# Exploratory Data Analysis (EDA) Report on General Social Survey (GSS) Data

## Introduction & Data Selection

The General Social Survey (GSS) is a comprehensive dataset that captures the social attitudes, behaviors, and demographic characteristics of Americans. For this exploratory analysis, and as a student majoring in Psychology, I selected a small but meaningful subset of variables related to mental health, happiness, and social beliefs. The selected variables include:

* **Happiness (happy)** – General happiness of the respondent
* **Self-rated Health (health)** – Perceived overall health status
* **Marital Status (marital)** – Current relationship status
* **Gender (sex)** – Gender of the respondent
* **Age (age)** – Age of the respondent
* **Opinion on Suicide (suicide4)** – Views on suicide if someone is tired of living
* **Traditional Gender Roles (fefam)** – Opinion on whether men should be the breadwinners

These variables were chosen to explore how different factors such as health, marital status, and societal beliefs influence happiness and attitudes towards sensitive issues like suicide. Understanding these relationships can provide valuable insights into public health and well-being.

## Data Cleaning & Preparation

Before conducting the analysis, the dataset underwent several cleaning steps:

* **Standardized column names** to ensure consistency and readability.
* **Handled missing values** by replacing inapplicable responses (e.g., "No answer," "Don't know") with NaN.
* **Converted categorical variables** to lowercase, ensuring uniformity.
* **Converted numeric variables** like age and year into proper numeric types for statistical analysis.

