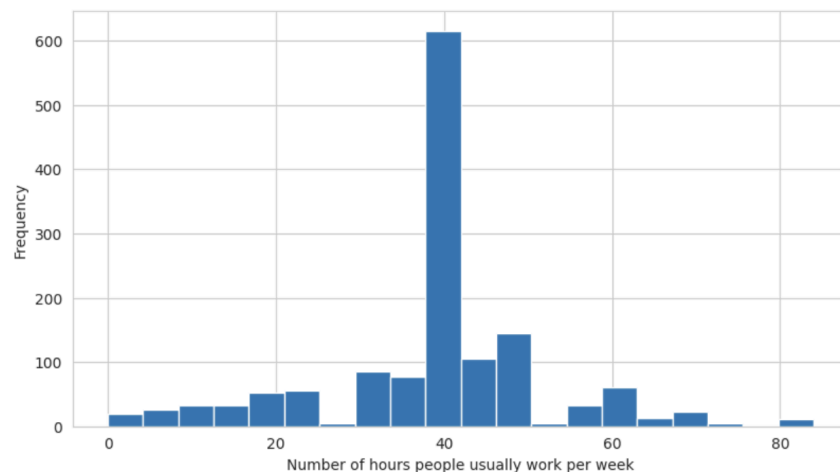1. **Data description**

   As the General Social Survey has been a reliable source of data to help monitor and explain trends in American behaviors, demographics, and opinions, I was interested in exploring education and career aspects within the family or household context. 7 variables of interest were selected and extracted from the GSS Data Explorer to drive potential insights, including hrs2, commute, educ, speduc, res16, hompop, and income, which will be illustrated in detail below.
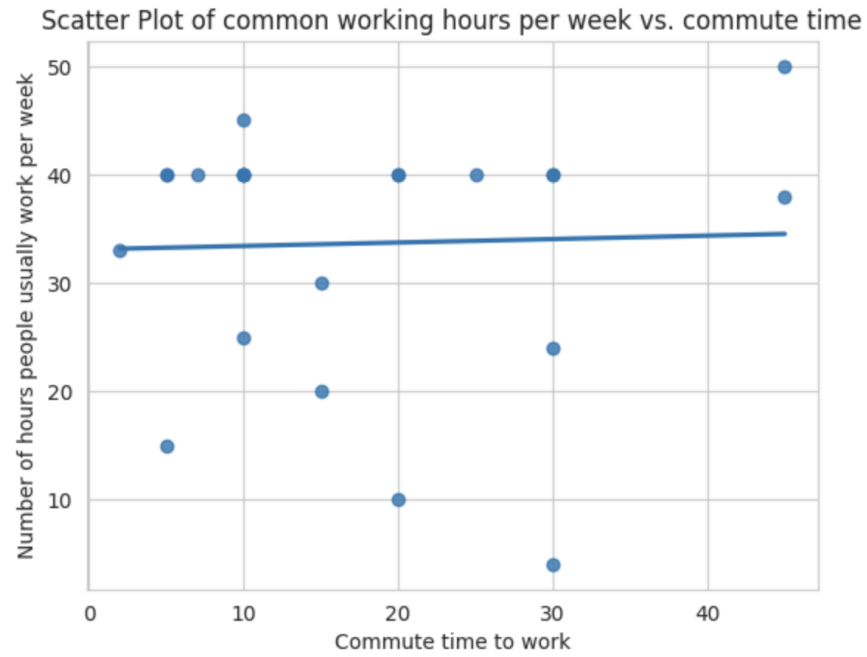
   a. Hrs2: number of hours usually work a week
   b. Commute: travel time to work
      i. Expected to correlate with "hrs2" as longer commute time might cause people to reduce working hours for convenience
   c. Educ: highest year of school completed
