2. Write a short description of the data you chose, and why.

The GSS is a nationally representative survey in the United States that gathers data on a wide range of social, economic, and political factors. The GSS was first conducted in 1972, and it has provided insights into how people's beliefs, behaviors, and experiences evolve over time. The dataset allows us to explore key aspects of income, education, job satisfaction, political ideology, and happiness, which will help us understand the relationships between these factors in American society.

For this analysis, I selected 10 key variables that aim to provide a well-rounded view of individual and societal outcomes. These variables cover demographic details, economic factors, work-life balance, and political and social perspectives.

The demographic variables include the respondent's age, gender, and racial identity, which help me understand how different groups experience social and economic conditions. Highest level of education completed and inflation-adjusted family income will allow me to analyze the impact of education on financial well-being. Hours worked last week and job satisfaction will provide insight into work-life balance and employment conditions.

Additionally, marital status, political ideology, and general happiness help me explore how personal relationships and political beliefs might be linked to well-being. By examining these variables together, I can discover meaningful patterns in how social and economic factors interact in people's lives.

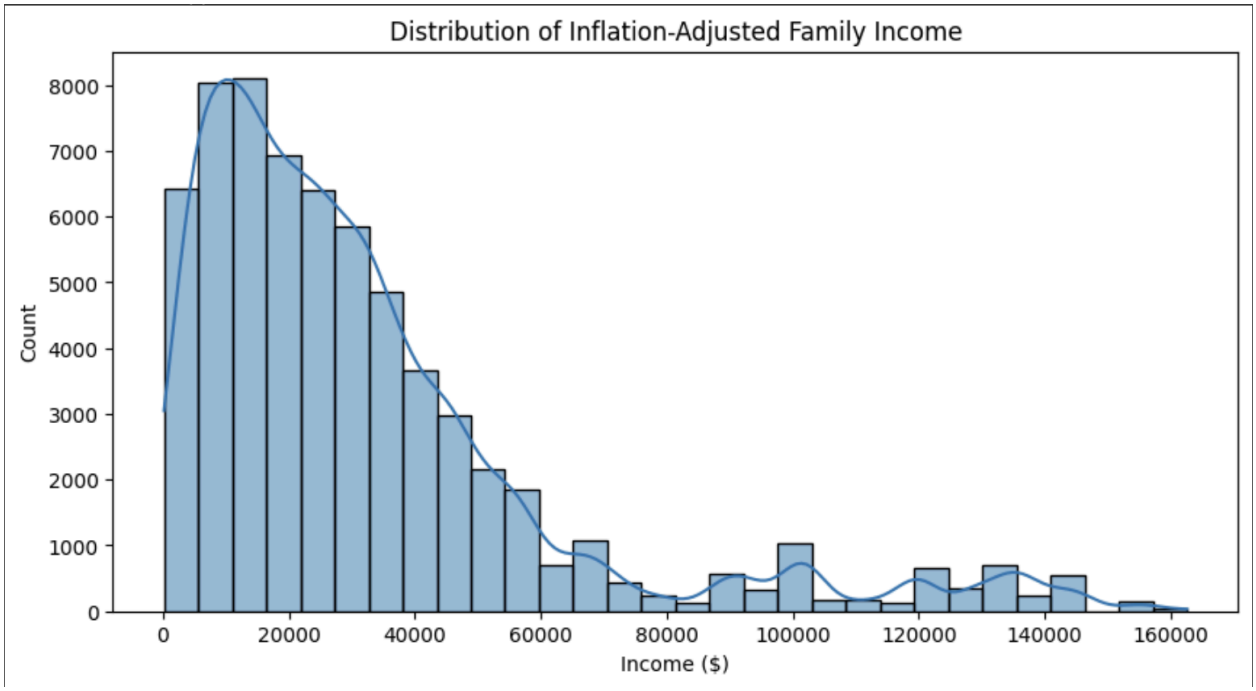
This analysis aims to answer several important questions about personal and economic well-being. First, I want to determine whether higher education leads to higher income and whether people who earn more money tend to be more satisfied with their jobs. I also want to see if people who report higher job satisfaction are generally happier in life.

