

## Data Appendix

### Introduction

The purpose of this project is to determine what factors impact individuals' opinion on legalizing abortion. This project is targeted towards campaigners so they can better understand voter habits and be effective in their campaigns.

### DF Dataset Overview

- Entries and Variables: The dataset contains 72,390 entries, each representing a survey response. There are 11 variables in total, which include a mix of demographic information, political views, religious affiliations, and opinions on specific issues such as pro-choice and pro-life stances.
- Time Frame: In the original DF dataset all the data span the year 1972 to 2021.
- Dataset CSV: The dataset was saved to a csv called “df.csv” .

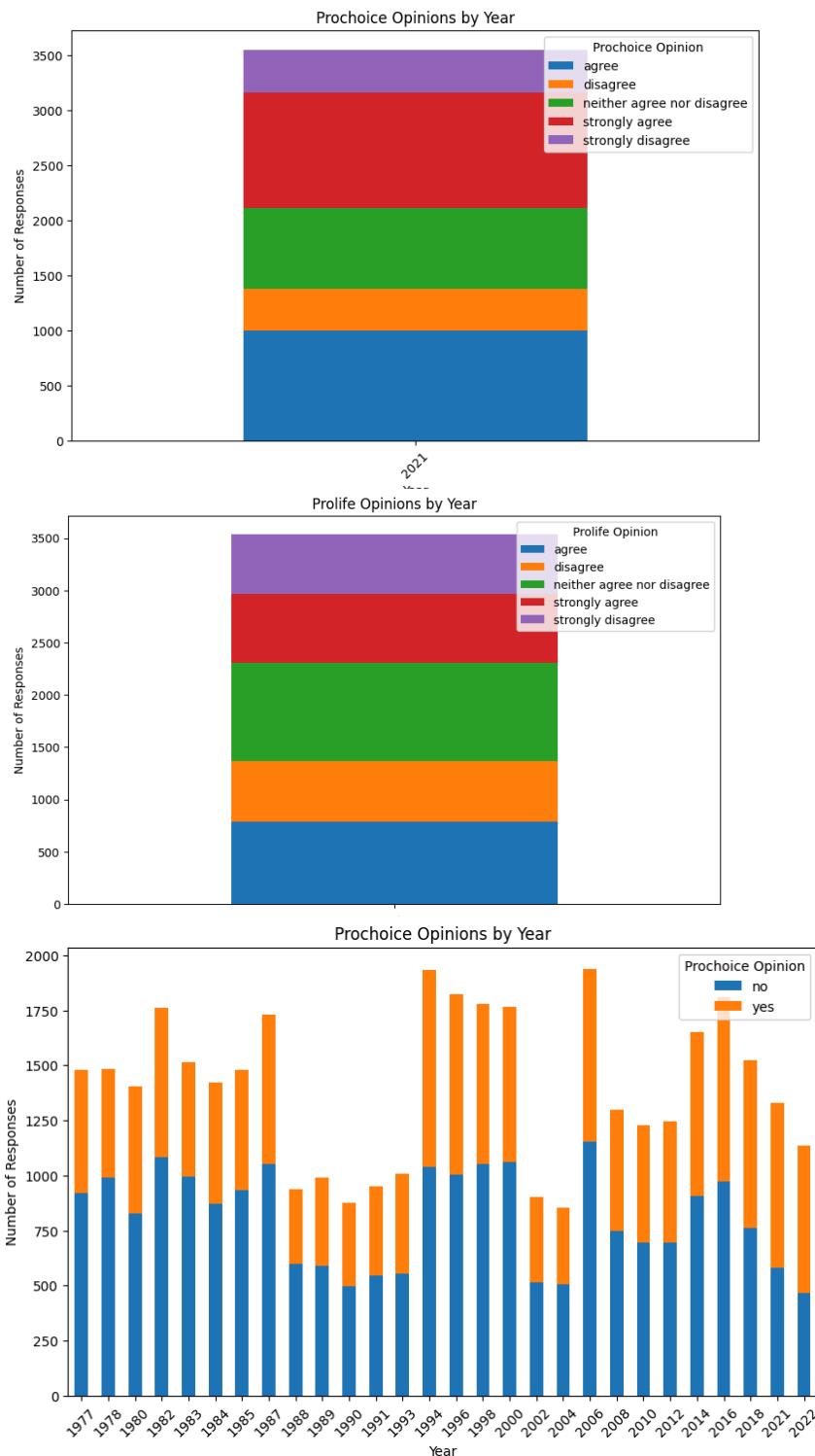
### DF Variable Descriptions:

