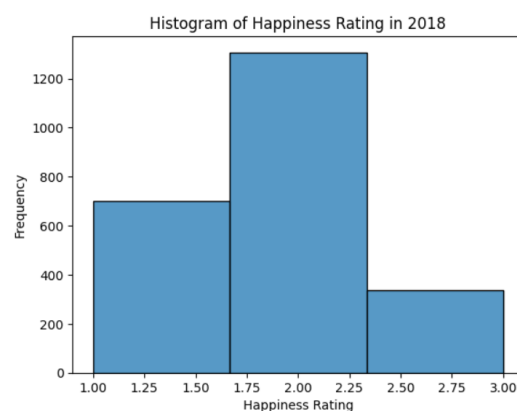
The final step to preparing our data for analysis was merging the two data frames into

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”.

### III. RESULTS AND ANALYSIS

#### 2018 Happiness

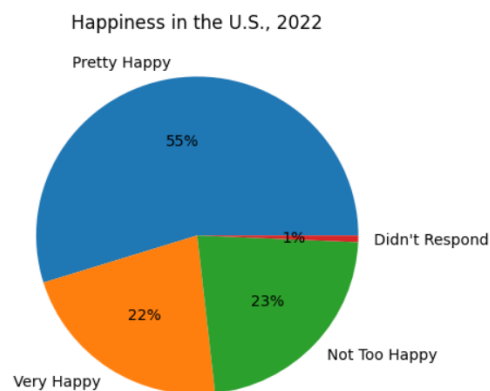
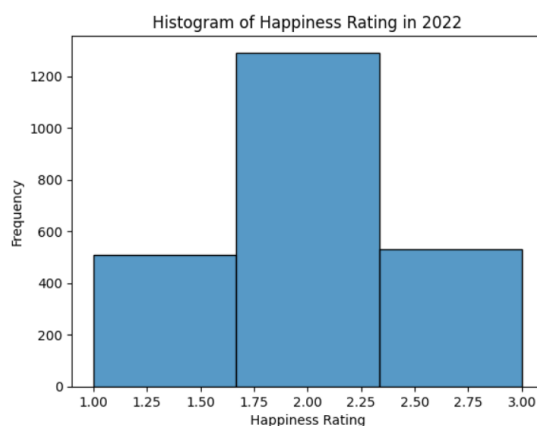
In 2018, two years before the COVID-19 Pandemic, there were 2,348 respondents to the GSS, of whom all 2,348 respondents answered the happiness question. Of the 2,344 within the scope of our analysis, 701 respondents (29.91%) answered 1: “Very Happy”, 1307 (55.76%) answered 2: “Pretty Happy”, and 336 (14.33%) answered 3: “Not Too Happy”. The other four (0.17% of respondents) answered “Don’t Know” and have been omitted from our analysis. No respondents chose not to answer. The numerical and percentage distributions are shown in the histogram and pie chart displayed to the right. The mean respondent to the survey, with a 1.844 mean happiness score, was slightly more than



“Pretty Happy”. Furthermore, the ratio of respondents who said they were “Very Happy” versus “Not Too Happy” was 2.086 - over twice as many people.

### 2022 Happiness

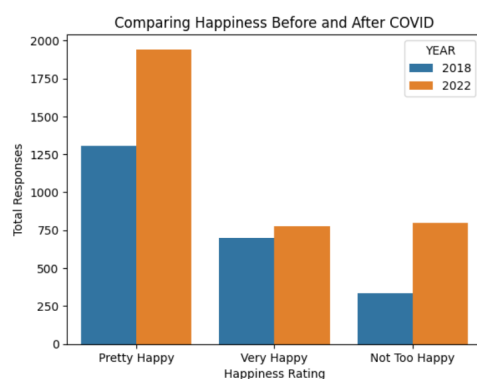
In 2022, two years after the beginning of the COVID-19 pandemic, there were 3,544 respondents to the GSS, including online respondents, introduced between 2018 and 2022. In addition to being able to select “no answer” or “Don’t Know”, online respondents also had the ability to skip questions. 13 respondents either chose not to answer or skipped the question, and 11 answered “Don’t Know”. While it is interesting that there was a proportional increase in people who did not answer or were unsure, this is a very small proportion of the overall dataset and not to be considered statistically robust, and these 24 values (0.68%) have been omitted from our analysis. Of the 3,520 responses within the scope of our analysis, 779 (22.13%) answered “Very Happy”, 1,942 (55.17%) answered “Pretty Happy”, and 799 (22.70%) answered “Not Too Happy”. See the charts to the right for numerical (above) and percentage (below) distributions. The mean respondent to the survey scored just barely over a 2, at 2.006,



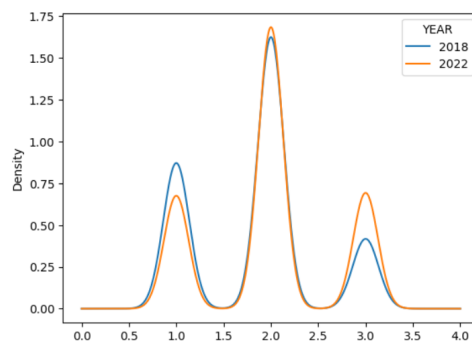
making the mean respondent marginally less than “Pretty Happy”. The ratio of respondents who characterized themselves as “Very Happy” versus “Not Too Happy” in 2022 was 0.975 - though we must keep in mind that the mean here is “Pretty Happy”, which is not necessarily a happiness-neutral term, it is still quite impactful that more respondents in 2022 considered themselves less than “Pretty Happy”.