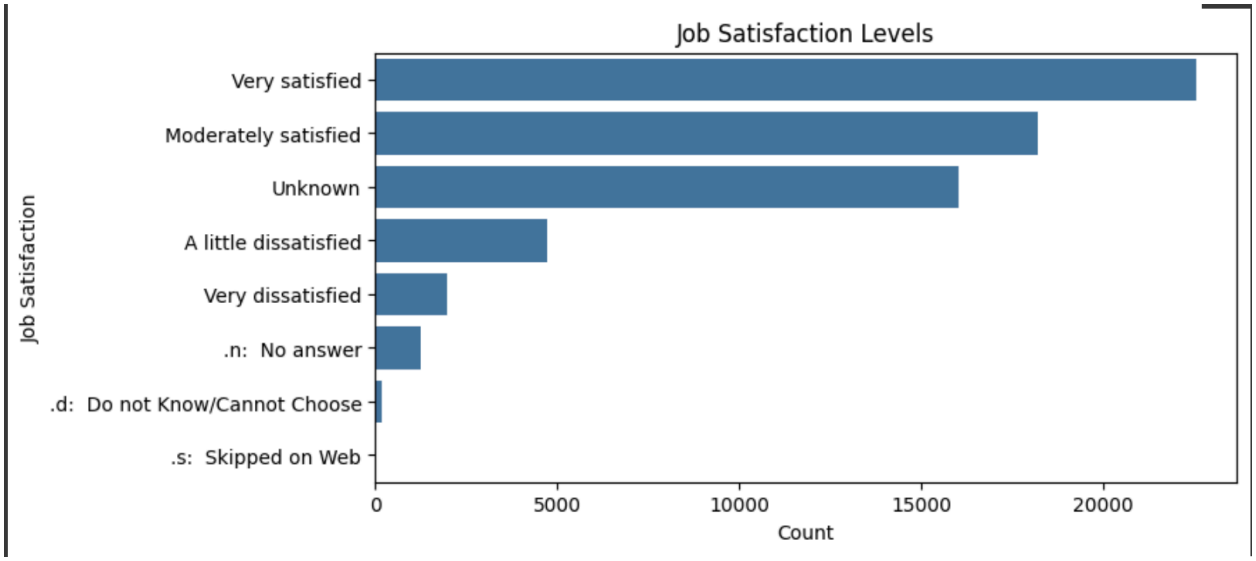
Beyond economics, I am interested in the role of political ideology in shaping people's experiences. Do political views correlate with income levels or happiness? Finally, I will examine potential disparities by looking at differences in work hours across gender and racial groups. These questions will help me better understand the factors that contribute to economic success, job satisfaction, and overall happiness.