- prochoic: Categorical variable indicating respondents' agreement or disagreement with pro-choice views. It has categories like 'strongly agree', 'disagree', 'agree', 'neither agree nor disagree', and 'strongly disagree'.
- partyid: Categorical variable representing respondents' political affiliation, with options such as 'independent, close to democrat', 'not very strong republican', 'strong democrat', and more.
- year: Integer variable indicating the year of data collection, which is 2021 for all entries.
- age: Numeric variable (float64) representing respondents' age. The dataset includes a wide age range from 18 to 89 years.
- income: Categorical variable describing respondents' income levels, with ranges from 'under \$1,000' to '\$25,000 or more'.
- polviews: Categorical variable capturing political views on a scale from 'conservative' to 'extremely liberal'.
- relig: Categorical variable indicating respondents' religious affiliations, including 'protestant', 'muslim/islam', 'none', 'catholic', etc.
- region: Categorical variable showing the geographic region of respondents, such as 'middle atlantic', 'pacific', 'east north central', etc.
- abany: Categorical variable regarding opinions on the right to an abortion, with responses 'yes' or 'no'.
- prolife: Similar to prochoic, this categorical variable reflects respondents' pro-life stance, with similar categories for agreement and disagreement.
- sex: Categorical variable indicating respondents' sex, with options 'female' and 'male'.

## Missing Values and Data Integrity

- The dataset initially contained missing values in several columns, such as prochoic, partyid, age, income, polviews, relig, abany, and prolife. After graphing the three categorical variables on individual people's opinions on abortion (prochoic, prolife, and abany) it was revealed that the variables prochoic and prolife were only recorded in 2021 which was the major cause of the missing variables.

Figures 1-3: Responses for variables Prochoic, Prolife, and Abany based on year



### **Cleaned DF Dataset Overview**

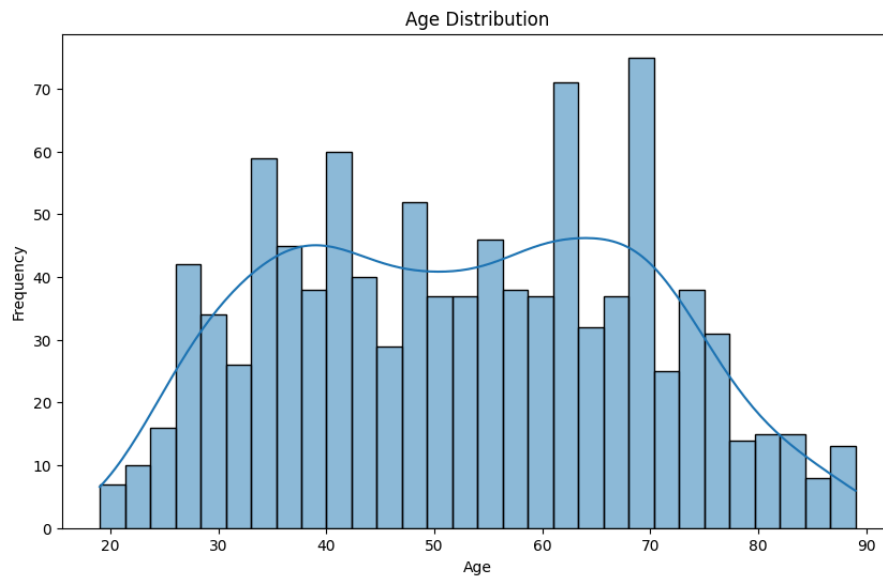
- **Entries and Variables:** The dataset contains 1027 entries, each representing a survey response. There are 11 variables in total, which include a mix of demographic information, political views, religious affiliations, and opinions on specific issues such as pro-choice and pro-life stances.
- **Changes in the dataset:** The data set was cleaned to only contain data recorded from the year 2021. During the cleaning process duplicates and rows with missing variables were dropped. The categorical variables were formatted to prevent typographical errors or variations in capitalization that might introduce redundancy.
- **Time Frame:** In df dataset all the data is from the year 2021.
- **Dataset CSV:** The cleaned df dataset was saved the a csv file called “cleaned\_data.csv”

