

Introduction - Chloe Stogsdill

For this analysis, I selected the following variables: year, ballot, wrkstat, child, hrs2, spwrksta, family16, and health. Each of these variables was chosen to provide details and draw conclusions about how an individual's life and social environment can impact their health. The "year" variable was added to help track trends or changes in responses over time and the "ballot" variable was added to help make sure that different versions of the survey are accounted for to account for potential biases in survey responses.

The "wrksta" and "spwrksta" variables were chosen to provide information about both the respondent's and their spouse's labor force status, offering insights into their employment status or- potentially- the absence of a partner to contribute to the financial aspects of the family. Additionally, "hrs2" measures the number of hours a respondent usually works per week, which can be used to analyze job demand in regards to health status.

The "child" variable captures the number of children a respondent has, and the "family16" variable, which indicates whether the respondent lived with their parents at age 16, both provide context for potential stressors and economic status. Lastly, the "health" variable assesses the respondent's self-reported health condition, which allows us to view all of the other variables in comparison to an individual's health.

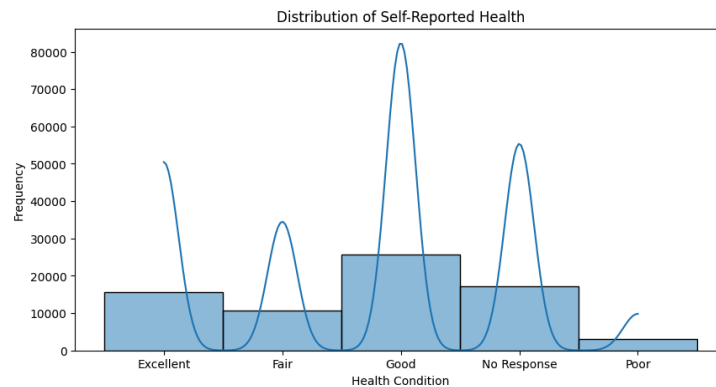
These variables were chosen because they collectively offered a large view of work, family, and health. By analyzing these factors together, we can explore relationships between employment status and family dynamics to a respondent's overall well-being.

Numerical & Graphical Summary

The dataset consists of three numerical variables. Below are the key numeric summaries for those variables:

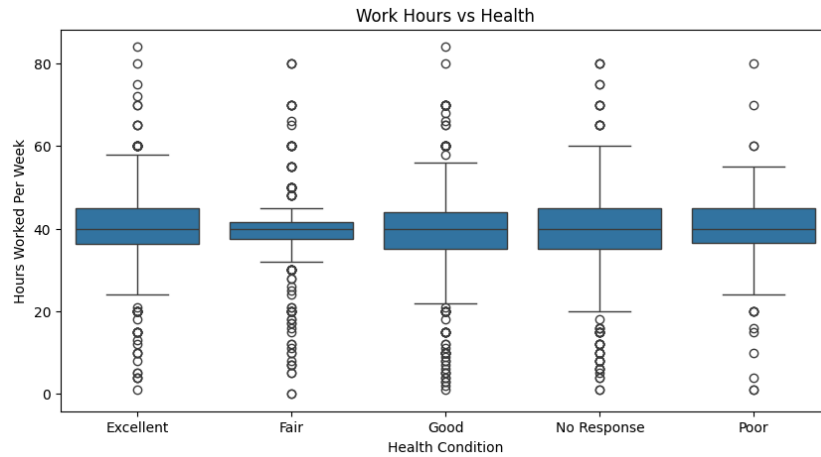
Statistic	year	childs	hrs2
count	72390.000000	71128.000000	1397.000000
mean	1997.715541	1.830925	39.063708
std	15.109995	1.615946	13.303363
min	1972.000000	0.000000	0.000000
25%	1985.000000	0.000000	35.000000
50%	1998.000000	2.000000	40.000000
75%	2010.000000	3.000000	45.000000
max	2022.000000	7.000000	84.000000

Looking at our categorical data, we can analyze the total self-reported health distribution:

