

Speed-Dating Analysis

Edward Yao

October 16, 2017

Introduction

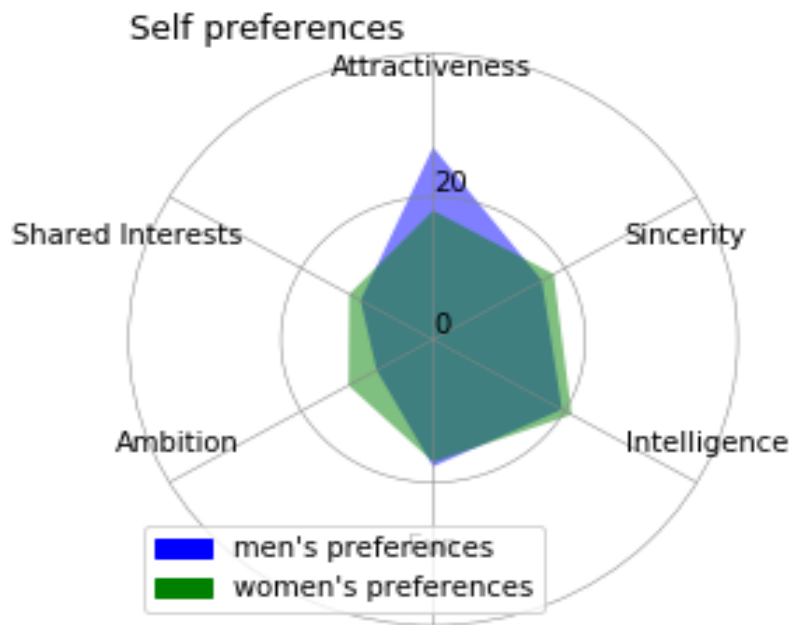
The data used in this analysis is taken from a Kaggle dataset, which originates from a paper by Fisman and Iyengar at Columbia university exploring gender preferences in dating. The dataset was collected by exposing different college students to different "speed-dates" lasting 4 minutes long. Both candidates from each date were required to fill a survey before starting the dates, complete a survey during the date, and respond to surveys couple of times after the day of their dates as well. The original authors of this paper focused along several dimensions to explain whether candidates were able to achieve second dates: commonality shared between potential dates in their race, geographical background, job, as well as SAT scores, income, and the relative levels of different traits.

In my analysis, I examined the data moreso along the dimension of desired and undesirable traits, as indicated by the participants in the study themselves. Rather than looking for particular factors that traditional theories might predict play a large role on how successful speed-dating occurs, I was more interested in the self-perceptions among the candidates insofar as how they characterized their own criteria in what factors determined whether they would submit a request for a second date with different candidates.

Analysis

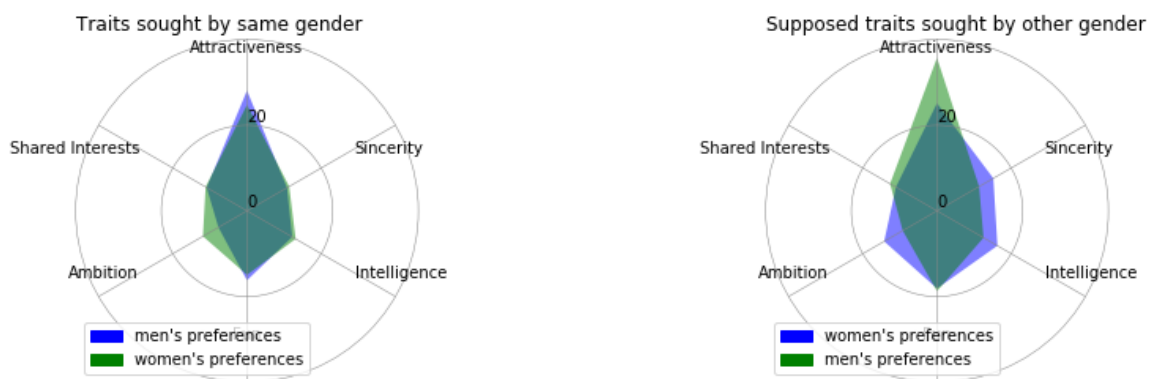
As a disclaimer to any of the results presented here, keep in mind that there is a very high threshold that results must be passed to draw causal inferences from data, and the analysis presented below in a semi-formal manner will certainly not be rigorous enough to be able to draw such conclusions. Nevertheless, I hope it will be thought-provoking and raise some questions that are worth exploring more in-depth.