### **Cleaned DF Variable Descriptions:**

- **prochoic:** Categorical variable indicating respondents' agreement or disagreement with pro-choice views. It has categories like 'strongly agree', 'disagree', 'agree', 'neither agree nor disagree', and 'strongly disagree'.
- **partyid:** Categorical variable representing respondents' political affiliation, with options such as 'independent, close to democrat', 'not very strong republican', 'strong democrat', and more.
- **year:** Integer variable indicating the year of data collection, which is 2021 for all entries.
- **age:** Numeric variable (float64) representing respondents' age. The dataset includes a wide age range from 18 to 89 years.
- **income:** Categorical variable describing respondents' income levels, with ranges from 'under \$1,000' to '\$25,000 or more'.
- **polviews:** Categorical variable capturing political views on a scale from 'conservative' to 'extremely liberal'.
- **relig:** Categorical variable indicating respondents' religious affiliations, including 'protestant', 'muslim/islam', 'none', 'catholic', etc.
- **region:** Categorical variable showing the geographic region of respondents, such as 'middle atlantic', 'pacific', 'east north central', etc.
- **abany:** Categorical variable regarding opinions on the right to an abortion, with responses 'yes' or 'no'.
- **prolife:** Similar to prochoic, this categorical variable reflects respondents' pro-life stance, with similar categories for agreement and disagreement.
- **sex:** Categorical variable indicating respondents' sex, with options 'female' and 'male'.