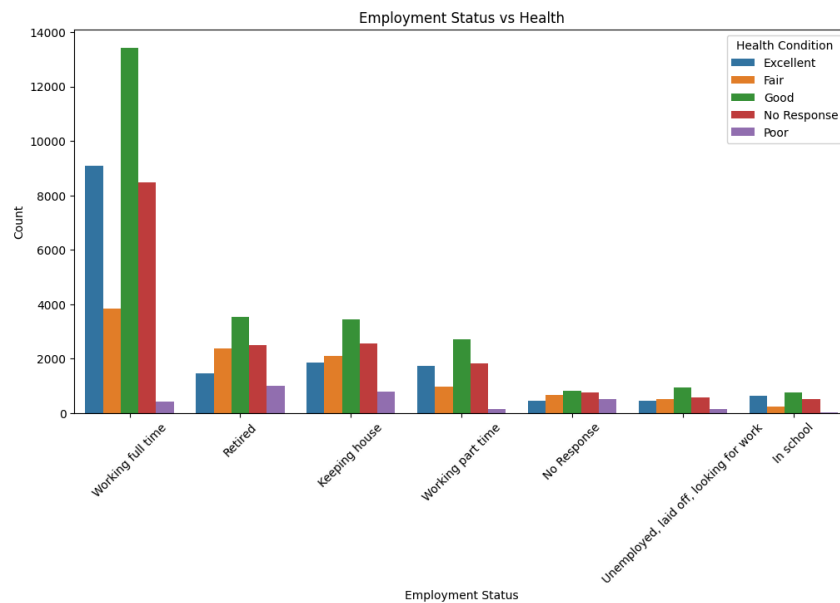


From here, we can see around 825,000 total responses we given. The high percentage of “No Response” indicates possible non-random missing data, which may impact analysis.

Now, we can visualize the correlation between employment status and health:

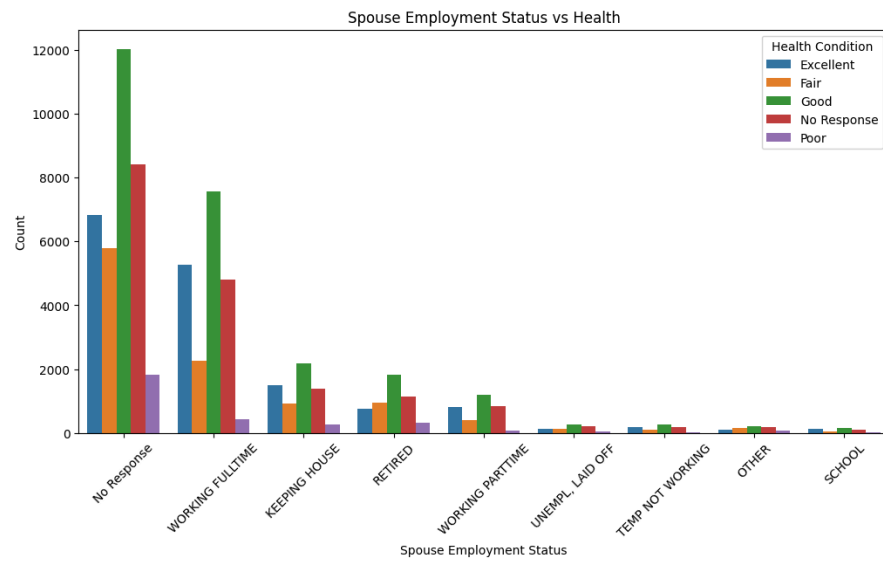


We can see from the graph above that the majority of respondents are employed full-time. There is a strong presence of individuals classified under “No Response,” which is comparable to the full-time working group. As a note, possible outliers for each group may include retired workers, part-time employees, or students and at-home parents. We can look at employment in more detail:

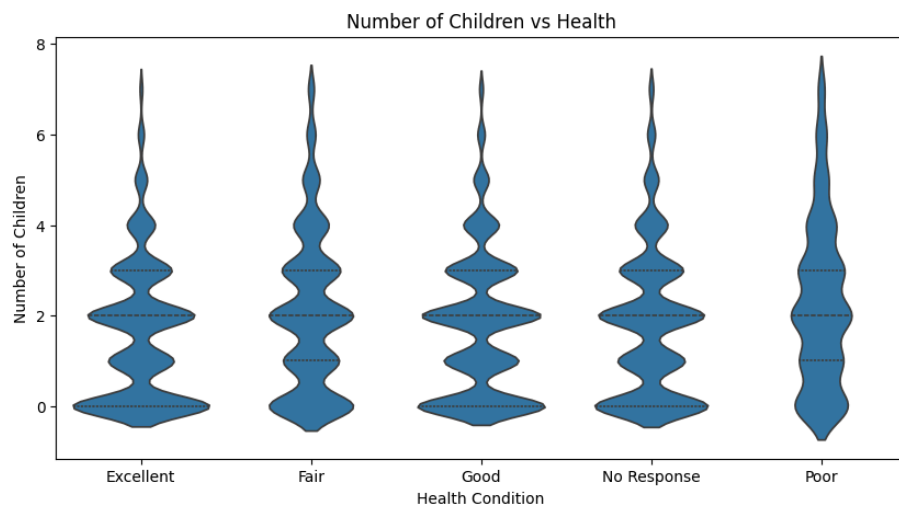


From this, we can see a breakdown of different job statuses and their health breakdowns.

Similarly, we can look at spousal employment status and see if similar trends occur:



Additionally, we can look at the number of children in comparison to reported health:



Analysis & Conclusion

Using the graphs and summaries above, we can provide several different analyses.

From the numerical summary table, we can see that these summaries indicate respondents generally report working around 40 hours per week, with a standard deviation of 13.3 hours. The number of children varies from 0 to 7, with a median of 2— all of which align with the expected population's responses for such questions.

Next, from the total self-reported health distribution, we can see around 825,000 total responses we given: 21% reported “Excellent” health, 15% reported “Fair” health, 33% reported “Good” health, 6% reported “Poor” health, and 24% reported with no response. The high percentage of “No Response” indicates possible non-random missing data, which may impact analysis. The majority of respondents report “Good” or “Excellent” health, with a significant portion not providing a response.

Additionally, using the correlation between employment status and health from the “Work Hours vs. Health” graph, we can see that the vast majority of respondents are employed full time, as every distribution has its median at exactly 40 hours which, when taken in context of the majority of adults, logically stands. The similarity between each group's graph does not lead to many additional conclusions.

However, if we look into other aspects of employment- specifically in the “Employment vs. Health Status” graph- we can see some trends emerge. The proportion of “Excellent” and “Good” health responses for full-time employees differs significantly from any other category. It appears that the more secure your job is, the more likely you are to have good or excellent health. Similarly, full-time workers also report significantly less “poor” health responses in comparison- proportionally- to retired and keeping house groups. Altogether, these responses make logical

sense. Retired workers are likely older, which comes with health concerns of its own, and many workers are forced to retire due to bad health and the inability to continue working. The Keeping House group likely has similar reasons for a larger response of “Poor” health. Many adults who stay home do so due to disabilities or large, at-home duties— all of which have increased stressors. Conversely, “No Response” has a very large overall portion of the responses— due to the unknown nature of this group, it’s hard to draw conclusions about why it's distributed the way it is.

Following that, the “Spousal Employment vs. Health Status” graph gives similar results to its predecessor graph. These groups follow the same general distribution, likely for many of the same reasons. Full time work is stable and offers more benefits to a spouse, so their health is more likely to be “Good” or “Excellent”. Having a retired or stay home spouse likely leads to more stressors- from old age, to disabilities, to children. Again, the “No Response” group has a very large overall portion of the responses— however, due to the unknown nature of this group, it’s hard to draw conclusions about why it's distributed the way it is.

From the number of children vs. health graph, we can see that the number of children per respondent is generally stable across health groups. However, the “Poor” health group does have more responses with 4+ children, which leads to the conclusion that more children can affect health status negatively.

Overall, this analysis provides insights into self-reported health trends and various other variables. In particular, having a full-time job with a spouse who has a full time job and having less than four children will likely mean that you have better overall health than if you were retired or stay-home with a spouse who was retired or stay-home having more than four children.