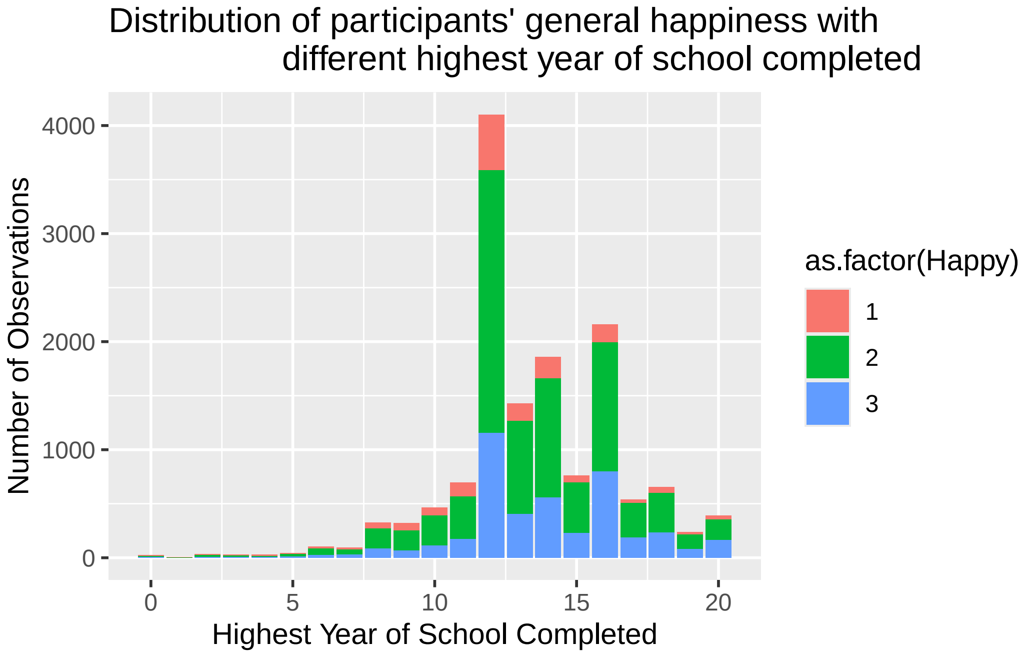
Project

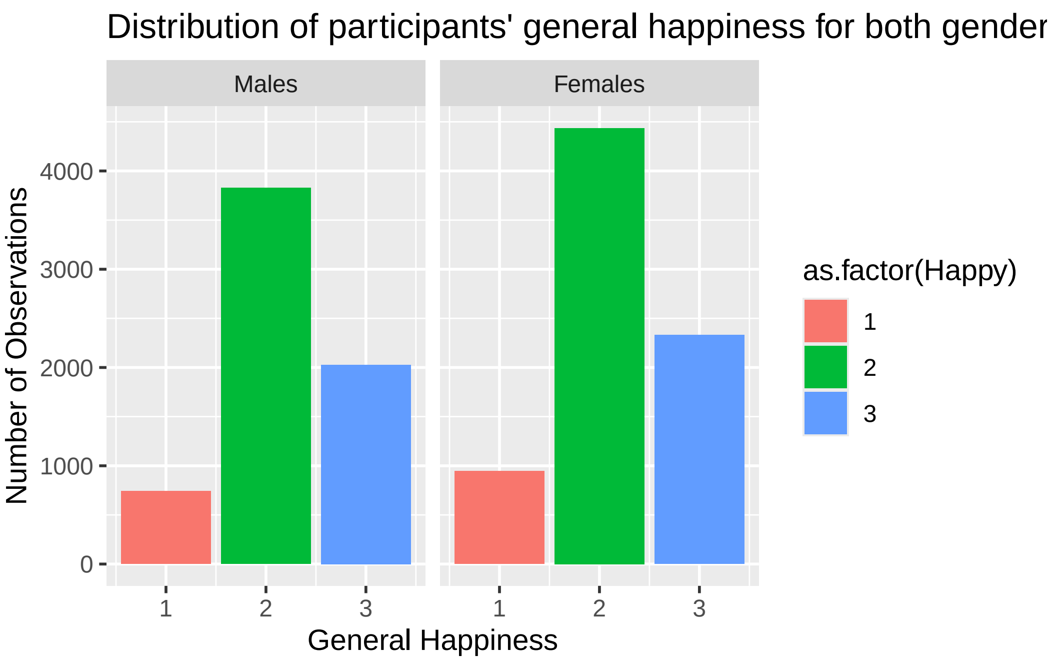
Tina Qian, Yanlin Li

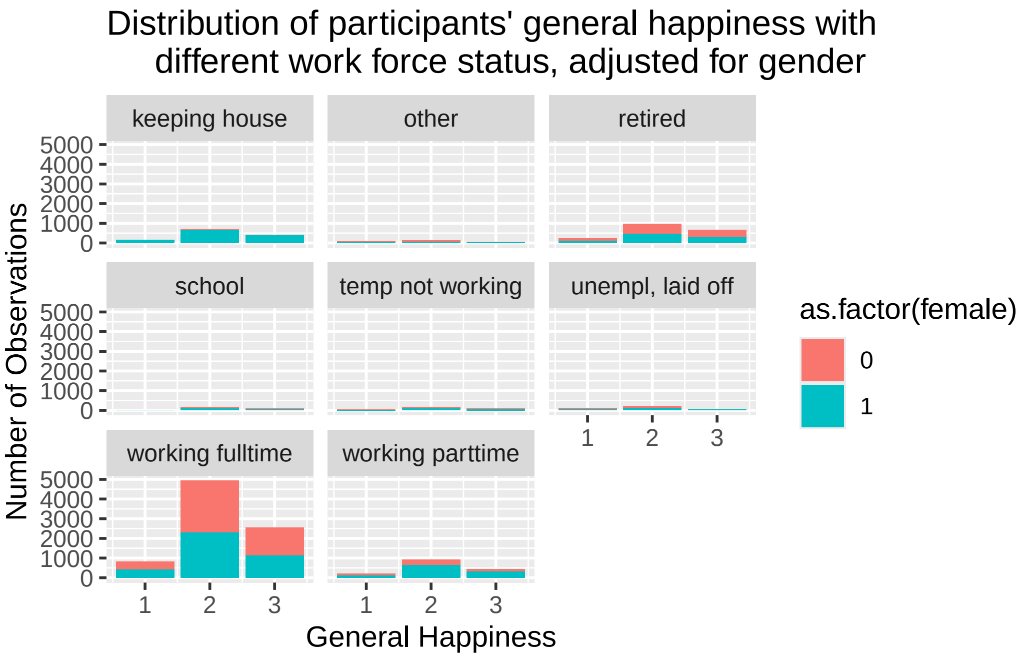
```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.5.0     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
-- Attaching packages -------------------------------------- tidymodels 1.1.1 --

v broom        1.0.5     v rsample      1.2.0
v dials        1.2.0     v tune         1.1.2
v infer        1.0.5     v workflows    1.1.3
v modeldata    1.2.0     v workflowsets 1.0.1
v parsnip      1.1.1     v yardstick    1.2.0
v recipes      1.0.8

-- Conflicts ----------------------------------------- tidymodels_conflicts() --
x scales::discard() masks purrr::discard()
x dplyr::filter()   masks stats::filter()
x recipes::fixed()  masks stringr::fixed()
x dplyr::lag()      masks stats::lag()
x yardstick::spec() masks readr::spec()
x recipes::step()   masks stats::step()
* Learn how to get started at https://www.tidymodels.org/start/

Rows: 17137 Columns: 34
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (9): workstat, divorce, widowed, reg16, income, region, attend, happy, ...
dbl (25): rownames, year, prestige, educ, babies, preteen, teens, tvhours, v...
```