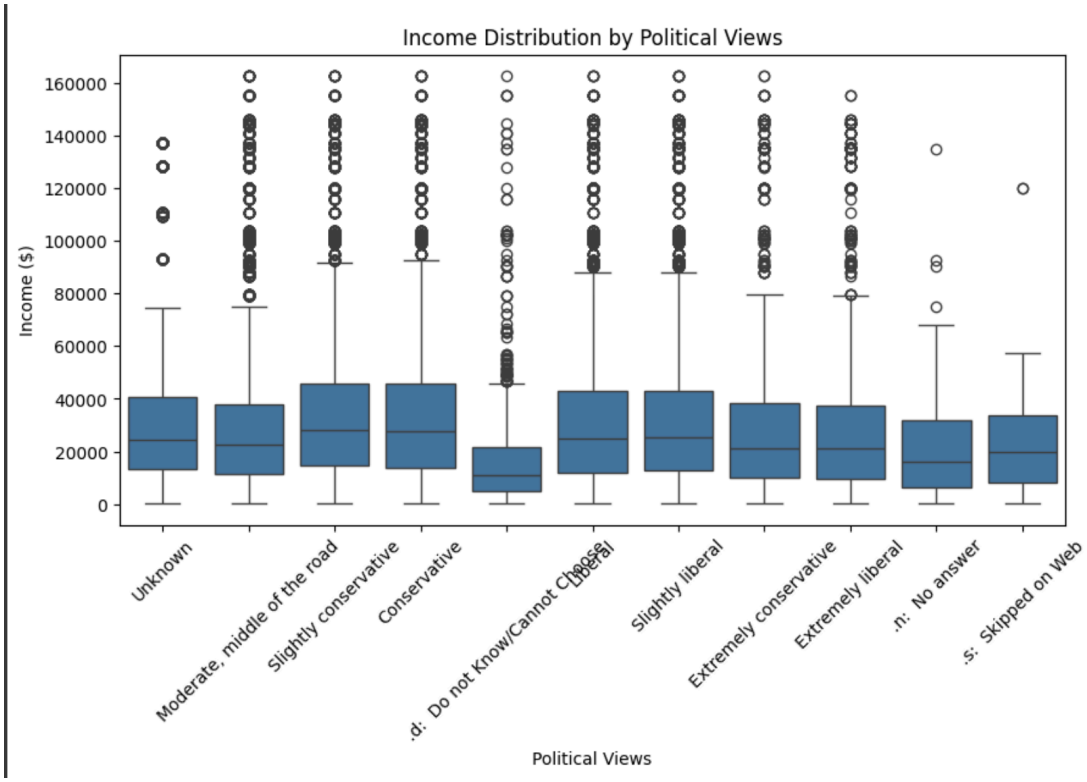
To explore these relationships, I will start with descriptive statistics, summarizing key trends using measures such as averages, medians, and distribution. I will then use data visualizations, including bar charts and scatter plots, to highlight patterns in the data. Finally, I will conduct correlation analysis to quantify the strength of relationships between variables such as education and income, job satisfaction and happiness, and political views and well-being.

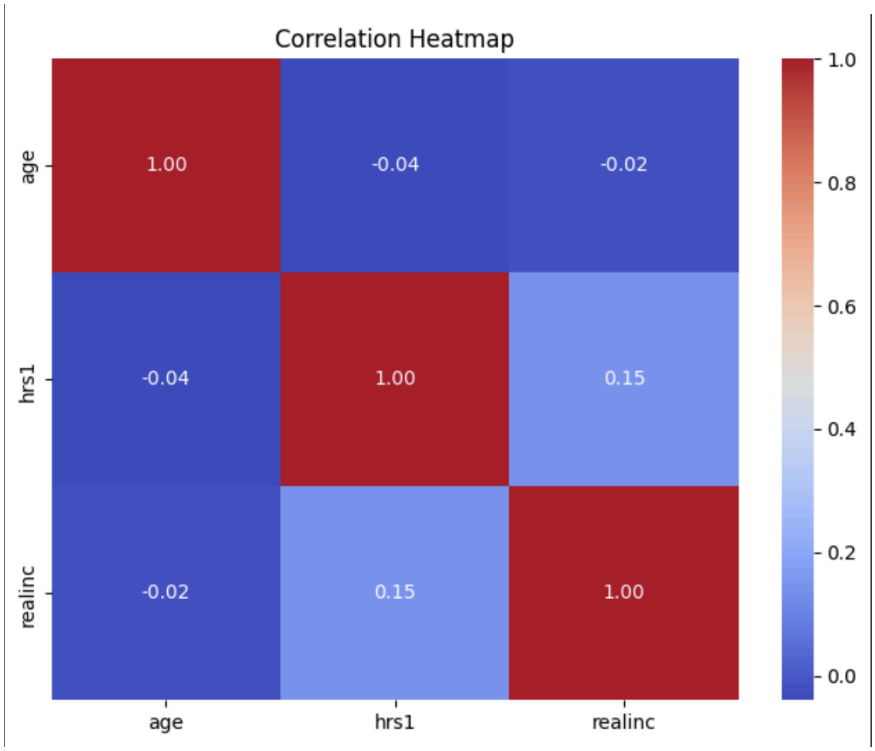
By applying these analytical techniques, I aim to uncover insights into the social and economic dynamics that shape people's lives. This study will provide a clearer picture of how education, income, work, and political ideology influence personal happiness and overall life satisfaction.