   d. Speduc: highest year of school completed, spouse
      i. Expected to correlate with respondent's own education level as people are more likely to marry to others with similar highest level of attained education
   e. Res16: type of place lived in when 16 years old
   f. Hompop: number of persons in household
      i. Expected to correlate with "res16" as individuals who live in more urban areas at 16 years old would be likely to have fewer household members for higher living costs in larger towns or cities
   g. Income: total family income
      i. Expected to correlate with "hompop" as number of household members intuitively would affect the total family income, which might be in both directions
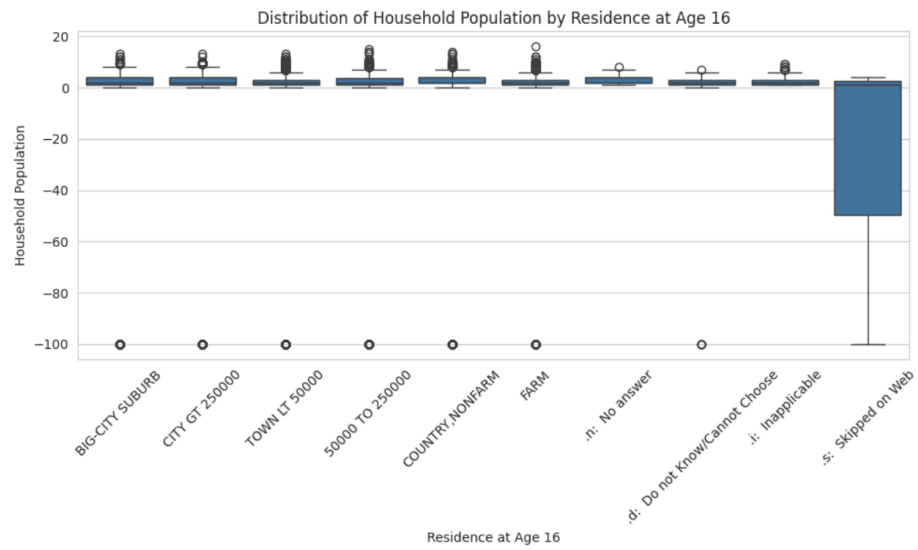
2. **Numeric summaries and visualizations**

