CMSC320 HW4

Vincent Strappelli

October 27, 2023

1. Data Cleaning

- Issue 1: Long column titles
    - One issue I noticed was that the names for these titles were the questions themselves, which were a few sentences long. This made my data very ugly and difficult to manage, so I renamed them to Q1, Q2, etc.
- Issue 2: Missing Ages
    - There were a few entries with missing ages in Max's dataset, so I used the "fillna" function and replaced missing values with the mode age.
- Issue 3: Missing rows
    - In both Max and Fardina's datasets, there were several rows/entries that were all or mostly blank. I decided to drop these rows because they do not contribute anything to my analyses and could possible skew my results. If someone responded to only a few questions but skipped the rest, it is most likely that they did not take the survey seriously and those responses could skew the results.
- Issue 4: Gender typo
    - In Fardina's dataset, a few respondents entered "Famale" instead of "Female" for their gender. I used the "replace" function to fix this.

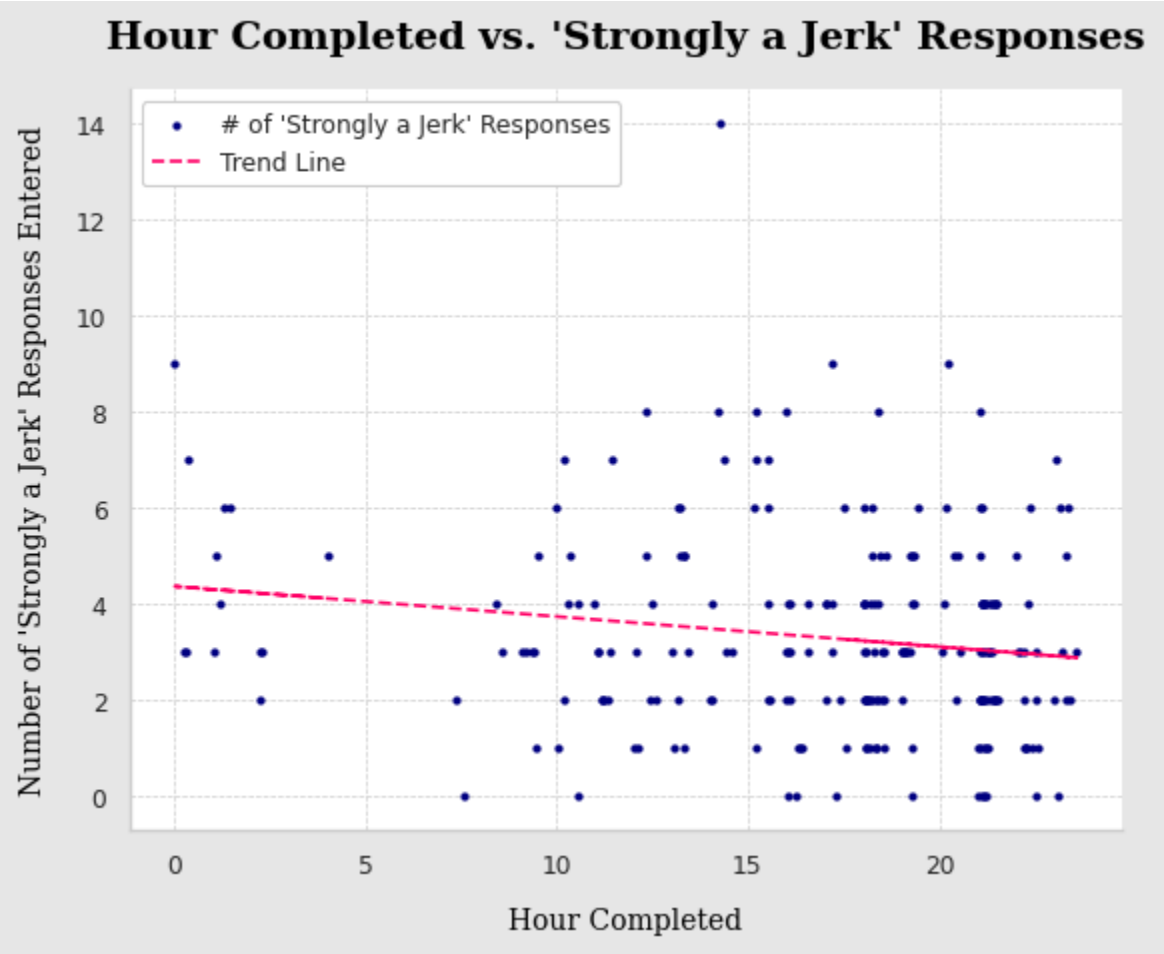
2. Questions

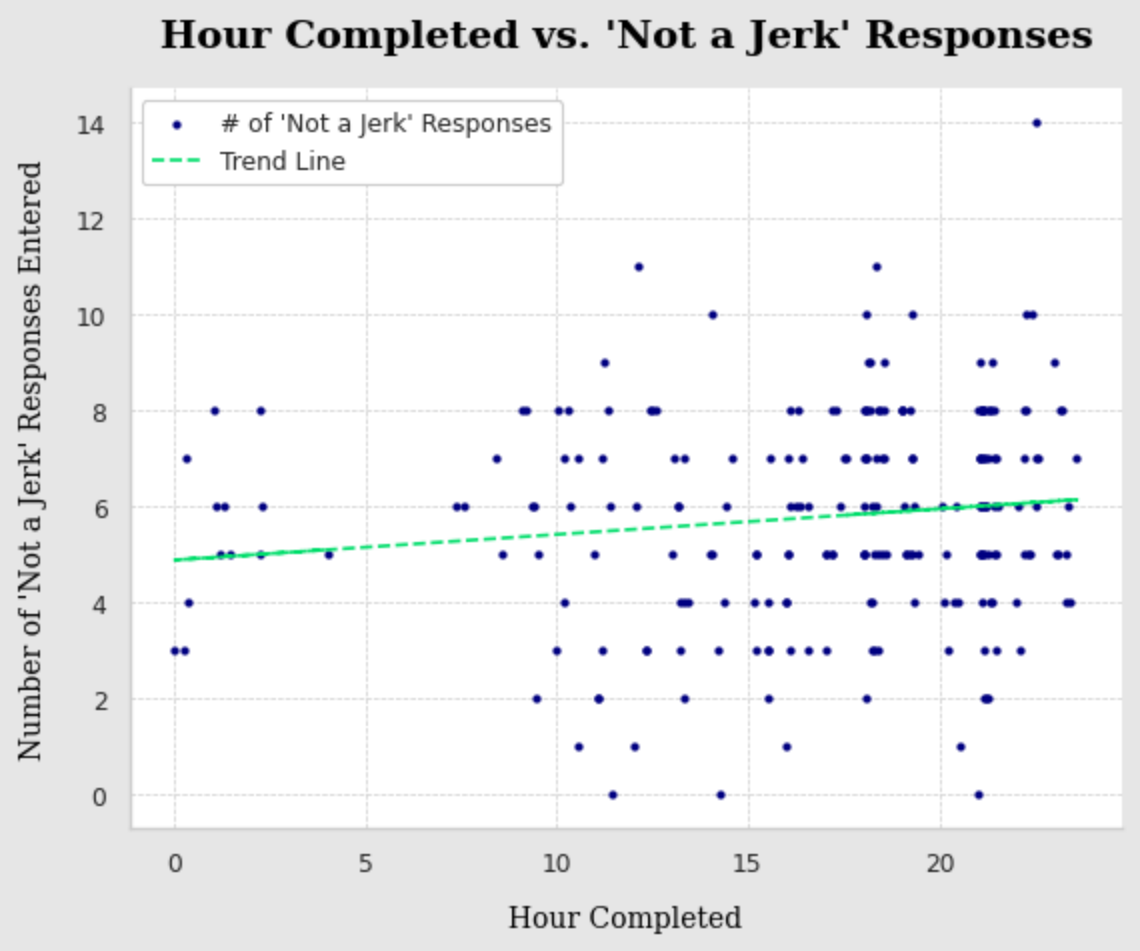
1. Do men tend to side with other men in the situations described in the questions? Do women tend to side with other women in these same situations?
    - I was interested in this question because it seems like men tend to side with men and women tend to side with women in most aspects of life, so I wanted to see if the same thing applied within this dataset.
    - In order to answer this question, I first found a few questions from the survey in which there was a clear and significant conflict between a man and a woman. The questions I chose were Question 1, Question 7, and Question 13. Then, for each of these questions, I created a table comparing gender and the answers to these questions. I dropped the "mildly a jerk" responses because they are neutral and make the table harder to work with for testing. I then performed a chi-squared test on each of these tables and observed the p-value returned from them. The chi-squared tests were used to determine the level of independence between gender and the respondents' answers to the given question. I used chi-squared tests in particular because the factors being compared are categorical rather than numerical.
    - From the testing I performed, there does not appear to be any strong evidence that suggests men tend to side with other men and women tend to side with

