

## Luke McMeans

### DS 3001 - GSS Lab

(Note: The data frame cleaning and visualizations/summaries were executed at the bottom of the `gss.ipynb` file.)

#### The Data

The retrieved data from the GSS codebook is as follows...

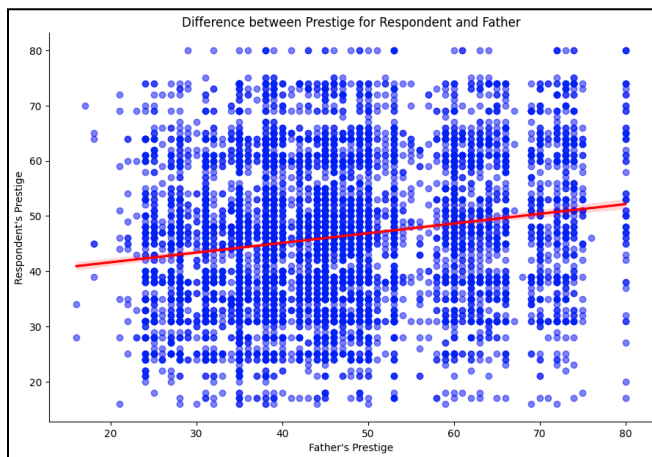
Column Name	Column Description
educ	highest year of school completed
paeduc	highest year of school completed, father
maeduc	highest year of school completed, mother
degree	r's highest degree
padeg	father's highest degree
madeg	mothers highest degree
prestg10	r's occupational prestige score (2010)
mapres10	mothers occupational prestige score (2010)
papres10	father's occupational prestige score (2010)
rincome	respondents income
finrela	opinion of family income
unemp	ever unemployed in the last ten yrs
satfin	satisfaction with financial situation

From this data, I wanted to find the relationships between the respondent's success and that of their parents. The biggest way I wanted to track their success was to see their degrees and the prestige of their jobs. Along with this, I wanted to see the effect of their parents' level of success on the respondent, seeing their income and perspective of money and their status from it. I also felt it was important to see if they had unemployment history as an extra detail. These variables

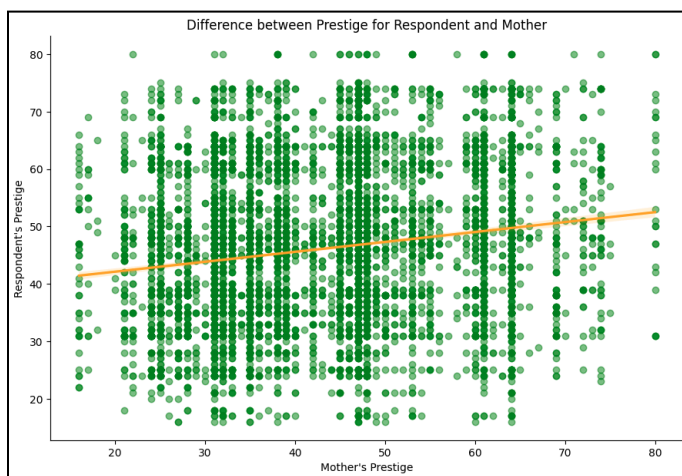
were chosen as a matter to see how legacy impacts the success of children. Some individuals who come from wealth and success could have situations handed to them, and possibly lead to a privileged perspective on money and their status. Of course, there are a strong number of those who have continued success out of their own talents, regardless of their parents status. This is an experiment to see if this relationship could be inferred with data, and not just from observation and narrative.