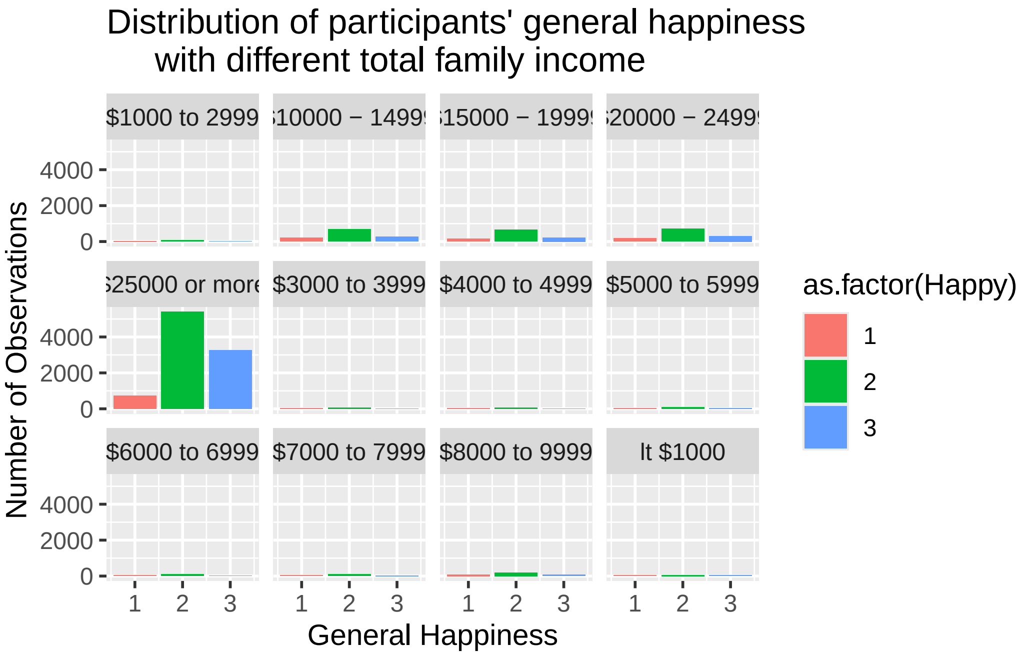
```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

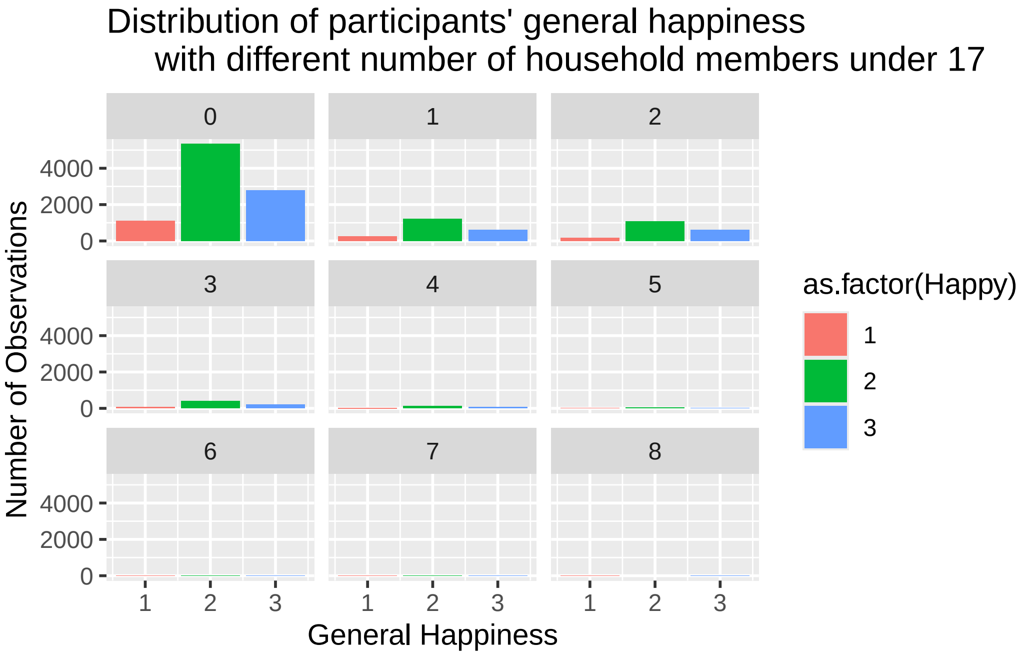
Data Exploratory Analysis




Distribution of participants' general happiness for both gender


Distribution of participants' general happiness with
different work force status, adjusted for gender





Concerning our response variable “happy,” we assigned numerical rankings of “1,” “2,” and “3” to its three categories—“not too happy,” “pretty happy,” and “very happy,” respectively—to facilitate analysis and model fitting.

Since we focus on the effect of socioeconomic status on happiness in our research question, we selected specific predictor variables: work force status (“workstat”), total family income (“income”), highest year of school completed (“educ”), gender (“female”), and the number of household members under 17 (“young_members”).

To create the predictor variable that records the number of household members under 17, which was not originally included in our dataset, we aggregated data from three existing variables: the number of babies, preteens, and teenagers in the household. We believe it’s more logical to consider these age groups collectively since each requires financial support and additional care from older members of the household, given that they are not yet fully independent.

Given the N/As in the “income” variable and the “young_members” variable, we used Multiple Imputation via Chained Equations (MICE) to perform multiple imputation to the data. We assume MAR is satisfied from our visual examination before.***

We generated visualizations to explore the relationships between the predictor variables and the response variable. Since most of our variables, including the response variable, are categorical, we opted for bar plots for all our exploratory data analyses. The first bar plot examines the distribution of observations across different levels of highest year of school completed in the sample, accounting for their overall happiness. The majority of participants in the sample have completed 12-16 years of education, roughly equivalent to middle school and a high school diploma. Across almost every education level, more individuals selected “pretty happy” than “very happy” or “not too happy,” possibly influenced by a tendency to prioritize these responses in general. The second plot analyzes the distribution of