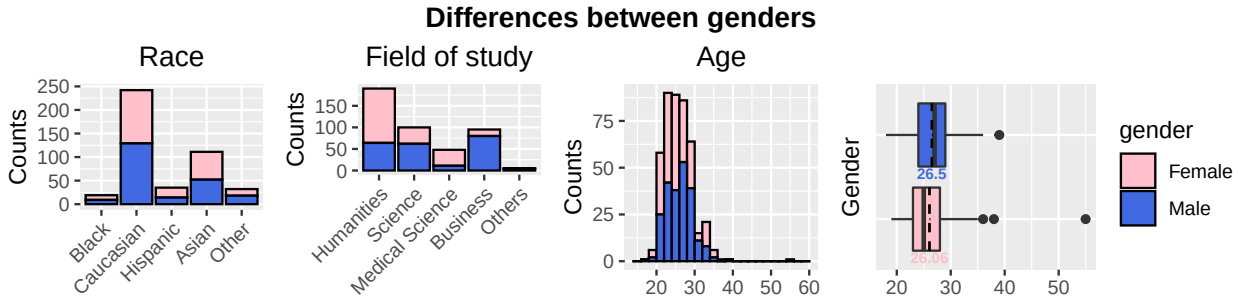
Do not judge a book by its cover... but a person maybe!

Evidence from a speed dating dataset reveals how to bag yourself a match

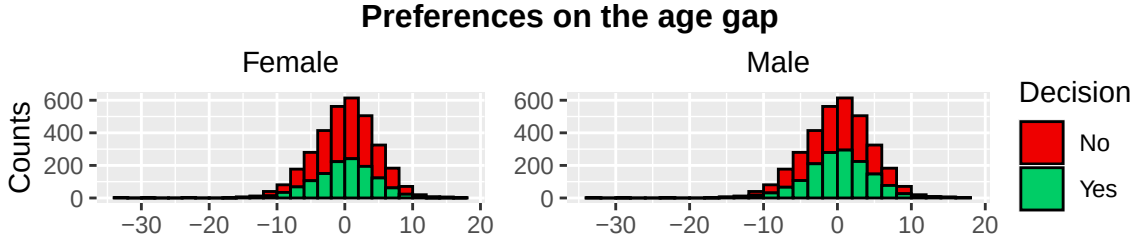
Everyone wants to find *the one*. However, approaching an attractive stranger at the bar seems so frightening (and somehow old-fashioned). Nowadays, indeed, the quest for love is pursued online: Tinder, Hinge, and other dating apps have become a sort of norm to meet potential partners. Before the arrival of smartphones, a good compromise was *Speed Dating*: started as a programme for Jewish singles in Los Angeles, it soon spread all over the world as a new way for single people to meet each other. The format is simple: a group of singles gathers at a cafe or similar venue. Armed with a scorecard and their sparkling personality, they are paired up and their first date begins. After four minutes of conversation, a bell rings and the men proceed to the next lady. Participants mark on the scorecard whether they would have an interest in meeting their date again. If there is mutual interest, the organizers will provide the couple with the contact information. From that point on, everything is up to you. The main goal is therefore to get a match. The question is: what makes you click with a person in a few minutes? We would like to have a better insight on how individual choices among random participants are made. To give an answer, we will explore the data behind the paper "Gender differences in mate selection: evidence from a speed dating experiment". This data collects the surveys filled out by the participants in Speed Dating events held at Columbia University from 2002 to 2004. Our goal is to select the main qualities we look for when selecting a partner.