The first thing other than a cursory glance over the dataset was how I could measure candidate's assessments of their own dating preferences throughout this study. The most convenient way seemed to be via the results of the pre-date surveys that candidates filled out, indicating the traits they sought from the other party. They assigned relative importance to six separate categories in the pre-date survey, which consisted of intelligence, attractiveness, sense of fun, ambition, sincerity, and level of common interests. Graphing the relative preferences split by gender, there were several differences.



As we can see from the above diagram, it seems like men self-reported placing much more emphasis on the attractiveness of the opposite party. In contrast, the self-reported preferences of women seem to be much more balanced across the 6 traits. Note that for these plots, the axes are labelled by percentages weight important for each trait, normalized by 100 (i.e. the weights were made to sum to 100).

Next, I also graphed what each candidate rated their own gender as desiring and the opposite gender as desiring.



The above graph reveals some interesting conclusion. When asked what their own gender preferred in the opposite sex, participants both heavily valued attractiveness as the most important feature, while fun and ambition seem to be the second most perceived categories for women perceived other women.

On the other hand, when asked what the other gender was likely to prefer, women reported men as heavily weighting attractiveness and less so looking for intelligence and ambition compared to the self-reported values by men. On the other hand, men also thought that women place increased importance on sincerity and ambition compared to the actual self-reported values by women. For more analysis of theory related to these traits, refer to Fisman and Iyengar's original 2006 paper that discusses psychological theories explaining several of these patterns by gender. However, I will not attempt to further explain these patterns here as doing so would take another paper by itself to flesh out the complications in such characterizations of desirable dating behavior.

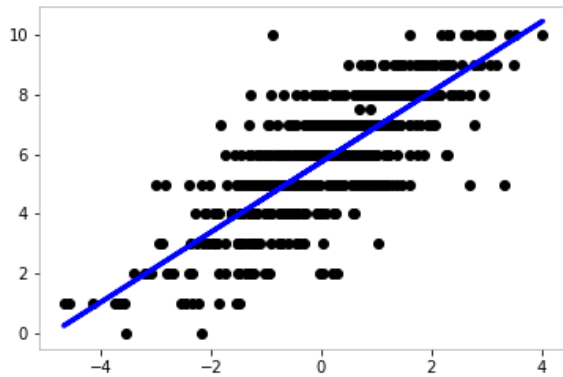
From random subsamples of the whole dataset, I examine scatter plots of the variable 'int_corr' of a particular other party and whether candidates preferred a second date with the corresponding other party, omitted here. 'int_corr' serves as measure of how similar the two party's interests were, normalized to be between -1 and 1. In both the following regressions, note that I set apart 20% of the data set to use as a testing set and trained with the other 80% of samples. Training a logistic regression on 'int_corr' and whether the candidate was interested in the other party or not, separately on both genders, the testing error was large. The regressions for men and women in this case were able only to correctly classify 64 and 53 percent of the data points in the testing set, respectively, indicating weak explanatory power of common interests in explaining interests. What, then, might help us examine how people select candidates they like during the dating process?

I decided that perhaps the combination of traits that we examined from candidates before the dates might be more useful in predicting relationships between candidates during the date. Training a plain logistic regression with each of the 6 features mentioned before as the dependent variables and the binary second-date feature as the independent variable, I found a stronger explanatory power for requested second dates with these variables, around 73% accuracy for both genders this time. Examining the coefficients, I found that the coefficients for attractiveness and fun were the two largest coefficients predicted by the logistic regression, which agrees with our earlier plots from the pre-dating surveys. It seems, then, that the combination of desirable features serves as a decent predictor of what both genders look for in dating partners.

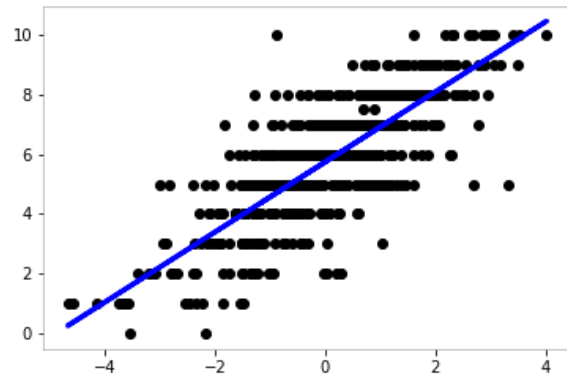
How does this relate to candidates own assessments of what they desired, though? Did candidates indeed stick to what they thought was important during the decision process, or were those decision rationale thrown into the wind during during the dates? To find these out, I created a new metric to measure a weighted combination of each other party's attractiveness based on the pre-reported desires of each person. This score metric was calculated by finding the sum of how much the opposite party deviated from the median value for that particular trait multiplied by how important the candidate previously said that trait was to them, then normalized. In this way, a positive number for this value would indicate a desirable other party based on what each candidate said in the pre-dating survey, and a negative number for undesirable candidates.