observations for males and females separately, considering their overall happiness. Overall, participants tend to choose “pretty happy” more frequently than “very happy” or “not too happy,” and there are more female participants in the dataset than male. The third plot examines the distribution of observations across different work force statuses, factoring in gender and overall happiness. Most participants in the dataset are employed full-time, with only a small proportion in school, temporarily unemployed, or not working. More female participants are homemakers or work part-time compared to male participants, while gender distribution is relatively even across other work force statuses. The overall pattern of happiness preference remains consistent across genders. The fourth plot illustrates the distribution of observations across different household income levels, considering overall happiness. The majority of participants have a household income above $10,000, especially those with incomes exceeding $25,000, who strongly favor “pretty happy” and “very happy” over “not too happy.” The general pattern of happiness preference persists across all income levels. The fifth plot displays the distribution of observations for participants with different numbers of young household members, adjusting for overall happiness. Most participants do not have anyone under 17 in their household, while those who do typically have one or two such members. The overall pattern of happiness preference remains consistent across these groups.

Testing Interaction & Assumptions

Among these predictors, we examined two possible interactions: the interaction between work force status and gender, and that between household income and the count of household members under 17. Given the societal expectation for women to bear children and the stereotype of women primarily responsible for childcare, we anticipate a stronger correlation between being female and working part-time. Additionally, we anticipate that households with higher incomes will generally be able to support more members under 17, necessitating greater economic assistance from other family members.

F test: (all coef associated with the interaction term are equal to 0)

Hypothesis test: H0: there is no interaction between working status and whether one is woman. Ha: there is an interaction between working status and whether one is woman.