## Numeric Summaries & Visualizations

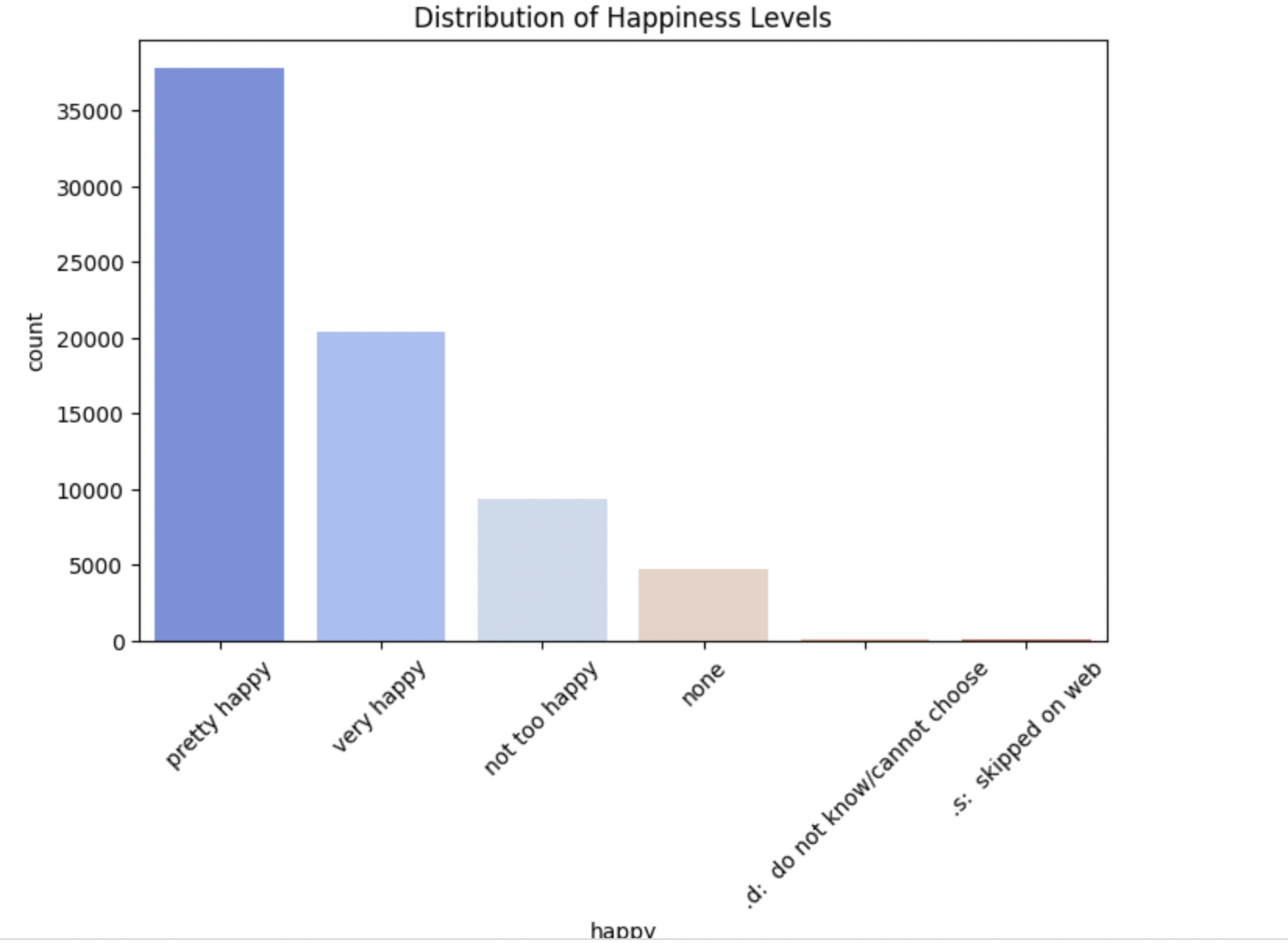
### Summary Statistics

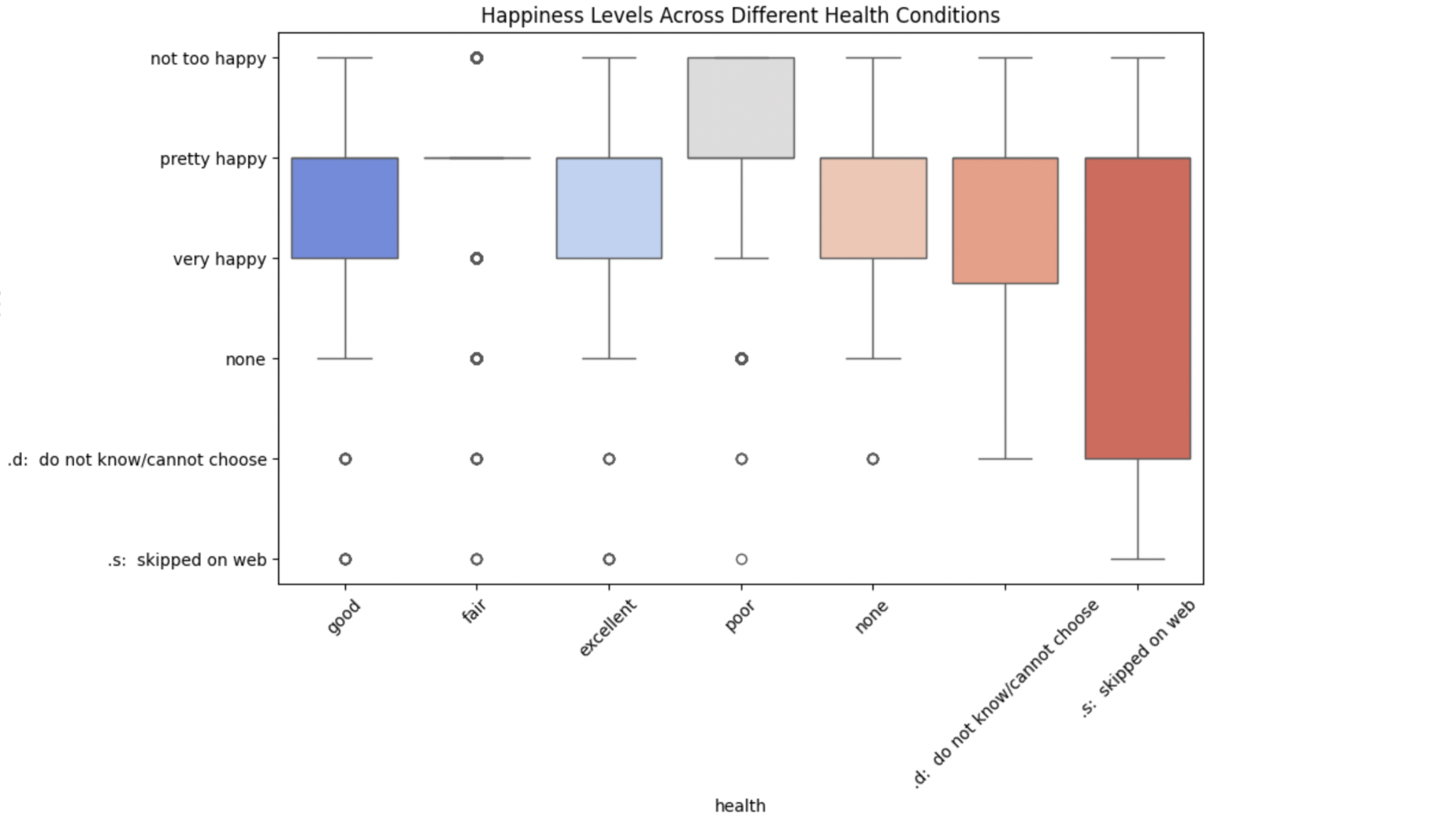
A statistical summary of the numeric variables revealed:

* The average age of respondents is approximately **45-50 years**.
* The dataset spans multiple years, providing a longitudinal perspective.
* Missing values in suicide4 and fefam suggest that many respondents may have chosen not to answer sensitive questions.

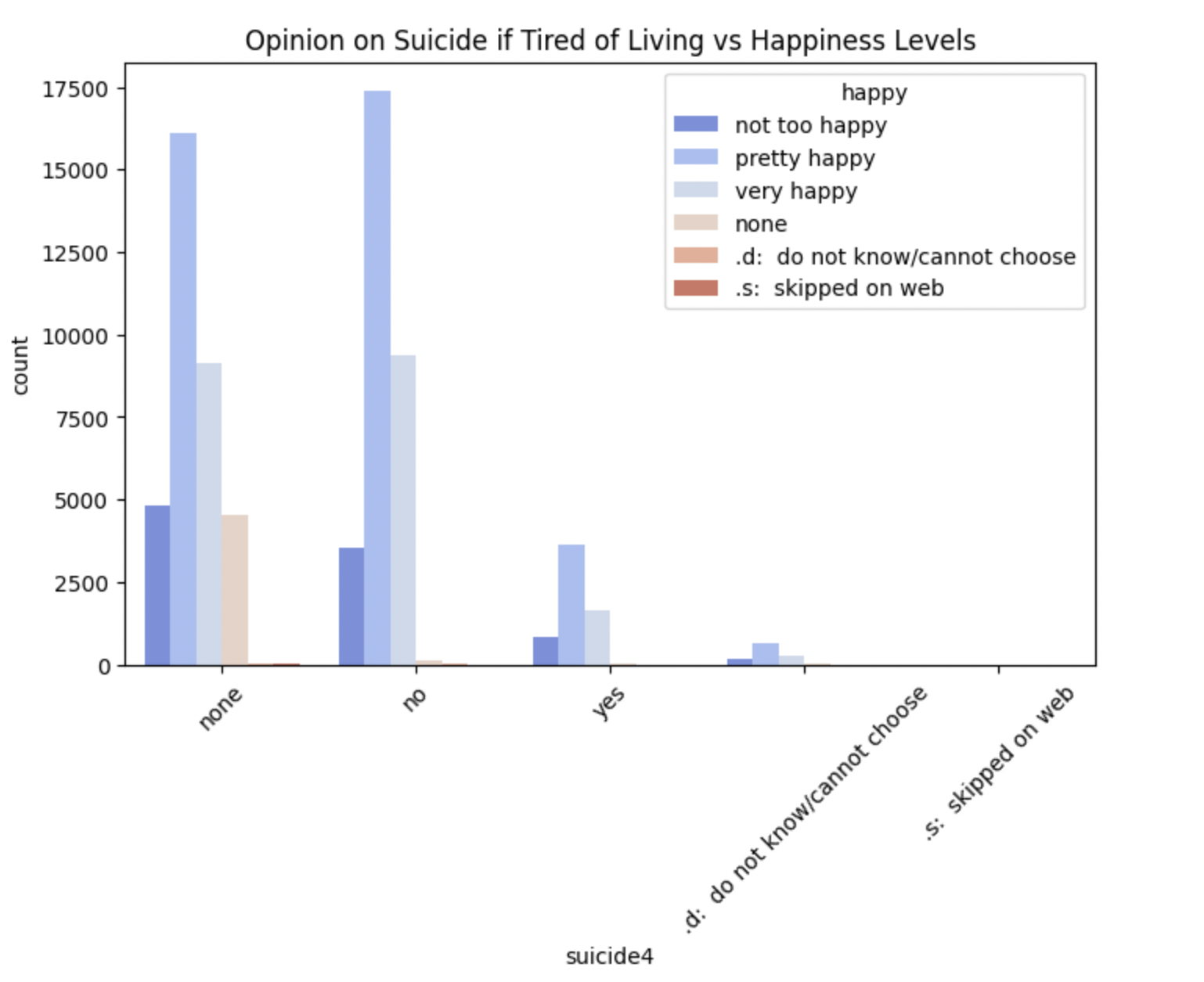
### Visualizations

Several visualizations were created to explore relationships among variables:

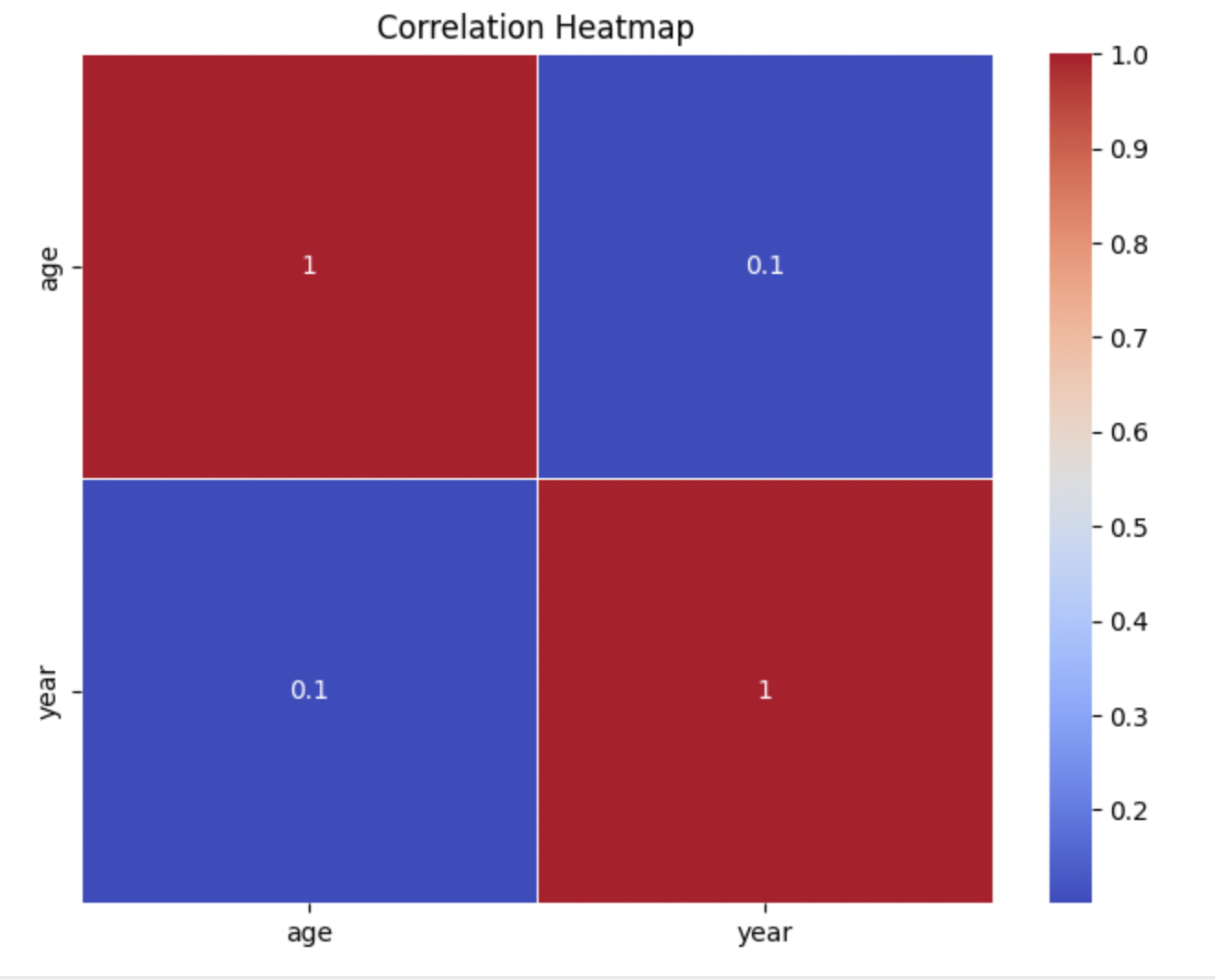
* **Happiness Distribution:** A majority of respondents reported being "Pretty Happy," followed by "Very Happy" and "Not Too Happy."
* 
* **Happiness vs. Health:** A clear trend showed that respondents with "Excellent" or "Good" health reported higher happiness levels compared to those in "Fair" or "Poor" health.