other women. The reason I think this is because each chi-squared test I performed returned a large p-value, averaging about 0.5-0.6. This indicates that there is not nearly enough evidence to prove there exists a significant correlation between gender and the answers to these questions. Furthermore, it is simple to tell just by looking at the tables that women and men alike answered very similarly.

2. Which has a greater impact one one's views on sharing money, political affiliation or level of spirituality?
    - I was interested in this question because both of these factors seem like they would influence someone's views on money and sharing money.
    - My investigation process was very similar to the previous question, but with more tests. In order to answer this question, I first chose a few questions (Q3, Q4, and Q10) that had to do with money and sharing money. Then, for each of these questions, I created a table comparing political affiliation and the answers to these questions and a table comparing spirituality and the answers to these questions. I dropped the "mildly a jerk" responses because they are neutral and make the table harder to work with for testing. I then performed a chi-squared test on each of these tables and observed the p-value returned from them. The chi-squared tests were used to determine the level of independence between gender and the respondents' answers to the given question. I used chi-squared tests in particular because the factors being compared are categorical rather than numerical. I averaged the p-values for the political affiliation tests and the spirituality tests.
    - From the testing I performed, I found that spirituality level has a greater impact on one's views on money. I believe this because the average p-value for the political affiliation tests was greater than the average for the spirituality tests. This is a bit surprising to me because politics and finances are more related than spirituality and finances.
3. Does the time of day impact how compassionate/forgiving respondents are?
    - I was interested in this question because I often feel like the time of day impacts how nice I am to those around me, so I wanted to see if the data reflected this.
    - In order to answer this question, I first created four new columns. Three columns (Strongly a Jerk Count, Mildly a Jerk Count, and Not a Jerk Count) stored the number of each response for each respondent. The fourth column stored the hour that the response was submitted. I did this because the date-time format is not able to be graphed. I then used this updated data frame to plot two scatter plots: hour completed vs. 'Strongly a Jerk' responses and hour completed vs. 'Not a Jerk' responses. I also plotted a trend line to see if there were any trends or patterns in the graph.
    - Based on my plots, I would say that the time of day does impact how compassionate/forgiving respondents are. I believe this because the trend lines of both graphs indicate that respondents earlier in the day gave more 'Strongly a Jerk' responses and less 'Not a Jerk' responses. Although this trend line is not

steep, it is noticeable and worth considering. However, one problem that may have impacted this result is that, unsurprisingly, the vast majority of respondents answered between 10am and midnight. These scatter plots can be seen below:




Hour Completed vs. 'Not a Jerk' Responses
of 'Not a Jerk' Responses
Trend Line
Number of 'Not a Jerk' Responses Entered
Hour Completed
0
2
4
6
8
10
12
14
0
5
10
15
20