### Summaries and Visualizations

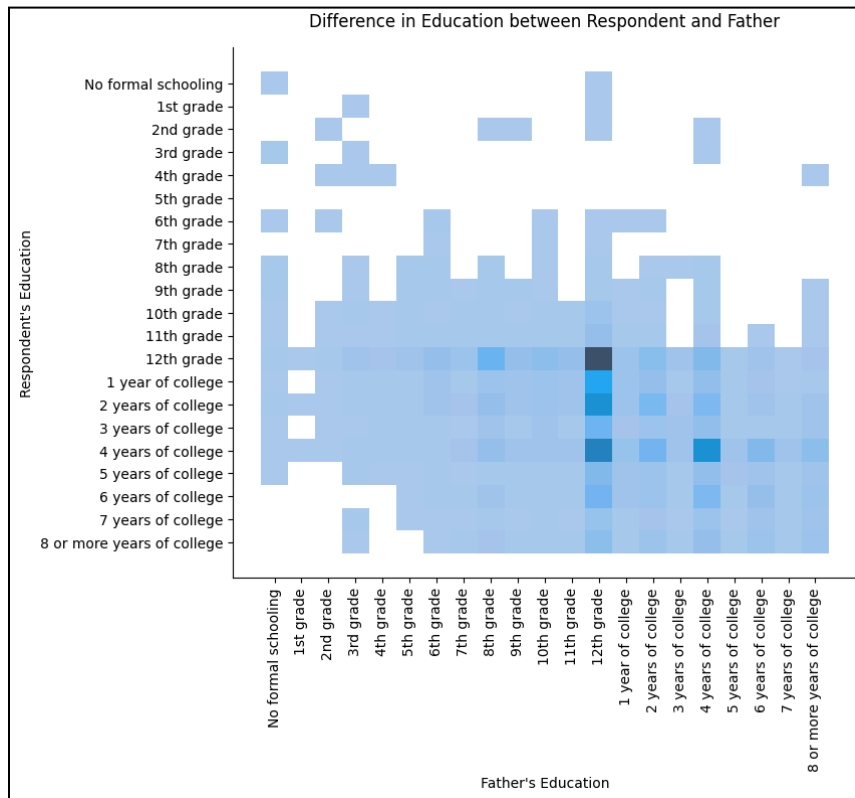
#### *Comparing 'prestg10' and 'papres10'*



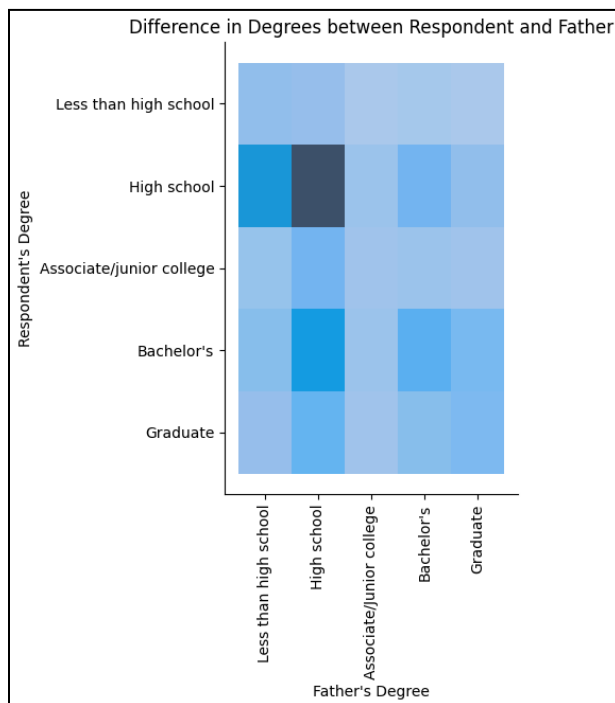
#### *Comparing 'prestg10' and 'mapres10'*