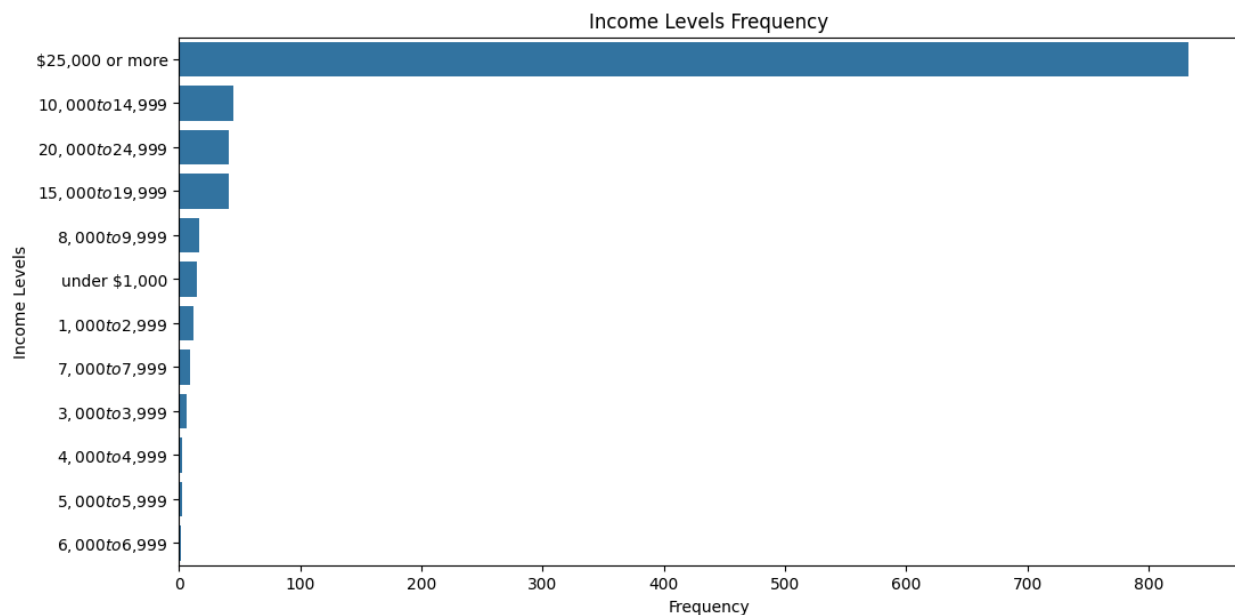
### **Detailed Statistical Analysis:**

#### **Age Distribution of Survey Respondents**



The age distribution of respondents shows a median age of 35 years, with the youngest participant being 18 and the oldest 89. The interquartile range, from the first quartile (Q1) at 25 years to the third quartile (Q3) at 50 years, suggests that half of the survey participants are within this age bracket. The distribution is slightly right-skewed, indicating a younger respondent base, with outliers suggesting a smaller proportion of participants older than 70.

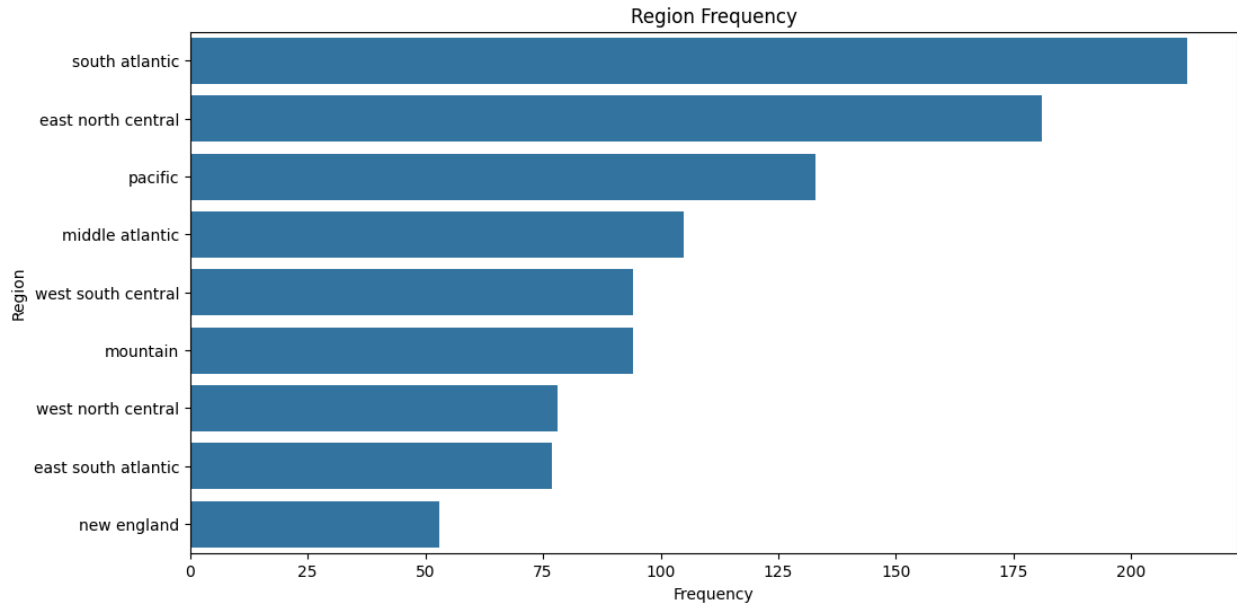
### Income Level Distribution Among Respondents



The income level distribution reveals a varied economic background among the respondents. The most common income bracket reported is '\$25,000 or more,' representing approximately 30% of responses, suggesting a significant portion of the population falls within a higher income range.