### Comparison: Happiness, 2018 to 2022

We created a pair of visualizations to better understand the happiness change over these four years: a combined bar graph and an overlaid kernel density plot. They can be seen to the right. We found the kernel density plot to be a far more useful visualization because while it is imperfect for evaluation of such a strongly categorical variable, it effectively controls for the difference in the sizes of the 2018 and 2022 datasets, with the 2022 dataset being 50.2% larger than the 2018 dataset.



The kernel density plot reveals much of what has already been discussed and visualized in our evaluations of the individual datasets. Though the percentage of people who characterized themselves as “Pretty Happy” remained largely unchanged, with a delta of less than one percent of total respondents, the two extremes of this survey experienced a much wider fluctuation, with a clear increase in



respondents who were “Not Too Happy” and corresponding decrease in respondents who were “Very Happy”. The proportion of respondents who were “Not Too Happy” increased by a factor of 1.584, while the proportion of respondents who were “Very Happy” decreased by a factor of 0.740. Evaluating the change over time in mean happiness score, there was an increase - keeping in mind that the higher the number, the less happy the respondent is - of 0.161 in the mean happiness score between 2018 and 2022.

From these statistics and visualizations, it is clear that there was a relatively dramatic downturn in perceived happiness between the four years in between 2018 (the last GSS survey pre-COVID) and 2022, two years after the beginning of the pandemic and a year in which the majority of the population had already been vaccinated, a year in which life had largely returned to the pre-COVID status quo. Though it requires further analysis of potential confounding variables, this evidence does support our hypothesis that the COVID-19 pandemic had a predominantly, though not exclusively, negative impact on Americans’ overall well-being and perceived happiness.

The major drawback to this analysis is the over 50% increase in respondents between 2018 and 2022. We had hoped that the continuity of respondents - the same households year after year - might help to insulate our analysis from sampling problems and provide an effective microcosm of the American experience of the COVID-19 pandemic, but unfortunately, a large number of additional respondents entered the dataset between 2018 and 2022. We would argue, however, that this drastic shift in the extreme points of analysis, between “Very Happy” and “Not Too Happy” corresponds strongly with what we know of the COVID-19 pandemic: though

it did not touch every household, the major lifestyle changes, including social isolation, necessitated by the government's response to the pandemic, as well as the negative impact upon those families who lost family members or their own lives to the COVID-19 virus or who developed extreme complications, significantly affected the happiness of certain elements of the population.

#### IV. CONCLUSION

This project was designed to compare results 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?'. This question, represented with the variable 'HAPPY' was included in both the 2018 and 2022 General Social Survey, which allowed for a fair comparison of results across the two years. Respondents could choose between '1', '2', and '3', correlating to 'very happy', 'pretty happy', or 'not too happy', in that order. We were interested in this variable for a particular reason: the Covid-19 pandemic. Covid-19, which began to spread on a global scale in 2020, was harmful for too many reasons to count. People all over the world got sick, lost family members, lost their jobs, and had no choice but to change their daily routines to protect themselves and their communities from the virus. For this reason, we predicted that respondents' overall happiness would have decreased between the years 2018 and 2022.

It is important to note that the total number of responses in 2018 and 2022 were not the same, so the following data represents a comparison of the percentages of people who answered '1', '2', or '3' for each individual year. A straightforward response count comparison



would have been inaccurate, because the total number of responses were not the same. After cleaning and modeling these data, we saw that our prediction was largely supported. More respondents in 2018 selected ‘very happy’ in 2018 than in 2022. Additionally, fewer respondents selected ‘not too happy’ in 2018 than in 2022.

While it is possible that these results were not a direct result of the Covid-19 pandemic, we would argue that many factors that would decrease a person’s general happiness were negatively impacted by the virus. For this reason, we think this comparison is worthwhile in helping understand the lasting impact of the pandemic.

This project could be expanded with the inclusion of different variables relating to the respondents. We think that ‘age’ would be a particularly interesting consideration, as it would allow us to compare how people of different ages and generations handled the pandemic in terms of their own happiness. Are young adults feeling less happy in 2022 than elderly people because of the ‘loss’ of crucial developmental time during lockdown? This question and other similar comparisons would shine light on more specific Covid-19 impacts, rather than just highlighting a nation’s general mood before and after.