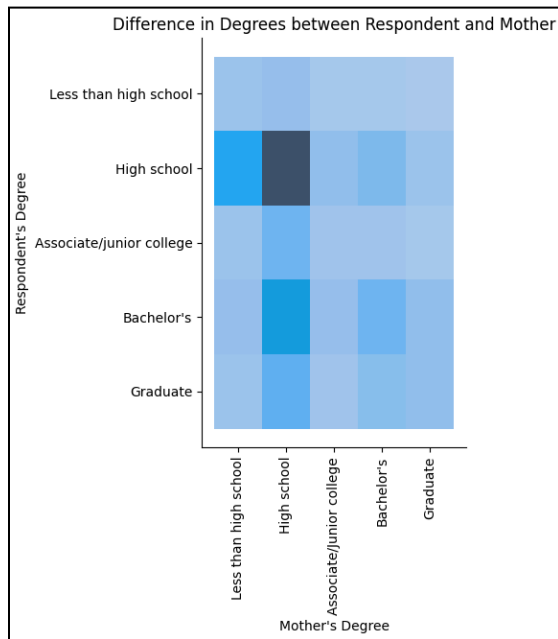
### Comparing 'paeduc' and 'educ'



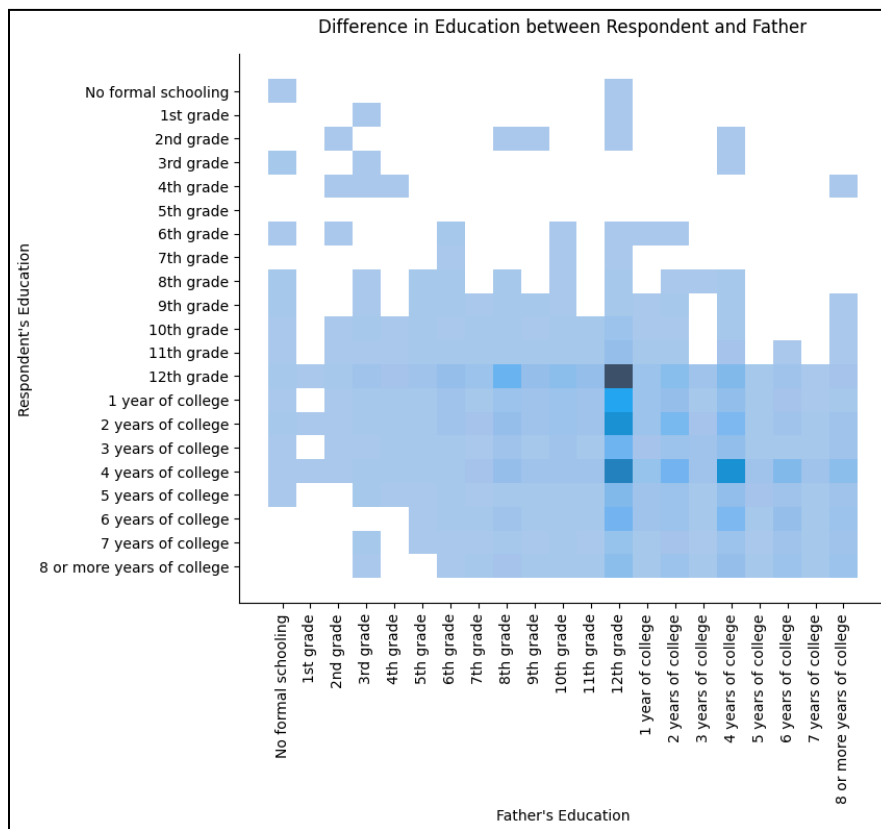
### Comparing 'degree' and 'pdeg'