   a.

Scatter Plot of common working hours per week vs. commute time

b.

Distribution of Household Population by Residence at Age 16
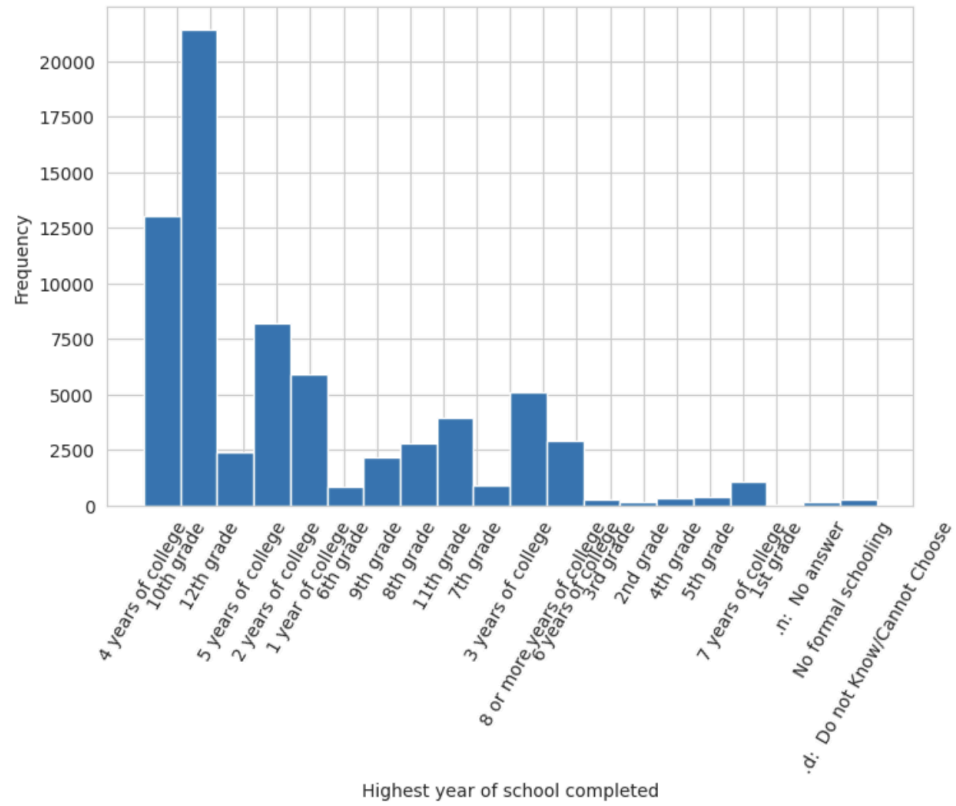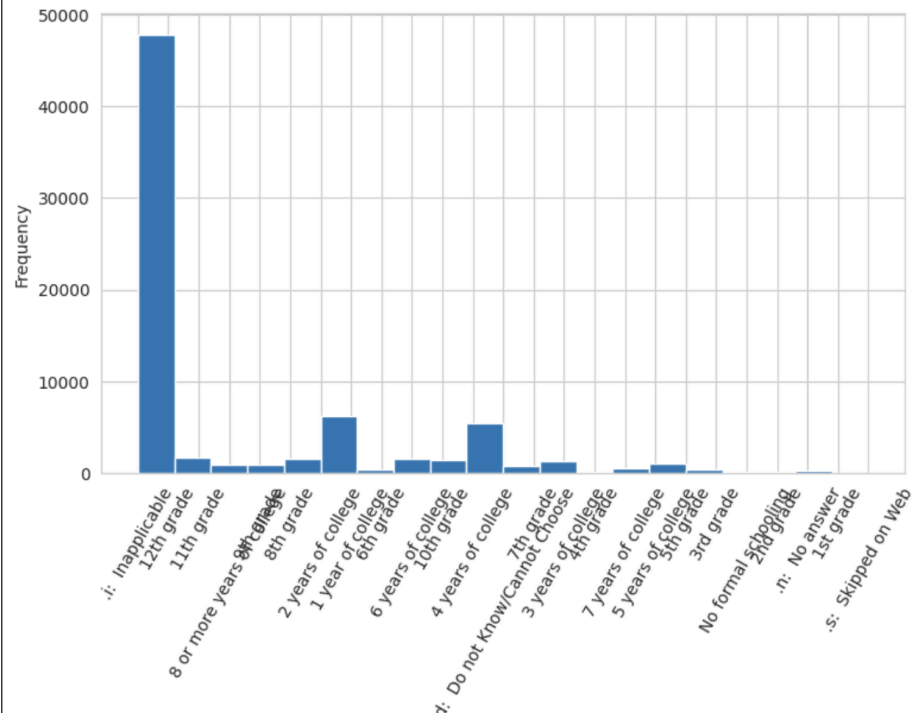
c.

Histogram of respondent's education

d.

Histogram of respondent's spouse education

e.

3. Output discussion

a. The number of hours people usually work per week demonstrates a normal distribution trend, with a mean of 39.06 and a median of 40. This could suggest that American workers have a standard and healthy work-life balance in that people who responded to the survey generally had normal working hours. However, the data extracted only got 1397 number of responses, which might not be representative enough of the American population.

b. The correlation coefficient between the number of hours people usually work per week and commute time is 0.033, suggesting a weak relationship between the two variables. From the scatterplot above, the relationship also seems to be random across the plot.

c. By conducting the contingency table and chi-square statistic for the respondent education and spouse education, the p-value (<0.05) suggests that the two variables are not independent, confirming the expectation that respondent's own highest education could determine or predict their spouse's education. Based on the histograms, the recorded highest education levels completed for both the respondents and the spouses are largely different, which could be attributed to the fact that most of the responses for spouse education are inapplicable, which could suggest that people are not usually willing to disclose or value spouse education while participating in the survey.