Exploratory analysis of the data

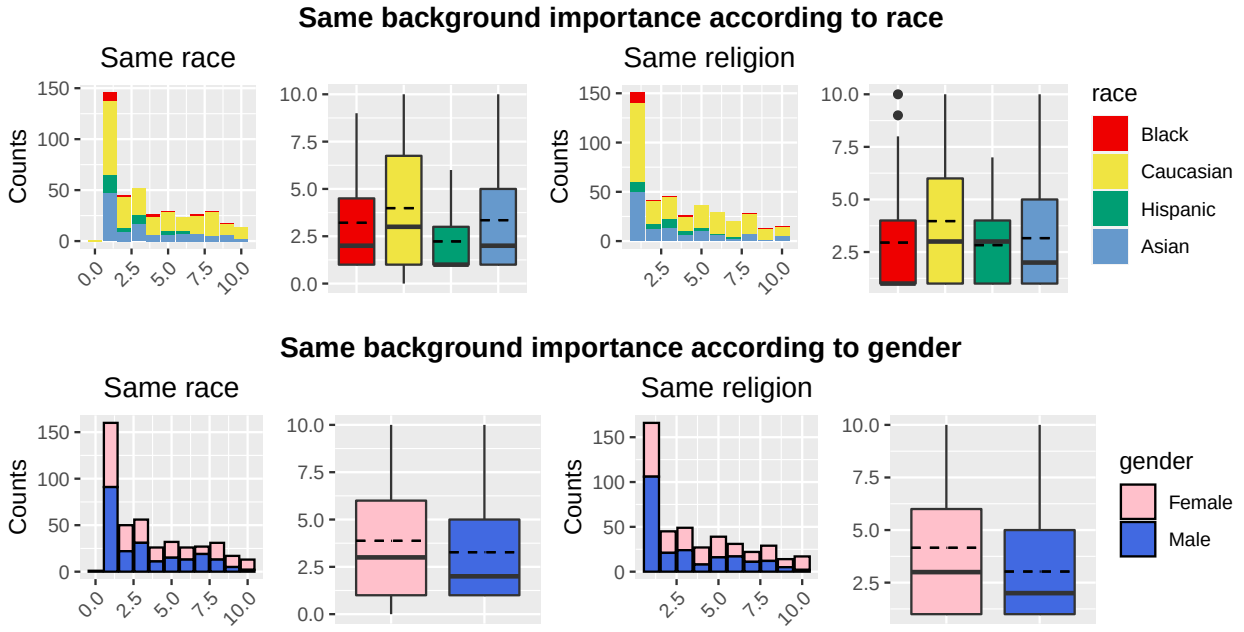
The dataset is reasonably even on the genders: 49.43% of the participants is indeed female. Gender is also balanced among races: most of the participants are Caucasian, but the proportion of Asians is also relevant. There is however a difference on the fields of study: while there are more females studying Humanities or Medical Sciences, most males are attending Business or Science courses. Most of the participants are in their twenties and males are in mean older of a few months.



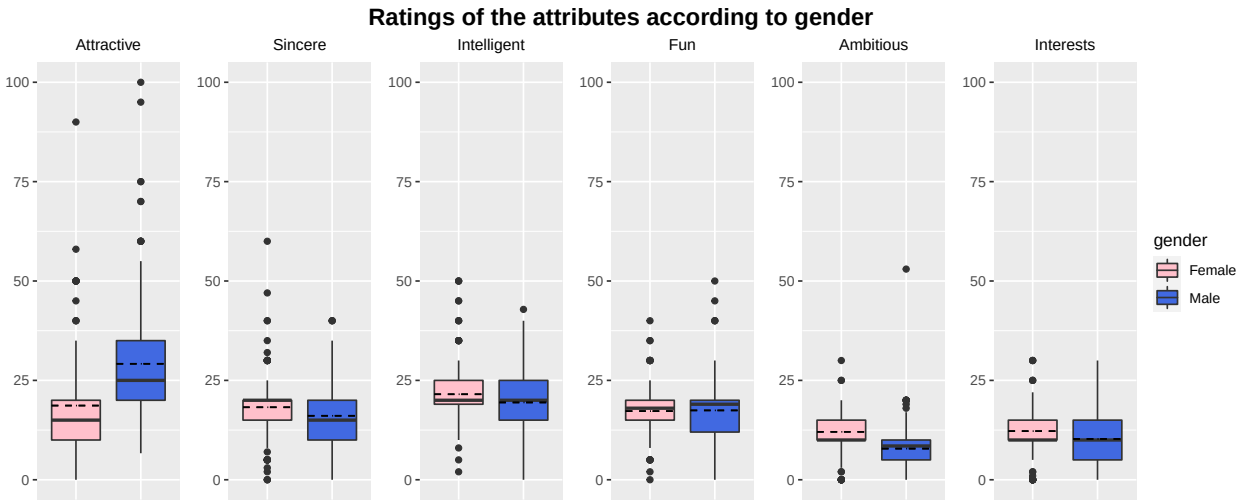
On the other hand, it could be more of interest to consider the age gap between the male and the female of each couple, as it could have an influence over the decision. Ages are quite balanced between the genders: the graph shows that in 52.32% of the dates the male is older. However, the decision variable follows the shape of age gap both for both gender, hinting at the fact that age difference might not be really taken into consideration when selecting a partner. At the end of the day, age is really just a number.



Also issues such as race and religion play much less of a role than expected. When registering to the Speed Dating events, participants were asked to give also some information on the importance of having the same background as the partner. We can see from the histograms that overall they do not attach great importance to having the same race and professing the same religion. However, it seems to be slightly more relevant to females than males and to Caucasian and Asian people.