```
Call:
lm(formula = Happy ~ as.factor(workstat) + prestige + educ +
    young_members + as.factor(income) + female, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4921 -0.2852 -0.1584  0.6806  1.4445

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          1.8252333  0.0583202  31.297  < 2e-16 ***
as.factor(workstat)other            -0.2342678  0.0402578  -5.819 6.04e-09 ***
as.factor(workstat)retired           0.0504177  0.0231948   2.174 0.029747 *
as.factor(workstat)school           -0.0105032  0.0402072  -0.261 0.793922
as.factor(workstat)temp not working -0.1285373  0.0395207  -3.252 0.001147 **
as.factor(workstat)unempl, laid off -0.3179799  0.0347838  -9.142  < 2e-16 ***
as.factor(workstat)working fulltime -0.0703945  0.0193490  -3.638 0.000276 ***
as.factor(workstat)working parttime -0.0635071  0.0231750  -2.740 0.006145 **
prestige                             0.0015089  0.0004417   3.416 0.000638 ***
educ                                 0.0101345  0.0021157   4.790 1.68e-06 ***
young_members                        0.0070600  0.0049212   1.435 0.151421
as.factor(income)$10000 - 14999      0.0841454  0.0541535   1.554 0.120247
as.factor(income)$15000 - 19999      0.0946902  0.0544857   1.738 0.082251 .
as.factor(income)$20000 - 24999      0.1336330  0.0541163   2.469 0.013547 *
as.factor(income)$25000 or more      0.2840411  0.0518941   5.473 4.49e-08 ***
as.factor(income)$3000 to 3999      -0.0129948  0.0729625  -0.178 0.858645
as.factor(income)$4000 to 4999      -0.0217046  0.0738324  -0.294 0.768784
as.factor(income)$5000 to 5999       0.1012660  0.0687914   1.472 0.141023
as.factor(income)$6000 to 6999      -0.0867617  0.0693227  -1.252 0.210750
as.factor(income)$7000 to 7999      -0.0512538  0.0673556  -0.761 0.446703
as.factor(income)$8000 to 9999       0.0095334  0.0602832   0.158 0.874346
as.factor(income)lt $1000            0.0336651  0.0719559   0.468 0.639894
female                              -0.0004968  0.0107011  -0.046 0.962974
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6065 on 14287 degrees of freedom
Multiple R-squared:  0.05251,   Adjusted R-squared:  0.05105
F-statistic: 35.99 on 22 and 14287 DF,  p-value: < 2.2e-16