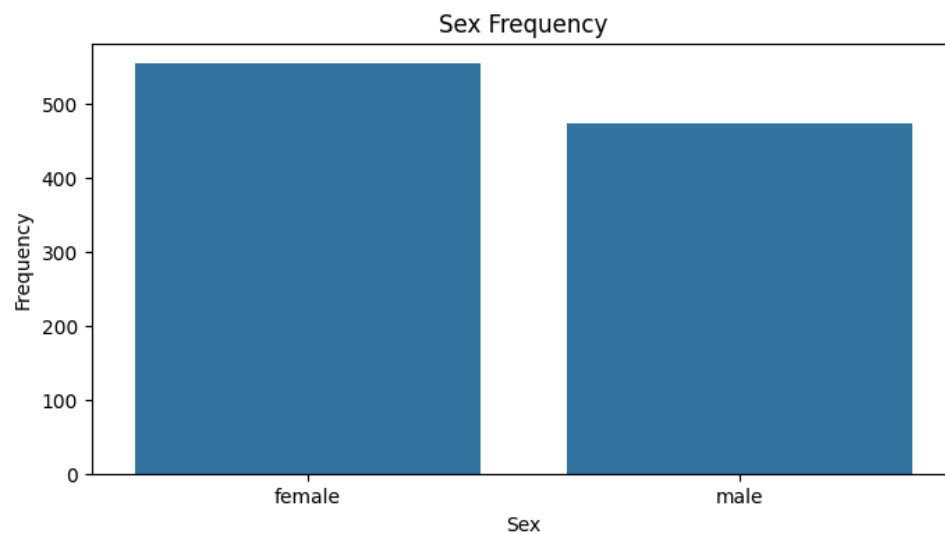
The least reported income bracket is 'under \$1,000,' indicating fewer participants at the lowest income level. This distribution highlights the economic diversity of the survey participants.

### Geographic Distribution of Survey Participants



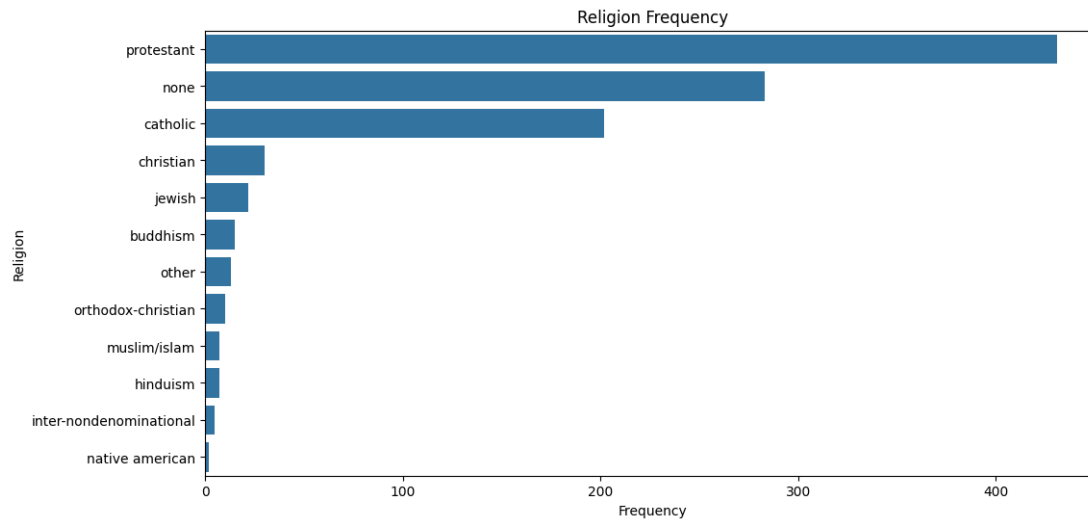
Region-wise distribution of respondents shows a higher concentration in the 'east north central' and 'south atlantic' regions, each accounting for roughly 20% of the total responses. The 'mountain' region has the fewest respondents, suggesting geographical disparities in participation or population density across regions.