Participants had also to rate the importance of six attributes in a potential partner by distributing 100 points among them. The boxplots of the ratings on *attractive* immediately catch the eye: it is clear indeed that males seem to put great emphasis on the physical appearance. However, males' ratings of the other attributes seem to be quite even, with the exception of *ambitious* which does not seem quite relevant. Females' ratings are instead more balanced, suggesting that all qualities matter for them: how picky! Last but not least, men decided to stay in touch with the partner 46.59% of the times, while women 38.73%. These percentages are not low, suggesting that speed dating could really help to find the other half of the apple.



It is time to drop the big gun: logistic regression

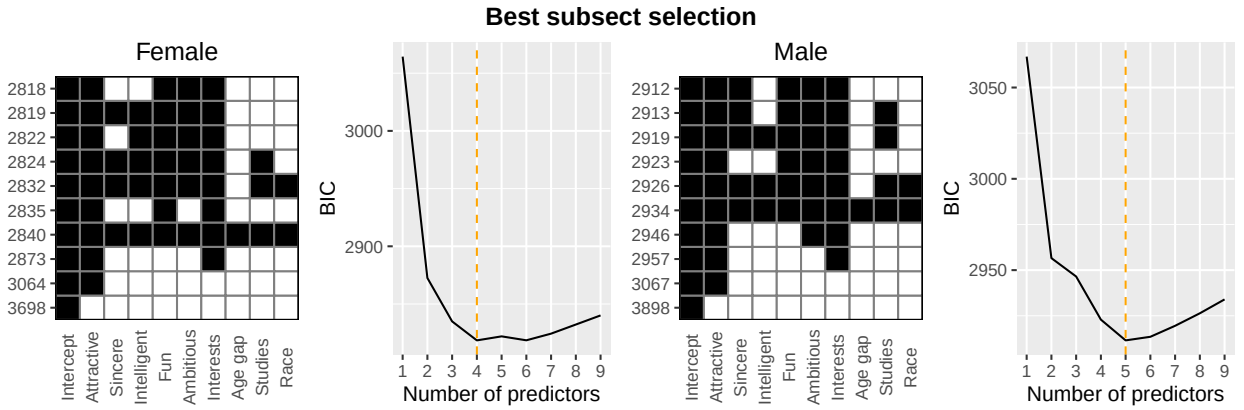
Logistic regression Now we would like to support our deductions with some Statistics. In particular, we would like to regress the variable *decision* with a generalized linear model: the logistic regression. By doing so, we will measure the relationship between the scores that one assigned to the partner and the decision made. As suggested from the data exploration, the six attributes could have an influence on the choice of a potential partner. If we have a look at the correlation between decision and each one of these variables, we see that some of them could be more relevant: we will investigate this later with the regression.

	Attractive	Sincere	Intelligent	Fun	Ambitious	Interests
Decision	0.502	0.19	0.201	0.391	0.209	0.387

The evaluation of a partner could take also into consideration the age gap, the cultural background and the field of study. We fit therefore a logistic regression model on the following attributes:

	Intercept	Attractive	Sincere	Intelligent	Fun	Ambitious	Interests	Age gap	Studies	Race
Female	-5.51 *	0.41 *	-0.13 *	0.17 *	0.29 *	-0.2 *	0.3 *	0	-0.15	0.02
Male	-5.15 *	0.67 *	-0.13 *	-0.07	0.26 *	-0.15 *	0.26 *	-0.01	0.26 *	-0.1

At a first glance, we can see that physical appearance is crucial when selecting a possible partner, especially for males. Being fun and sharing same interests also seem to be relevant for both genders, whereas attending lectures in the same field of study appears to be taken into consideration only by males. We apply best subset selection (BSS) to understand which variables really matter for the regression.



BIC criterion for BSS suggests the use of two different sets of predictors according to the genders:

- Attractive, Fun, Ambitious and Shared interests for **Females**
- Attractive, Sincere, Fun, Ambitious and Shared interests for **Males**

Moreover the best subsets are nested, inducing a natural order of relevance in the prediction of the decision:

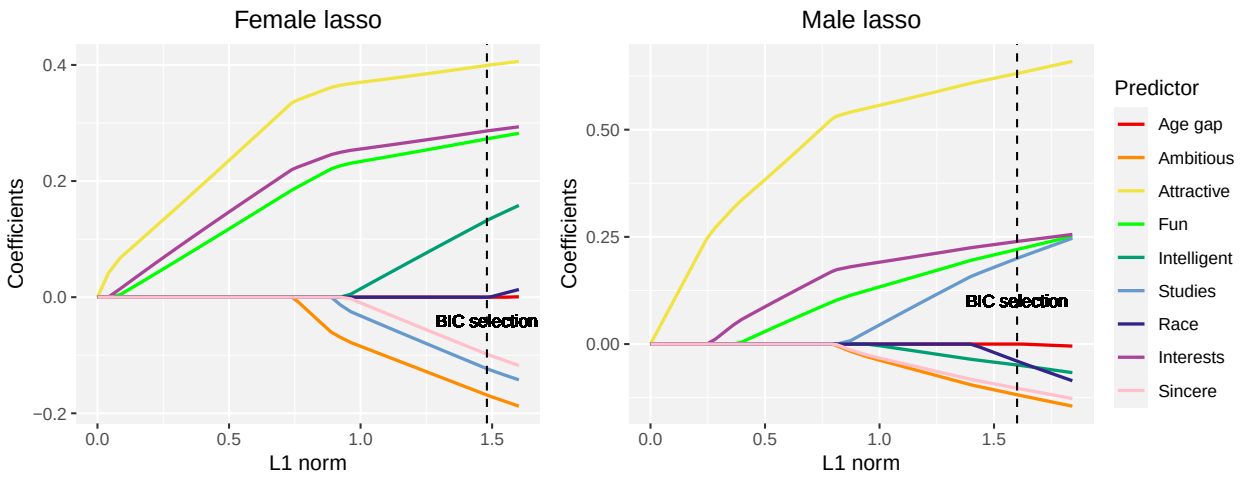
- Attractive, Interests, Fun, Ambitious, Intelligent, Sincere, Studies, Race and Age gap for **Females**
- Attractive, Interests, Ambitious, Fun, Sincere, Studies, Intelligent, Race and Age gap for **Males**

Performing further testing with anova, we reckon that being sincere is actually relevant only for males and adding other predictors to both models would not be that significant. We then fit a logistic regression with the selected predictors:

	Intercept	Attractive	Sincere	Fun	Ambitious	Interests
Female	-5.31 *	0.41 *	–	0.29 *	-0.17 *	0.28 *
Male	-5.25 *	0.65 *	-0.16 *	0.25 *	-0.17 *	0.27 *

We can therefore deduce that the final decision for both genders relies mostly on the physical appearance of the partner, in particular for males. It is indeed difficult to get a grasp on the personality of a complete stranger in four minutes. However, sharing common interests improves the chances of being chosen, as well as being fun. Yet, both genders think bitterly of ambition and males do not seem to appreciate sincerity: if you want to impress a stranger, it would then be better not to be blunt and reveal too much!

Lasso regression We now want to apply a penalty in order to shrink the coefficient estimates towards zero. We will consider Lasso technique as we want to get a sparse solution to simplify our model.



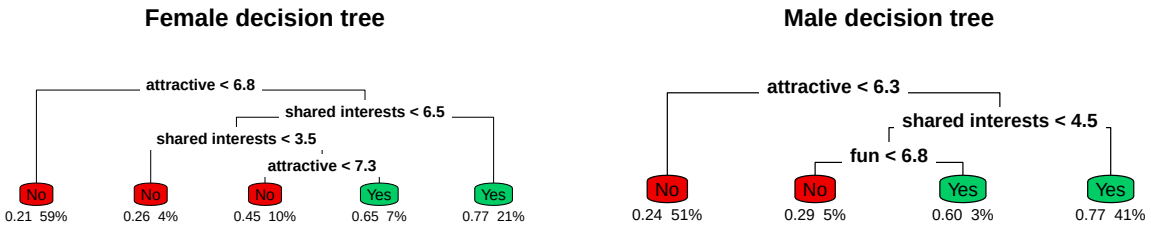
	Intercept	Attractive	Sincere	Intelligent	Fun	Ambitious	Interests	Age gap	Studies	Race
Female	-5.40	0.40	-0.1	0.13	0.27	-0.17	0.29	0	-0.12	0.00
Male	-5.16	0.63	-0.1	-0.05	0.22	-0.12	0.24	0	0.20	-0.04

We can observe a clear division of the predictors: Attractive, Shared Interests and Fun are the most relevant; Race and Age gap do not seem to affect the decision at all; all the remaining predictors have some impact, depending on the gender. Moreover, we can observe that the order of shrinkage of the coefficients agrees quite well with the ranking given by best subset selection.