Call:
lm(formula = Happy ~ as.factor(workstat) + prestige + educ +
    young_members + as.factor(income) + female + as.factor(workstat) *
    female, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5032 -0.2874 -0.1502  0.6749  1.4869

Coefficients:
                                             Estimate Std. Error t value
(Intercept)                                 1.7412459  0.0894188  19.473
as.factor(workstat)other                   -0.1580774  0.0874560  -1.808
as.factor(workstat)retired                  0.1469180  0.0732978   2.004
as.factor(workstat)school                   0.0322044  0.0896624   0.359
as.factor(workstat)temp not working        -0.0840885  0.0883513  -0.952
as.factor(workstat)unempl, laid off        -0.2880422  0.0811699  -3.549
as.factor(workstat)working fulltime         0.0299583  0.0712285   0.421
as.factor(workstat)working parttime        -0.0677263  0.0756881  -0.895
prestige                                    0.0015066  0.0004418   3.410
educ                                        0.0101790  0.0021160   4.811
young_members                               0.0050917  0.0049372   1.031
as.factor(income)$10000 - 14999             0.0865764  0.0541287   1.599
as.factor(income)$15000 - 19999             0.0968899  0.0544566   1.779
as.factor(income)$20000 - 24999             0.1331380  0.0540872   2.462
as.factor(income)$25000 or more             0.2818486  0.0518722   5.434
as.factor(income)$3000 to 3999             -0.0121795  0.0729260  -0.167
as.factor(income)$4000 to 4999             -0.0194333  0.0738241  -0.263
as.factor(income)$5000 to 5999              0.0961489  0.0687700   1.398
as.factor(income)$6000 to 6999             -0.0821911  0.0693002  -1.186
as.factor(income)$7000 to 7999             -0.0499575  0.0673389  -0.742
as.factor(income)$8000 to 9999              0.0098793  0.0602650   0.164
as.factor(income)lt $1000                   0.0401557  0.0719376   0.558
female                                      0.0909211  0.0726302   1.252
as.factor(workstat)other:female            -0.0761648  0.1020901  -0.746
as.factor(workstat)retired:female          -0.1177203  0.0778969  -1.511
as.factor(workstat)school:female           -0.0190918  0.1026223  -0.186
as.factor(workstat)temp not working:female -0.0167714  0.1016312  -0.165
as.factor(workstat)unempl, laid off:female  0.0301082  0.0941236   0.320
as.factor(workstat)working fulltime:female -0.1215027  0.0738657  -1.645
as.factor(workstat)working parttime:female  0.0390608  0.0798205   0.489
                                           Pr(>|t|)
(Intercept)                                 < 2e-16 ***
as.factor(workstat)other                   0.070704 .
as.factor(workstat)retired                 0.045046 *
as.factor(workstat)school                  0.719470
as.factor(workstat)temp not working        0.341239
as.factor(workstat)unempl, laid off        0.000388 ***
as.factor(workstat)working fulltime        0.674058
as.factor(workstat)working parttime        0.370905
prestige                                   0.000651 ***
educ                                       1.52e-06 ***
young_members                              0.302423
as.factor(income)$10000 - 14999            0.109742
as.factor(income)$15000 - 19999            0.075226 .
as.factor(income)$20000 - 24999            0.013846 *
as.factor(income)$25000 or more            5.62e-08 ***
as.factor(income)$3000 to 3999             0.867363
as.factor(income)$4000 to 4999             0.792371
as.factor(income)$5000 to 5999             0.162098
as.factor(income)$6000 to 6999             0.235636
as.factor(income)$7000 to 7999             0.458171
as.factor(income)$8000 to 9999             0.869788
as.factor(income)lt $1000                  0.576716
female                                     0.210650
as.factor(workstat)other:female            0.455647
as.factor(workstat)retired:female          0.130752
as.factor(workstat)school:female           0.852417
as.factor(workstat)temp not working:female 0.868929
as.factor(workstat)unempl, laid off:female 0.749065
as.factor(workstat)working fulltime:female 0.100009
as.factor(workstat)working parttime:female 0.624596
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.606 on 14280 degrees of freedom
Multiple R-squared:  0.05443,   Adjusted R-squared:  0.05251
F-statistic: 28.35 on 29 and 14280 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Model 1: Happy ~ as.factor(workstat) + prestige + educ + young_members +
    as.factor(income) + female
Model 2: Happy ~ as.factor(workstat) + prestige + educ + young_members +
    as.factor(income) + female + as.factor(workstat) * female
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  14287 5255.1
2  14280 5244.4  7    10.684 4.1559 0.0001407 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