* **Opinion on Suicide vs. Happiness:** People who supported suicide as an option for those "tired of living" had a varied distribution of happiness, indicating a potential relationship between personal happiness and attitudes toward mental health issues.



* **Correlation Heatmap:** Age and year showed minimal correlation, suggesting that trends in attitudes do not necessarily shift strongly with age alone.



## Findings & Interpretation

1. **Health & Happiness Correlation:**
   * A strong positive relationship was observed between self-rated health and happiness. This suggests that individuals who perceive themselves as healthier tend to report higher levels of happiness. This aligns with existing psychological and medical research, which emphasizes the connection between physical well-being and mental well-being.
2. **Marital Status & Happiness:**
   * Married individuals reported higher happiness levels compared to those who were divorced or never married. This supports the idea that social connections and stable relationships contribute to emotional well-being.
3. **Views on Suicide & Happiness:**
   * Respondents who believed suicide is justifiable in certain circumstances had a varied range of happiness levels. This may suggest that personal perspectives on mental health are shaped by personal experiences, cultural factors, and broader social support systems.
4. **Gender & Traditional Beliefs:**
   * Responses to the fefam variable revealed differences in perspectives on gender roles, with men being slightly more likely to support traditional roles compared to women. However, younger respondents were more likely to reject the idea that men should be the primary earners, highlighting a generational shift in attitudes.

## Conclusion

This exploratory analysis of the GSS dataset highlights how demographic and social factors influence happiness and mental health perceptions. Key findings include:

* **Health and marital status** are strong predictors of happiness.
* **Views on suicide** vary widely, and personal happiness levels may be associated with these attitudes.
* **Generational shifts in gender roles** indicate evolving social norms over time.

Future research could further explore these relationships by incorporating additional variables such as education level, income, or political beliefs. Understanding these dynamics is essential for policymakers, mental health professionals, and sociologists aiming to improve societal well-being.