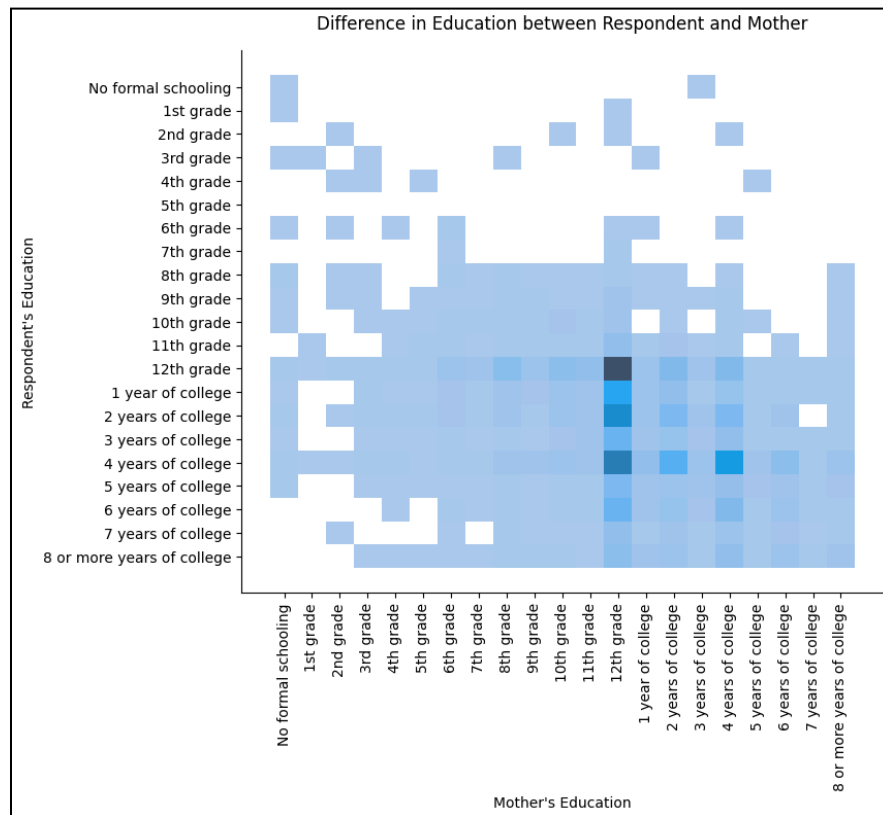
### Comparing 'degree' and 'madeg'



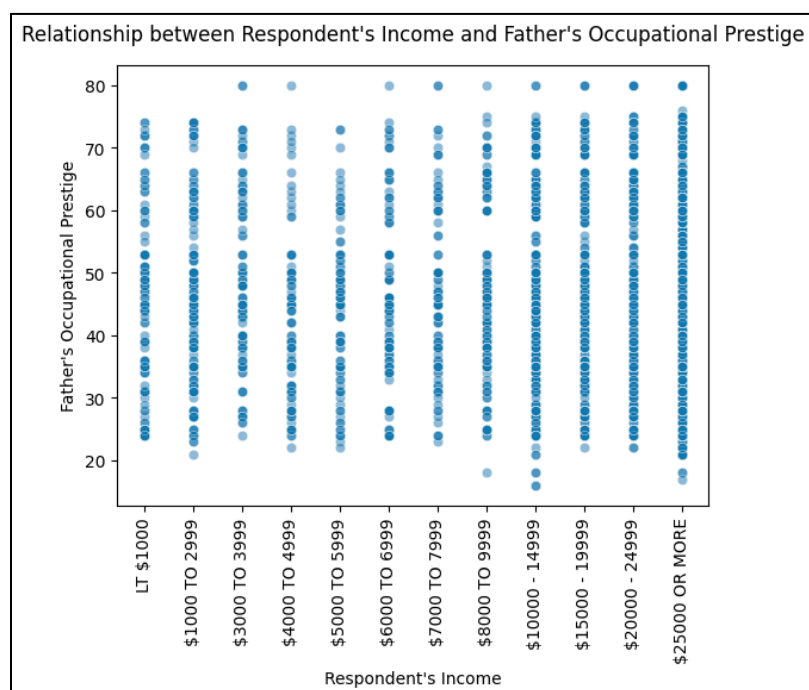
### Comparing 'paeduc' and 'educ'