interpretations:

The p value for interaction term between full-time working status and female is 0.0001, which is lower than a significance level of 0.05 and statistically significant. We have sufficient evidence to reject the null hypothesis, and thus the relationship between happiness and whether one is working full-time is depended on whether the respondent is female.

```
Call:
lm(formula = Happy ~ as.factor(workstat) + prestige + educ +
    young_members + as.factor(income) + female + +as.factor(workstat) *
    female + as.factor(income) * young_members, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4949 -0.2882 -0.1522  0.6742  1.4705

Coefficients:
                                                Estimate Std. Error t value
(Intercept)                                    1.7881651  0.0950471  18.813
as.factor(workstat)other                      -0.1695576  0.0875954  -1.936
as.factor(workstat)retired                     0.1387242  0.0734317   1.889
as.factor(workstat)school                      0.0218177  0.0897947   0.243
as.factor(workstat)temp not working           -0.0930247  0.0884772  -1.051
as.factor(workstat)unempl, laid off           -0.2968087  0.0813181  -3.650
as.factor(workstat)working fulltime            0.0222614  0.0713641   0.312
as.factor(workstat)working parttime           -0.0759437  0.0758132  -1.002
prestige                                       0.0015023  0.0004417   3.401
educ                                           0.0101622  0.0021162   4.802
young_members                                 -0.0524720  0.0462552  -1.134
as.factor(income)$10000 - 14999                0.0510833  0.0636973   0.802
as.factor(income)$15000 - 19999                0.0574295  0.0642329   0.894
as.factor(income)$20000 - 24999                0.1173130  0.0637335   1.841
as.factor(income)$25000 or more                0.2355338  0.0611926   3.849
as.factor(income)$3000 to 3999                -0.0567627  0.0859366  -0.661
as.factor(income)$4000 to 4999                -0.0062530  0.0856008  -0.073
as.factor(income)$5000 to 5999                 0.0756804  0.0793951   0.953
as.factor(income)$6000 to 6999                -0.0721486  0.0794574  -0.908
as.factor(income)$7000 to 7999                -0.0659425  0.0764985  -0.862
as.factor(income)$8000 to 9999                -0.0363617  0.0700780  -0.519
as.factor(income)lt $1000                     -0.0047674  0.0835263  -0.057
female                                         0.0856346  0.0727684   1.177
as.factor(workstat)other:female               -0.0683492  0.1022021  -0.669
as.factor(workstat)retired:female             -0.1152764  0.0780148  -1.478
as.factor(workstat)school:female              -0.0067301  0.1027312  -0.066
as.factor(workstat)temp not working:female    -0.0111511  0.1017338  -0.110
as.factor(workstat)unempl, laid off:female     0.0446995  0.0943323   0.474
as.factor(workstat)working fulltime:female    -0.1143895  0.0740069  -1.546
as.factor(workstat)working parttime:female     0.0457235  0.0799220   0.572
young_members:as.factor(income)$10000 - 14999  0.0508362  0.0492496   1.032
young_members:as.factor(income)$15000 - 19999  0.0577407  0.0499398   1.156
young_members:as.factor(income)$20000 - 24999  0.0194206  0.0488881   0.397
young_members:as.factor(income)$25000 or more  0.0671819  0.0465926   1.442
young_members:as.factor(income)$3000 to 3999   0.0652173  0.0707011   0.922
young_members:as.factor(income)$4000 to 4999  -0.0316060  0.0675854  -0.468
young_members:as.factor(income)$5000 to 5999   0.0181920  0.0668959   0.272
young_members:as.factor(income)$6000 to 6999  -0.0475456  0.0654595  -0.726
young_members:as.factor(income)$7000 to 7999  -0.0302088  0.0791735  -0.382
young_members:as.factor(income)$8000 to 9999   0.0725223  0.0565624   1.282
young_members:as.factor(income)lt $1000        0.0683617  0.0718340   0.952
                                              Pr(>|t|)
(Intercept)                                    < 2e-16 ***
as.factor(workstat)other                      0.052925 .
as.factor(workstat)retired                    0.058891 .
as.factor(workstat)school                     0.808030
as.factor(workstat)temp not working           0.293094
as.factor(workstat)unempl, laid off           0.000263 ***
as.factor(workstat)working fulltime           0.755090
as.factor(workstat)working parttime           0.316495
prestige                                      0.000673 ***
educ                                          1.59e-06 ***
young_members                                 0.256646
as.factor(income)$10000 - 14999               0.422584
as.factor(income)$15000 - 19999               0.371292
as.factor(income)$20000 - 24999               0.065689 .
as.factor(income)$25000 or more               0.000119 ***
as.factor(income)$3000 to 3999                0.508932
as.factor(income)$4000 to 4999                0.941769
as.factor(income)$5000 to 5999                0.340498
as.factor(income)$6000 to 6999                0.363885
as.factor(income)$7000 to 7999                0.388696
as.factor(income)$8000 to 9999                0.603856
as.factor(income)lt $1000                     0.954485
female                                        0.239291
as.factor(workstat)other:female               0.503656
as.factor(workstat)retired:female             0.139531
as.factor(workstat)school:female              0.947767
as.factor(workstat)temp not working:female    0.912720
as.factor(workstat)unempl, laid off:female    0.635613
as.factor(workstat)working fulltime:female    0.122209
as.factor(workstat)working parttime:female    0.567262
young_members:as.factor(income)$10000 - 14999 0.301988
young_members:as.factor(income)$15000 - 19999 0.247617
young_members:as.factor(income)$20000 - 24999 0.691192
young_members:as.factor(income)$25000 or more 0.149352
young_members:as.factor(income)$3000 to 3999  0.356316
young_members:as.factor(income)$4000 to 4999  0.640046
young_members:as.factor(income)$5000 to 5999  0.785668
young_members:as.factor(income)$6000 to 6999  0.467645
young_members:as.factor(income)$7000 to 7999  0.702800
young_members:as.factor(income)$8000 to 9999  0.199806
young_members:as.factor(income)lt $1000       0.341284
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6058 on 14269 degrees of freedom
Multiple R-squared:  0.05585,   Adjusted R-squared:  0.0532
F-statistic:  21.1 on 40 and 14269 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Model 1: Happy ~ as.factor(workstat) + prestige + educ + young_members + 
    as.factor(income) + female + as.factor(workstat) * female
Model 2: Happy ~ as.factor(workstat) + prestige + educ + young_members + 
    as.factor(income) + female + +as.factor(workstat) * female + 
    as.factor(income) * young_members
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)  
1  14280 5244.4                              
2  14269 5236.6 11    7.8535 1.9454 0.02954 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The f statistic value 0.03 for interaction term between household income (dollars per year) and number of children (young_members) across all levels are lower than 0.05, which is not statistically significant. We have sufficient evidence to reject the null hypothesis, and thus the relationship between score of general happiness and household income (dollors per year) is depended on the number of children in the household.

Ordinal Regression

In our study, we contemplated employing both multinomial regression and ordinal regression models for the categorical response variable. Ultimately, we opted for the latter. The decision was straightforward, as the response variable "happy" exhibits an ordinal nature, with discernible gradations from "not too happy" to "pretty happy" to "very happy."

After assessing and confirming all the above conditions, we went into fitting the ordinal regression model.

