GSS Exploration and Analysis

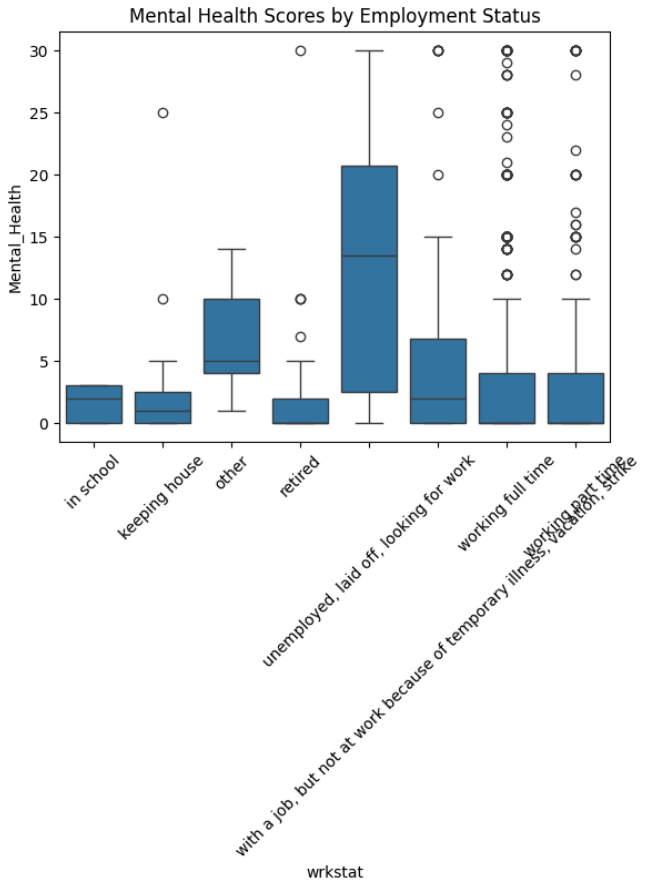
Description of Data

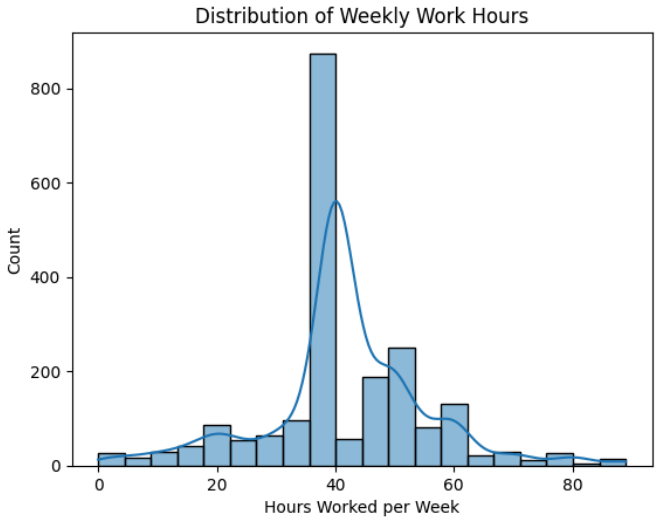
For the EDA of the GSS, I selected the following variables: year, wrkstat, hrs1, mntlhlth, happy, health, age, educ, degree, major1. The variable, **year**, indicates the year the survey was conducted. This was chosen to keep track of the time period in which the other variables in this specific dataset were recorded. Furthermore, **wrkstat** indicates the employment status (full-time, part-time, retired, keeping house, etc.) of respondents. The **hrs1** variable shows the number of hours worked the last week if applicable. Nextly, **mntlhlth** specifies the self reported days of poor mental health status for the last 30 days. The variable, **happy**, shows the self reported general happiness level (very happy, pretty happy, not too happy) of each respondent. Moreover, **health** indicates the self reported general health status (excellent, good, fair, or poor). **Age**, is self explanatory, specifying the age of the respondents, The variable, **educ** is indicative of the highest level of education completed by the respondent. **Degree**, similar to educ, shows the highest degree earned. Lastly, the variable, **major1**, indicates the college major of the respondent.

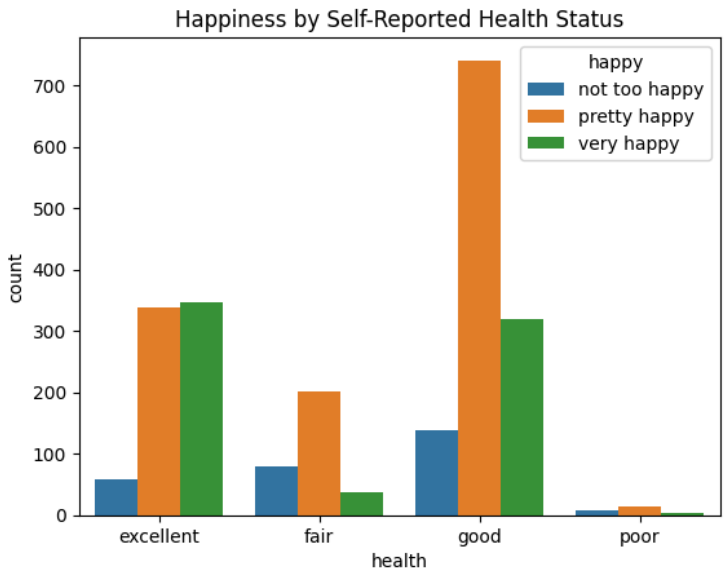
These variables were chosen to explore relationships between work, mental health, happiness, and education. Specifically, I am aiming to investigate how employment status and work hours relate to mental health. Moreover, I am aiming to explore the relationship between education, career choices, and happiness. Finally, I want to see how health and happiness interact across different demographic groups.