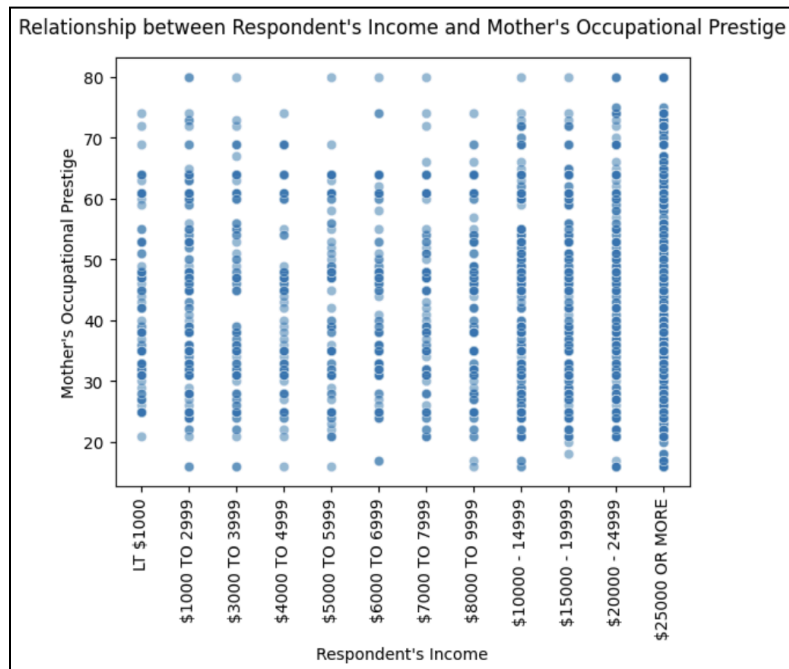
### Comparing 'maeduc' and 'educ'



### Comparing 'rincome' and 'papres10'



### Comparing 'rincome' and 'mapres10'



For the following proportions, respondents were split into three groups

- Both parents have occupational prestige over 50 (high)
- Neither parents have occupational prestige over 50 (low)
- Only one parent has occupational prestige over 50 (mixed)

*Proportion that has been unemployed in the last 10 years*

```
Proportion of unemployed for high prestige parents: 0.38493723849372385
Proportion of unemployed for low prestige parents: 0.35789473684210527
Proportion of unemployed for mixed prestige parents: 0.3657049979683056
```

*Proportion that are 'pretty well satisfied' with financial situation*

```
Proportion of strong satisfaction for respondents with high prestige parents: 0.34518828451882844
Proportion of strong satisfaction for respondents with low prestige parents: 0.2660954712362301
Proportion of strong satisfaction for respondents with mixed prestige parents: 0.3010971149939049
```

*Proportion that selected the highest income option*

```

Proportion of $25000+ income for respondents with high prestige parents: 0.698744769874477
Proportion of $25000+ income for respondents with low prestige parents: 0.5968176254589963
Proportion of $25000+ income for respondents with mixed prestige parents: 0.6607070296627388

```

**Findings**

Based on the analysis given by these plots and proportions, I don't see any significant evidence that could allude to the original stereotypes/reasoning being true. For the comparisons between prestige, both graphs have a correspondence value of about 0.17. Though that would imply a very slight positive correlation between the two, it's not strong enough to draw any conclusions (we'd need something close to 1 in order to do so). Another variable that could imply a privilege towards respondents with prestigious parents would be the final proportion above.

Approximately 70% of respondents with both parents of high occupational prestige have an income over \$25,000, while only 60% of respondents. While this divide at this value can definitely give some information that can help lead to a conclusion, higher ranges would be needed to properly assess the results and make conclusions. Our biggest divide would likely be the financial situation proportion, with there being an 8% difference between satisfaction for high and low prestige parent respondents.

However, no other plot/proportions give any hint of correlation to prove the original assumptions. Therefore, it'll be hard to make any real connection between our strongest proportion and the rest of our data. Our unemployment proportion had less than a 3% difference between high and low prestige parents, so I'd feel even less confident in making conclusions with these results. The rest of our plots seem to be scattered, and show no real pattern being shown (regardless of if we're comparing mother or father).