### Gender Distribution of Respondents



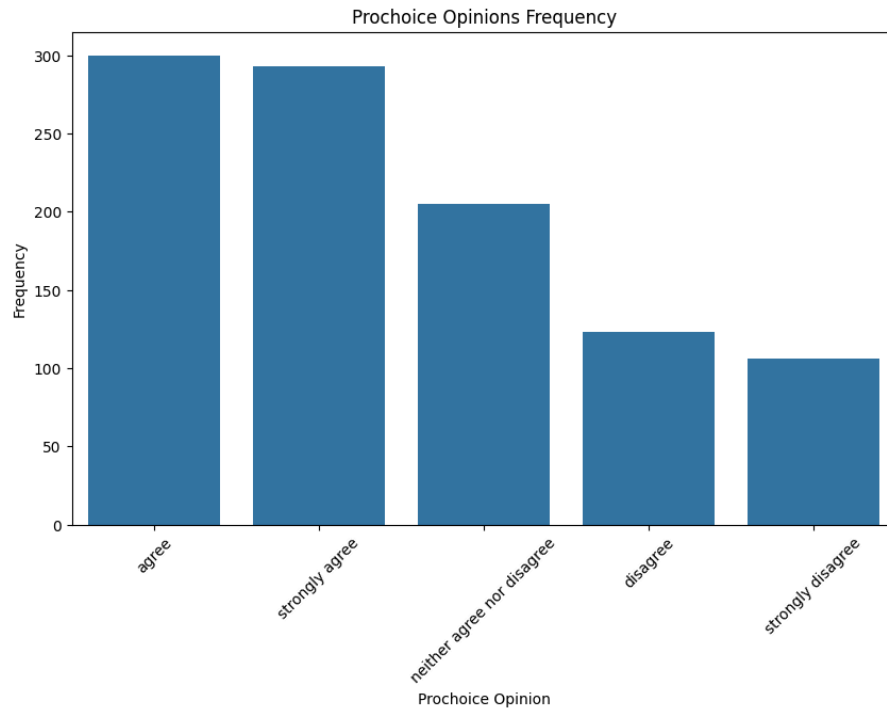
The gender distribution of respondents is nearly evenly split, with a slight majority identifying as female. This balanced distribution ensures that the survey results can be considered representative of both male and female perspectives on the issues explored.

### Religious Affiliation Among Respondents



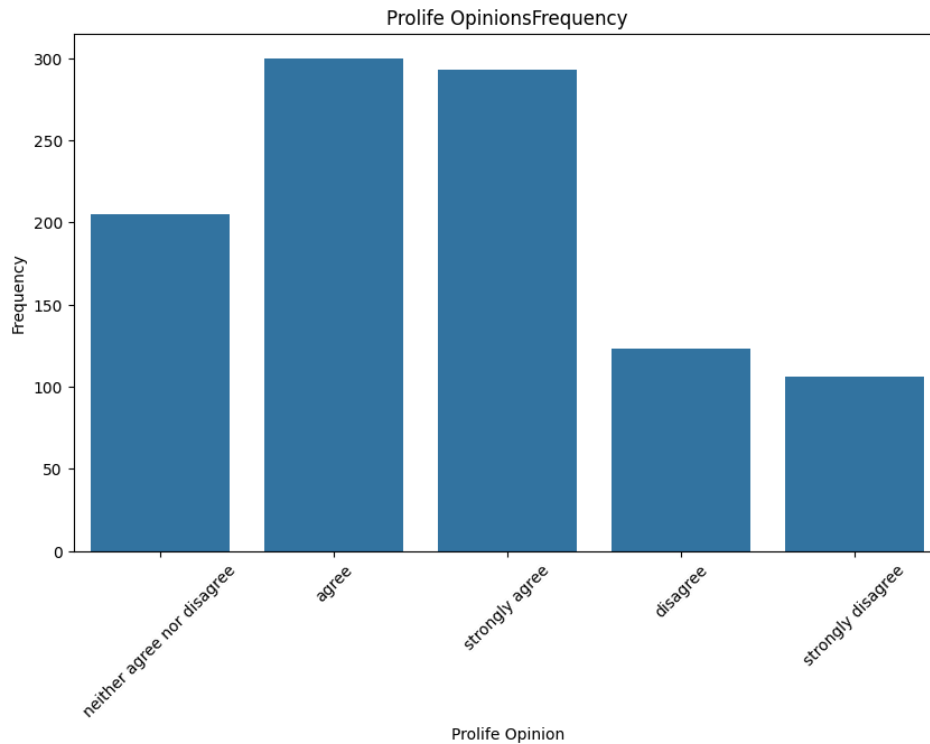
Religious affiliation among respondents is diverse, with 'protestant' being the most reported, followed by 'catholic' and 'none'. This indicates a mix of religious and non-religious perspectives within the survey population, offering insights into the role of religious beliefs in shaping opinions on surveyed topics.

