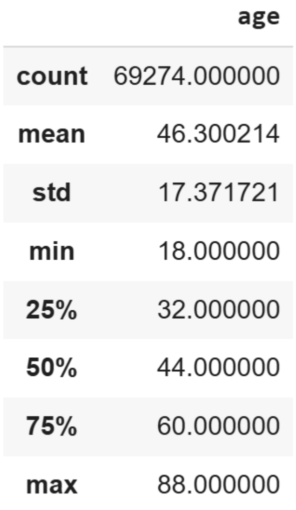
The General Society Survey provided data about many different questions, ranging from basic data like race, age, and sex to marital status, widowed, and spouse's occupation & education. The survey is described as "representative" of the American population. I was most interested in the less-seen trends, such as how family status influenced education or if there were any trends between work status and marital status. I mainly chose these variables to reduce the number of entries that were unknown, inapplicable, had no answers, skipped, or didn't choose. The first set of variables I chose, such as widowed status or commute time, had so many of these unknown entries that it dropped the number of observations from 70,000 to 19 when they were marked as NaN and dropped.

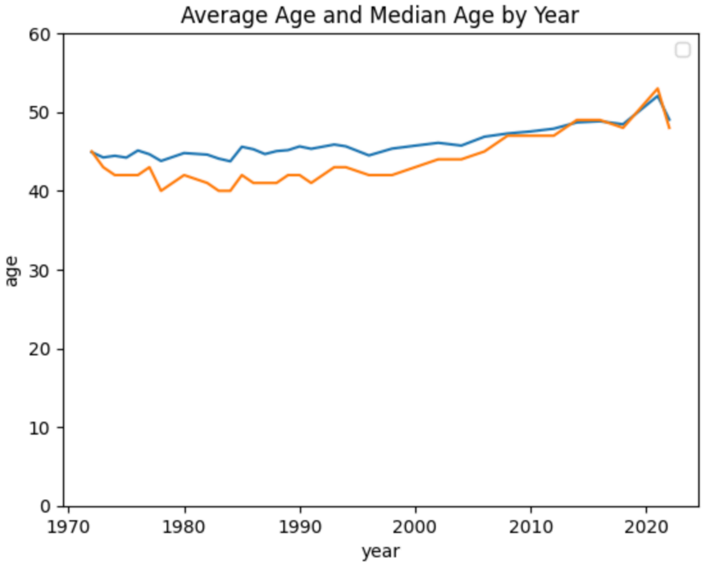
The variables I settled on were marital status, work status, age, education degree, sex, race, place of residence until sixteen, and family status at 16. Work status, as implied, explores a respondent's work status, be it retired, unemployed, in school, working full-time, etc. The variable res16 is the respondent's type of residence when they were 16 years old, be it countryside, city, farm, etc. Finally, the variable family16 records who the respondent lived with up to sixteen years old, ranging from parents to one or two relatives. It also records the gender of the respondent's relative(s). I found all of the response types that would indicate a non-answer like "Did Not Answer", ".s: Skipped on Web", and ".i: Inapplicable", replaced them as NaN null values, and dropped them. That only dropped the number of responses from 72,000 to 69,000. Looks like the survey creators did a good job of reducing the potential number of null responses, and the large majority of respondents were able to find an answer applicable to them.

The first question I wanted to know was the average age of respondents. I have heard professors and statisticians often say that young adults tend to ignore survey requests that just pop into their email inboxes, the mail, or from people standing on the roadside. I wanted to see if that was true using this survey with massive responses. Turns out, that statement is not exactly true. The chart on the right shows the overall average age of respondents while the line plot right below shows the average age of respondents, with the orange line being the median

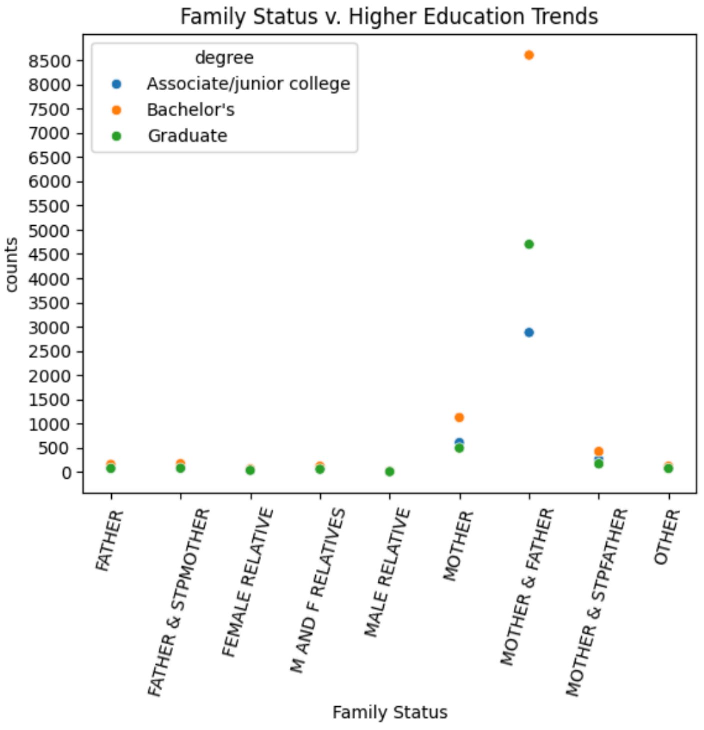


	age
count	69274.000000
mean	46.300214
std	17.371721
min	18.000000
25%	32.000000
50%	44.000000
75%	60.000000
max	88.000000