```r
library(MASS)
```

```

Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select
```

```r
m3 <- polr(as.factor(Happy) ~ as.factor(workstat) + educ + young_members + as.factor(incom
           data = Happiness, Hess = T)
```

```r
summary(m3)
```

```
Call:
polr(formula = as.factor(Happy) ~ as.factor(workstat) + educ + 
    young_members + as.factor(income) + female, data = Happiness, 
    Hess = T)

Coefficients:
                                        Value Std. Error t value
as.factor(workstat)other            -0.773263   0.135589 -5.7030
as.factor(workstat)retired           0.193274   0.075922  2.5457
as.factor(workstat)school           -0.064594   0.131860 -0.4899
as.factor(workstat)temp not working -0.404080   0.131059 -3.0832
as.factor(workstat)unempl, laid off -1.061233   0.115523 -9.1863
as.factor(workstat)working fulltime -0.240466   0.063385 -3.7938
as.factor(workstat)working parttime -0.227947   0.075832 -3.0060
educ                                 0.044899   0.006142  7.3098
young_members                        0.021251   0.016034  1.3253
as.factor(income)$10000 - 14999      0.279715   0.181923  1.5375
as.factor(income)$15000 - 19999      0.313509   0.182711  1.7159
as.factor(income)$20000 - 24999      0.455530   0.181669  2.5075
as.factor(income)$25000 or more      0.952520   0.174331  5.4639
as.factor(income)$3000 to 3999      -0.063785   0.246791 -0.2585
as.factor(income)$4000 to 4999      -0.095107   0.248295 -0.3830
as.factor(income)$5000 to 5999       0.332730   0.230074  1.4462
as.factor(income)$6000 to 6999      -0.313717   0.231487 -1.3552
as.factor(income)$7000 to 7999      -0.174160   0.226302 -0.7696
as.factor(income)$8000 to 9999       0.028142   0.203049  0.1386
as.factor(income)lt $1000            0.079274   0.247880  0.3198
female                               0.006769   0.034898  0.1939

Intercepts:
    Value   Std. Error t value
1|2 -0.9535  0.1948    -4.8953
2|3  2.0130  0.1955    10.2967

Residual Deviance: 25906.53 
AIC: 25952.53 
```

```r
exp(coef(m3))
```

```
           as.factor(workstat)other          as.factor(workstat)retired
                          0.4615045                           1.2132158
          as.factor(workstat)school as.factor(workstat)temp not working
                          0.9374476                           0.6675908
as.factor(workstat)unempl, laid off as.factor(workstat)working fulltime
                          0.3460287                           0.7862611
as.factor(workstat)working parttime                                educ
                          0.7961665                           1.0459224
                      young_members     as.factor(income)$10000 - 14999
                          1.0214784                           1.3227522
    as.factor(income)$15000 - 19999     as.factor(income)$20000 - 24999
                          1.3682176                           1.5770087
    as.factor(income)$25000 or more      as.factor(income)$3000 to 3999
                          2.5922347                           0.9382064
     as.factor(income)$4000 to 4999      as.factor(income)$5000 to 5999
                          0.9092753                           1.3947708
     as.factor(income)$6000 to 6999      as.factor(income)$7000 to 7999
                          0.7307258                           0.8401622
     as.factor(income)$8000 to 9999           as.factor(income)lt $1000
                          1.0285413                           1.0825009
                             female
                          1.0067915
```

workstat: According to the ordinal regression model, a person who retired is predicted to have 1.213 times the odds of being in the next higher score of happiness category compared to a person who keeps house, while adjusting for years of education, household income, the number of children they have, and whether being a woman. In the work force status categories, people with all other status are predicted to have lower possibilities of being in the next higher score of happiness category compared to those who keep houses, while adjusting for years of education, household income, the number of children in their household, and whether being a woman.

educ: a person who has one more year of education is predicted to have 1.05 times the odds of being in the next higher score of happiness category compared to a person who has less years of education, while adjusting for work force status, household income, the number of children in their household, and whether being a woman.

young_members: a person who has one more number of children in their household is predicted to have 1.021 times the odds of being in the next higher score of happiness category compared to a person who has less numbers of children in their household, while adjusting for work force status, household income, years of education, and whether being a woman.

female: a person who is a woman is predicted to have 1.007 times the odds of being in the next higher score of happiness category compared to a person who is a man, while adjusting for years of education, household income, and number of children in their household.