To measure the explanatory power of this metric, I chose to regress it two separate measures - the same 'dec_o', indicating whether a candidate was flagged for a second date by their partner, as well as 'like_o', a measure on a scale of 0-10 of how likeable they were by their partner. To make sure that both of these metrics hold some weight, I regressed the decision on the likability metric with a logistic function, and found that for both genders, the likeability factor was able to explain just above 70% of decisions for second dates, showing at least that they are correlated. The weighted score metric when regressed on the decision binary variable was able to explain around 73% of decisions

as well, a bit higher than the likeability metric, indicating a fairly strong correlation between what candidates self-reported as being important and which candidates they wanted a second date with, for both genders. I also ran a linear regression of the likability score itself on the score metric, from which a strong positive coefficient on the score resulted, found in both genders. The regression for women is on the left, and for men on the right.



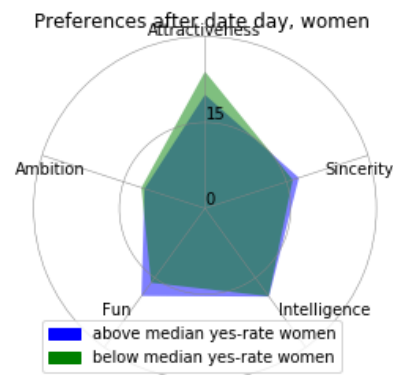
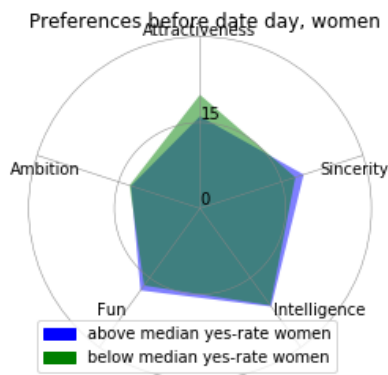
Likeability vs Score, Women



Likeability vs Score, Men

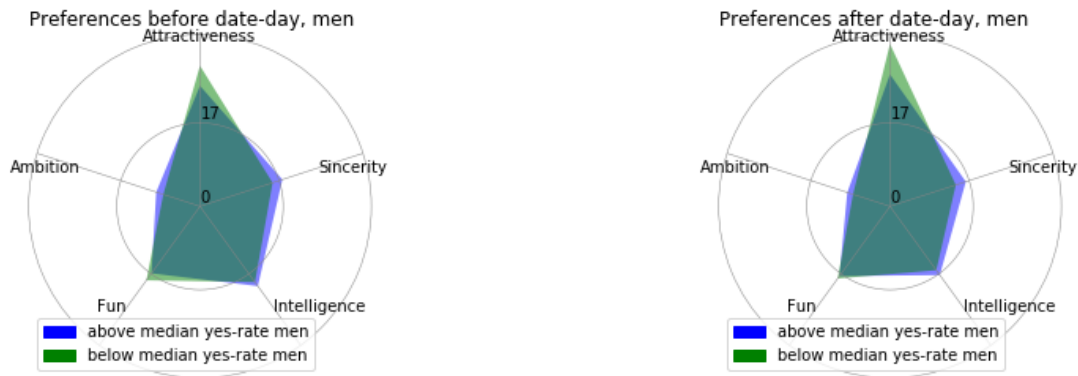
From the above regressions, it seems that candidates did indeed follow their self-reported criteria in selecting potential candidates for a second date and for reporting overall likeability of corresponding opposing candidates.

Using data from after the date, I then compared the top half of people to the bottom half of their gender in terms of percent second dates requested of them. Below are the plots of how the relative importance shifted for those two groups of the same women before and after the day of dating.



What we can see from these two graphs is that women who were more successful at garnering second date requests from men on that date day ended up valuing fun comparatively more than they did before. At the same time, those that were less successful at garnering more interest placed an increased amount of importance on attractiveness and comparatively less importance on fun. Of course, we should keep in mind that the above plots are far from being able to establish causality between importance of attractiveness and being able to get a second date, but nevertheless there seems to be a correlation that seems to be worth investigating in greater details elsewhere. The same plots for men are plotted below, and again we see that the candidates receiving less requests for second

dates also ended up placing more importance on attractiveness a day after the date than before the date. Interestingly, attractiveness also grew in importance for candidates in the upper-half of the second-date receiving men, but not as much as it had for those below median. Again, we see that compared to before, the men who received more requests for second dates placed comparatively more importance on fun than they had before the date after the date.



All in all, it seems like across all of the test subjects, the importance of attractiveness grew after the dates, but it grew much more for candidates receiving less second dates, whereas the importance of fun grew much more for those candidates who received more requests for second dates.