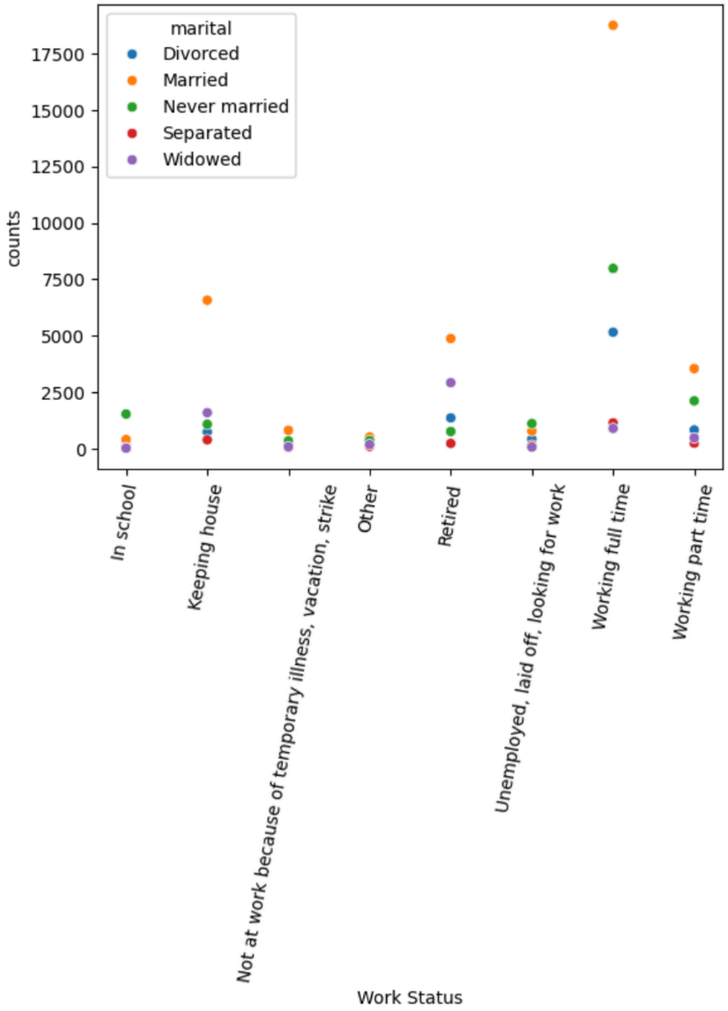
age and the blue line being the mean age. It’s indeed mostly middle-aged folks filling out this survey but the line plot displays that the median and mean age don’t differ by much. When calculated, the mean age is highly influenced by major outliers, meaning if there are somehow any ten-year-olds and below filling out this survey for example, the mean will be dragged down to be lower than it really is. On the other hand, the median is more resistant or robust to it. If the two are close together, it means that most of the people who took this age are centered around the age of 40 to 50 and not older or younger. This means that no matter the reason, while fewer younger folks filled out this survey, there weren’t many older respondents either. This implies that it might be less so that young people don’t fill out surveys and more so that middle-aged people are just more likely to do it or that there’s a greater general population of people around the 40 - 50 age range in the US.



Two variables I wanted to analyze are family status and educational degree, specifically what kinds of family environment best fosters a child to pursue higher education. After isolating variables to higher educational degrees, I created a scatterplot based on how many people got bachelor’s, associate/junior college, or graduate degrees relative to their family status at 16 years. Respondents living with their mother & father had an overwhelming majority, leaving all other categories in the dust. That was expected since a large majority of



respondents marked that they lived with their parents. Parents usually push for their children to pursue higher education for more future career options that bring financial stability, to follow in their footsteps, or to support their child's dreams. What did surprise me, though, is that respondents who only lived with their mother had a much higher postsecondary education completion rate than all other categories, second only to mother & father. From a societal perspective, mothers do tend to push more for higher education because degree=more success, which may be one of the reasons influencing the results.



The last question I analyzed is if marital status and work status have any correlation. From the scatterplot to the left, it's pretty clear that married people have the highest counts across all categories, but that only displays that married people were the most common respondents. Widowed respondents being 2nd highest retired makes sense since it's commonly old age and time that comes with both retirement and death. Keeping house (stay-at-home parent) still seems to be quite common in married couples. The other categories wouldn't normally allow for that because they are single and likely must support themselves. Not too many people said they were in school, which matches the median and mean age analyzed previously. It's mostly those who have never been married in school, which makes sense since being in school requires a lot of time, and many of those respondents might be young adults fresh out of school or going into graduate school.