4. Produce some numeric summaries and visualizations






Income Distribution by Political Views
Income ($)
160000
140000
120000
100000
80000
60000
40000
20000
0
Unknown
Moderate, middle of the road
Slightly conservative
Conservative
.d: Do not Know/Cannot Choose
Liberal
Slightly liberal
Extremely conservative
Extremely liberal
.n: No answer
.s: Skipped on Web
Political Views


Correlation Heatmap
age
hrs1
realinc
1.00
-0.04
-0.02
-0.04
1.00
0.15
-0.02
0.15
1.00
age
hrs1
realinc
1.0
0.8
0.6
0.4
0.2
0.0

5. Describe your findings.

The exploratory data analysis of the GSS dataset I narrowed down provides insights into various socioeconomic factors, including age, work hours, income distribution, job satisfaction, and political views. These findings help us understand patterns in employment, earnings, and personal well-being across different demographic groups.

One of the first observations is that the average age of respondents is 45.86 years, with most individuals falling between 30 and 50 years old. This suggests that the dataset includes a significant number of working-age individuals, making it suitable for analyzing labor market trends.The average reported workweek is around 41 hours, which aligns with standard full-time employment. However, the variation in work hours, with some respondents reporting as few as 0 hours and others as many as 88 hours per week, indicates that there are clearly differences in employment status, which may range from part-time work, to self-employment, or multiple job holdings.

Income distribution in the dataset is right-skewed, which means that while most respondents earn on the lower end, a smaller group reports significantly higher incomes. The median inflation-adjusted income is $12,080, which is relatively low and suggests a large proportion of respondents earn below national averages. At the same time, a small percentage of individuals report annual earnings exceeding $100,000, highlighting the presence of income inequality. The wide income gap suggests that there may be many external factors such as education, experience, and industry type that are likely to play an important role in determining earnings.

Job satisfaction appears to be largely positive, with the majority of respondents reporting being either “Very Satisfied” or “Moderately Satisfied” with their jobs. However, a noticeable percentage of individuals report dissatisfaction, indicating that workplace conditions, wages, or job security could be contributing to lower satisfaction levels. Understanding the relationship between income and job satisfaction further insight into whether financial compensation plays a dominant role in workplace contentment.

When analyzing the relationship between income and political views, some variation is observed. Certain conservative and moderate groups tend to report higher income levels on average, though the differences are not extreme. This suggests that while political ideology and earnings may be loosely connected, other factors such as education, job industry, and geographic location likely have a more substantial influence on income levels.

The correlation analysis further reveals some interesting patterns. There is a weak positive correlation (0.15) between work hours and income, meaning that individuals who work more hours tend to earn slightly more, though the effect is not very strong. This indicates that while longer work hours may contribute to higher earnings, other factors such as skill level, job type, and education likely have a greater impact. Additionally, age and income show almost no correlation, suggesting that older individuals do not necessarily earn more than younger

respondents. This could be due to various reasons, including career stagnation, retirement, or wage disparities across generations.

In conclusion, this exploratory data analysis highlights several key socioeconomic trends, including income inequality, job satisfaction variations, and potential links between political views and earnings. While work hours and income show a slight correlation, it is evident that other underlying factors contribute significantly to economic outcomes. Further research could explore additional elements such as education levels, marital status, and racial disparities in income and employment to provide a more comprehensive understanding of these trends.