Henry Duke

2/17/2025

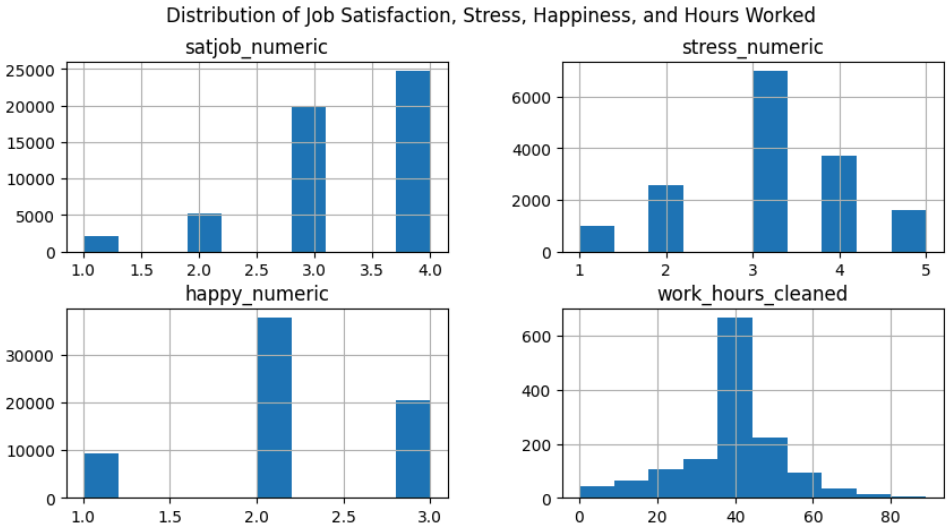
For this lab, I wanted to focus on fields that relate to jobs, satisfaction in those jobs, work life balance, etc. I feel it is a fitting subject for the current time period in my life. Being months away from finishing undergrad, it would be valuable to me to see real, well recorded data on what Americans feel about their work life. The first variable is SATJOB. The description reads, “On the whole, how satisfied are you with the work you do -- would you say you are very satisfied, moderately satisfied, a little dissatisfied, or very dissatisfied?” One of my fears is ending up in a job or career that I really can’t stand. Seeing this data would provide me with some really good insights. The next group of variables would be OBEY, POPULAR, THNKSELF, WORKHARD, HELPOTH. This group of variables is asked in a form of questions along the lines of the user ranking the variables in order of importance for a child to learn. In order, these variables are defined as “to obey”, “to be well-liked or popular”, “to think for himself or herself”, “to work hard”, and “to help others when they need help.” I find these variables interesting and want to see how the importance of these values change and relate to occupations. The next variable is STRESS, “How often do you find your work stressful?” WKVSFAM, “How often do you find your work stressful?” hrs2, “ IF WITH A JOB, BUT NOT AT WORK: How many hours a week do you usually work, at all jobs?” HAPPY, “Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?” The depth of the GSS data set is impressive and somewhat daunting. I feel that the combination of these variables will provide interesting insights that I am excited to uncover.

To first gain any insights from this data, I first had to do a fair amount of data wrangling. To pull the insights, I had to do a lot of dummy variable encoding. In addition, once the variables were encoded, I removed a lot of variables that did not provide any insights. These include entries like “no answer” or “inapplicable.” Once I had the data cleaned, I underwent some EDA to pull some insights from the data. I pulled the medians, means, and distributions of the job satisfaction, job stress, happiness, and work hours(Figures 1 & 2). In terms of the medians, for job satisfaction it was moderately satisfied. Job stress was sometimes. Happiness was pretty happy. Hours worked was 40 hours. In terms of means: 3.29, 3.15, 2.16, and 39.21, in the same order as the medians. Job satisfaction and job stress are on a 1-4 scale and happiness is on 1-3 scale. The distributions seem to all follow a somewhat normal distribution, besides job satisfaction. Job satisfaction seems to have a left skew. For my next plot(figure 3), I wanted to see the correlation coefficients between the four aforementioned variables. Expecting some high values, a lot of the correlation was weak. We saw that there is always no correlation between hours worked and job satisfaction. Of all the relationships, happiness and job satisfaction were the most correlated. Lastly, I analyzed the multi-variable question, asking the respondents what are the most important values to teach a child. There are definitely some evident trends. To think for him or herself seems to be the leading most important value. To work hard and help others seems to be close for number 2 and 3. To obey leads at 4th and to be popular leads at least important. I think this lab was great practice at creating different plots and doing some exploratory analysis on some real data. If I were to add something, I would analyze how these trends changed over time.