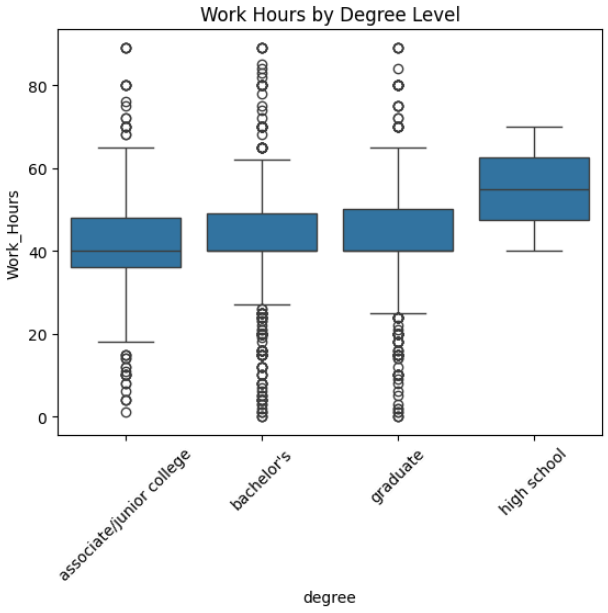
Numeric Findings and Data Visualizations

```
count    2106.000000
mean       41.819563
std        13.780937
min         0.000000
25%        40.000000
50%        40.000000
75%        50.000000
max        89.000000
Name: Work_Hours, dtype: float64
happy
pretty happy     1296
very happy        706
not too happy     284
Name: count, dtype: int64
health
good         1198
excellent     744
fair          318
poor           26
Name: count, dtype: int64
Education_Level
16.0    1003
18.0     303
14.0     273
20.0     239
17.0     234
19.0     117
15.0     101
13.0      15
12.0       1
Name: count, dtype: int64
degree
bachelor's                 1155
graduate                    695
associate/junior college    433
high school                   3
Name: count, dtype: int64
```


Mental Health Scores by Employment Status
Mental_Health
0
5
10
15
20
25
30
in school
keeping house
other
retired
unemployed, laid off, looking for work
working full time
working part time
with a job, but not at work because of temporary illness, vacation, strike
wrkstat


Distribution of Weekly Work Hours
Count
0
200
400
600
800
0
20
40
60
80
Hours Worked per Week


Happiness by Self-Reported Health Status
happy
not too happy
pretty happy
very happy
count
0
100
200
300
400
500
600
700
excellent
fair
good
poor
health


Work Hours by Degree Level
Work_Hours
0
20
40
60
80
associate/junior college
bachelor's
graduate
high school
degree

Findings and Conclusions

I am not too surprised by the results of the data, especially with regards to the visualizations. For instance, it would seem pretty obvious that someone who is unemployed, laid off, and looking for work is more likely to have lower mental health on average for the last 30 days. It also seems pretty obvious that a majority of the people surveyed work around 40 hours a week as that is the gold standard for work hours. One finding I would say was maybe a little surprising was the fact that the average number of work hours for someone with an education of just high school education was higher than the average number of any other education. However, in general, someone with a high school education might have to work multiple jobs, making their schedules more clogged with work. Overall, though, I am pretty unsurprised by what was revealed by these intersections of data, but I will go in more depth with my findings.

For the intersection of employment status and mental health, everyone who answered this survey seemed to be around the same average number of days. However, as previously stated, unemployed individuals (unemployed, laid off, looking for work) tend to report lower mental health more of the days of the month. Also, people working full time and part time had various outliers that skewed the data more to the right.

Next, for the intersection of happiness and health, there were four possible choices for health: “poor”, “fair”, “good”, and “excellent”. Of these, the choice, “good”, had the most number of people answer while “poor” had the least number of people answer. There were three answers for happiness: not too “happy”, “pretty happy”, and “very happy”. Of these, people

reported “pretty happy” the most overall. However, people who rate their health as “excellent” are much more likely to report being “very happy.”

Lastly, for the intersection of education and career choices, as previously stated, having only a high school education correlates with a higher average of work hours. However, there were no outliers for this degree level. On the other hand, every other degree level (associate/junior college, bachelor’s, graduate’s) has a multitude of outliers on both sides. As such, everyone with just a high school degree, worked over 30 hours while every other degree level had people working considerably less hours.

In conclusion, while there were some outliers, the results of the dataset were pretty unsurprising. It’s pretty expected for people who have excellent health to be healthier on average. At the same time, it’s expected for people to work 40 hour work weeks on average because that is what is standard for the United States of America. At the very least, though, this was a very educational exploration for me and I feel as though I learned a good amount about the GSS and cleaning/transforming data.