## Distribution of Prochoice Opinions



This figure illustrates respondents' opinions regarding prochoice, revealing a close division within the surveyed population. A total of 593 respondents expressed support for prochoice, with 300 agreeing and 293 strongly agreeing with prochoice principles. On the other hand, a combined total of 229 respondents expressed opposition, including 123 disagreeing and 106 strongly disagreeing with prochoice. Additionally, 205 respondents remained neutral, neither agreeing nor disagreeing with prochoice views. This distribution highlights a slight majority in favor of prochoice opinions among the participants.

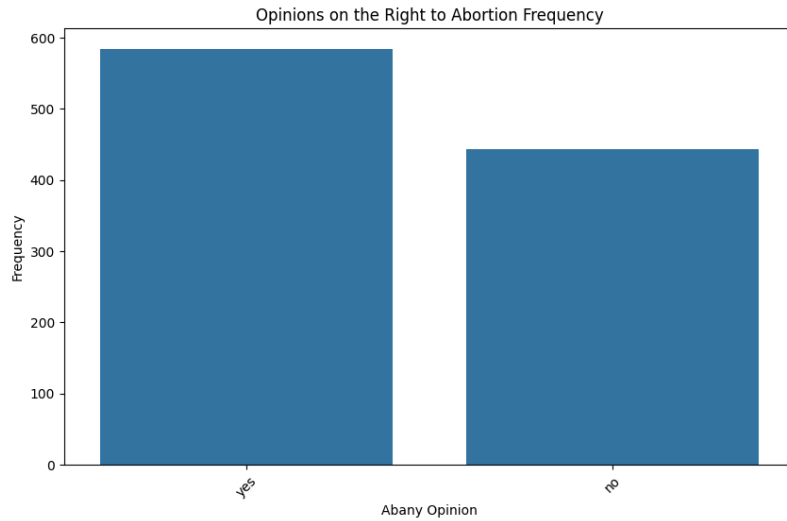
## Distribution of Prolife Opinions



The distribution of prolife opinions among survey participants indicates a relatively even spread across different viewpoints. The largest group, comprising 257 respondents, neither agreed nor disagreed with prolife statements, indicating a significant portion of neutrality. Those in agreement with prolife opinions (strongly agree and agree) totaled 409, with 182 strongly agreeing and 227 agreeing. Similarly, opposition to prolife opinions was nearly equal, with 181 disagreeing and 180 strongly disagreeing, totaling 361. This balance underscores the complexity and divisiveness of prolife sentiments within the surveyed group.



## Opinions on the Right to Abortion



This bar chart summarizes respondents' opinions on the right to abortion, offering a snapshot of societal attitudes towards this contentious issue. The frequencies are divided into 'yes' and 'no' categories, reflecting support or opposition to abortion rights. There is a slight skew towards support for abortion rights ('yes'), demonstrating a societal trend towards maintaining or expanding access to abortion services.

### References (IEEE format):

[1] "GSS 2021 Codebook," *gss.norc.org*.

<https://gss.norc.org/Documents/codebook/GSS%202021%20Codebook.pdf>

[2] "R3 Release Variables" *gss.norc.org*.

<https://gss.norc.org/Documents/other/R3%20Release%20Variables.pdf>