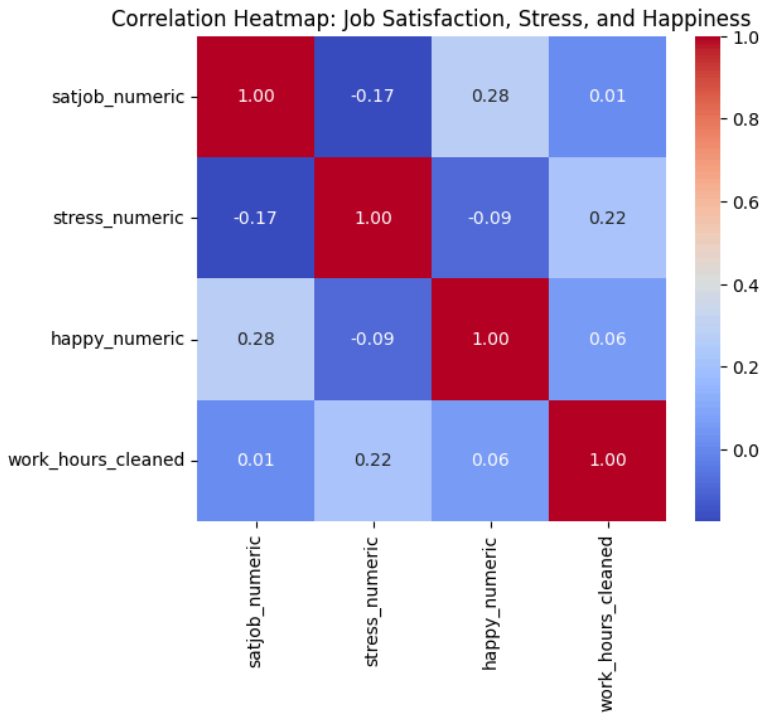
1.

	satjob_numeric	stress_numeric	happy_numeric	work_hours_cleaned
count	51887.000000	15879.000000	67588.000000	1401.000000
mean	3.293812	3.149569	2.162677	39.206281
std	0.809827	1.016608	0.643490	13.549101
min	1.000000	1.000000	1.000000	0.000000
25%	3.000000	3.000000	2.000000	35.000000
50%	3.000000	3.000000	2.000000	40.000000
75%	4.000000	4.000000	3.000000	45.000000
max	4.000000	5.000000	3.000000	89.000000

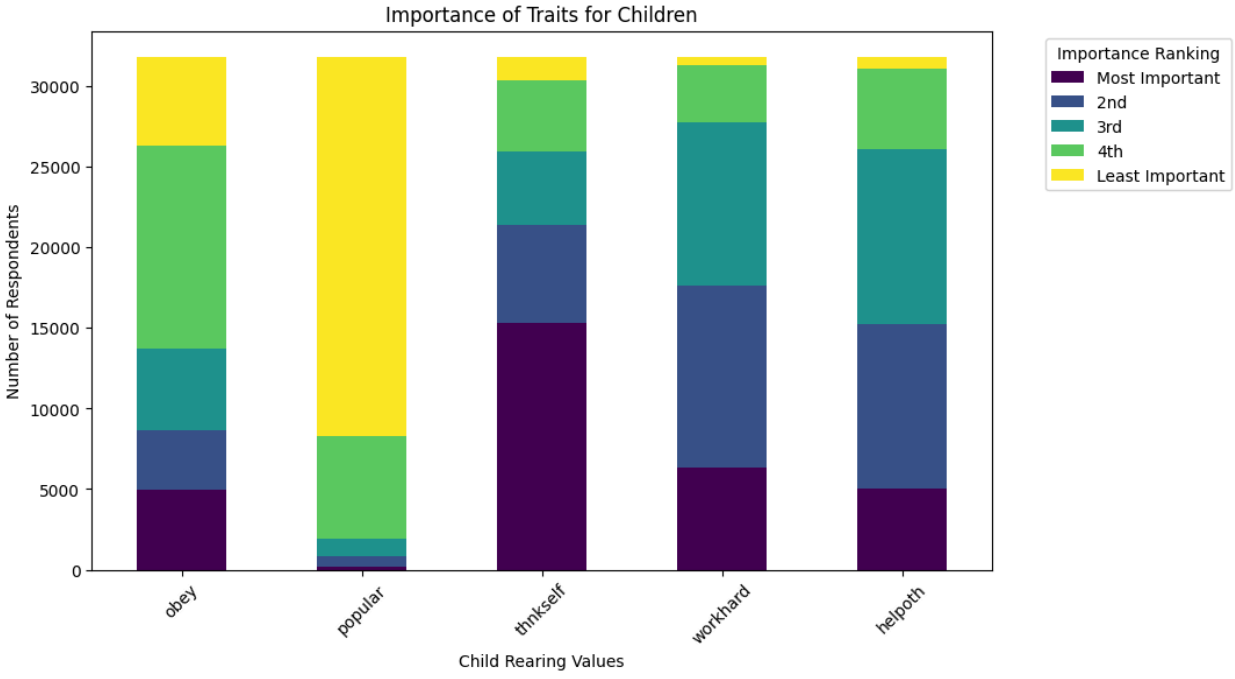
2.



3.


Correlation Heatmap: Job Satisfaction, Stress, and Happiness
satjob_numeric
stress_numeric
happy_numeric
work_hours_cleaned
1.00
-0.17
0.28
0.01
-0.17
1.00
-0.09
0.22
0.28
-0.09
1.00
0.06
0.01
0.22
0.06
1.00
1.0
0.8
0.6
0.4
0.2
0.0

4.


Importance of Traits for Children
Number of Respondents
0
5000
10000
15000
20000
25000
30000
obey
popular
thnkself
workhard
helpoth
Child Rearing Values
Importance Ranking
Most Important
2nd
3rd
4th
Least Important