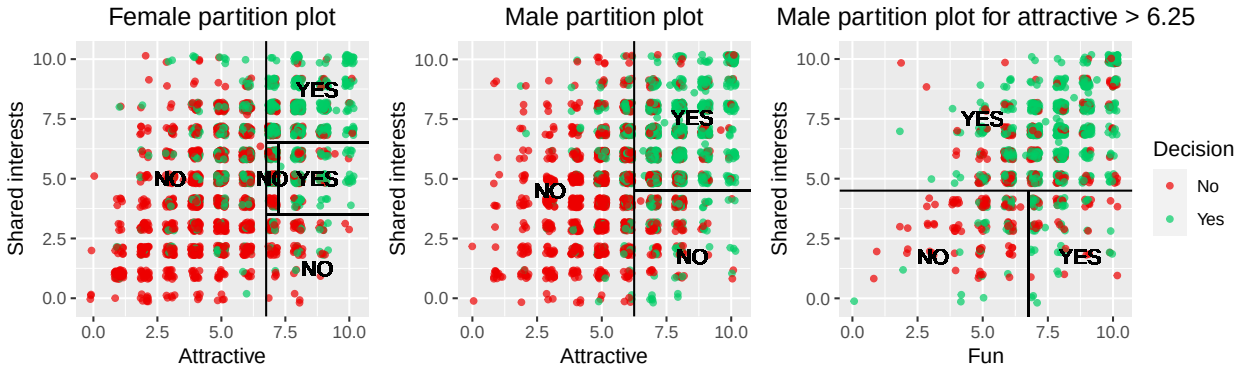
Comparison bewteen the models The tables below shows that there is almost no difference between the mean squared errors obtained through 5-fold cross validation: we prefer to use the model chosen by BSS for both genders as it is the most sparse.

	MSE	Intercept	Attractive	Sincere	Intelligent	Fun	Ambitious	Interests	Age gap	Studies	Race
FEMALES											
All	0.166	-5.51	0.41	-0.13	0.17	0.29	-0.20	0.30	0.00	-0.15	0.02
BSS	0.167	-5.31	0.41	0.00	0.00	0.29	-0.17	0.28	0.00	0.00	0.00
Lasso	0.166	-5.40	0.40	-0.10	0.13	0.27	-0.17	0.29	0.00	-0.12	0.00
MALES											
All	0.170	-5.15	0.67	-0.13	-0.07	0.26	-0.15	0.26	-0.01	0.26	-0.10
BSS	0.170	-5.25	0.65	-0.16	0.00	0.25	-0.17	0.27	0.00	0.00	0.00
Lasso	0.171	-5.16	0.63	-0.10	-0.05	0.22	-0.12	0.24	0.00	0.20	-0.04

Decision trees Now, let us support our analysis with some decision trees. They allow to have a better grasp on the classification by visualizing the process which brings to the final decision on the partner.



Once again, physical appealing is confirmed to be the most relevant attribute both genders look at when choosing a possible partner: if a person is not considered attractive enough, he/she is rejected. Shared interests are the second thing taken into consideration: it is indeed an indicator of compatibility, however the threshold required by females is far higher. If this criterion is not met, males look for a fun partner, whereas females require the date to be really attractive. Below we can find the partition plots of the trees: they partition the feature space into a number of smaller and non-overlapping regions with similar response values according to the splitting rules. As the male decision tree took into account three attributes, there are two plots: in the second one we are assumig that the score for attractiveness is higher than 6.25.



What did we learn?

Let us recall the chosen regression model on the decision after the speed date:

	Intercept	Attractive	Sincere	Fun	Ambitious	Interests
Female	-5.31 *	0.41 *	–	0.29 *	-0.17 *	0.28 *
Male	-5.25 *	0.65 *	-0.16 *	0.25 *	-0.17 *	0.27 *

It is all fun and games, but how can we use it? The model works in the following way: if all 0 scores are assigned to the partner, the odds of a positive decision are around 0 for both genders (duh!). Then, for each attraction point assigned by a woman, the odds of wanting to see the partner again become 1.51 times larger. Analogously, one point assigned to fun or shared interests corresponds to an increase rate for the odds of 1.34 and 1.33 respectively. On the other hand, one point of ambition makes the odds 1.18 times smaller. From the point of view of men instead, each attraction point makes the odds 1.92 times higher. Similarly, one point of fun or shared interests corresponds to an increase rate of 1.29 and 1.31, while one point of sincere or ambition makes the odds 1.17 and 1.18 times smaller respectively.
The take home is that what counts the most for university students is the first impression you give, as the time is limited. It is hard indeed to appreciate the intelligence of a stranger in 4 minutes. Hence, we rely mostly on our gut feelings, which are heavily influenced by how much we find that person attractive or compatible. If there is match, you will have plenty of time to see whether they checks all the boxes in the long list of attributes that describe our perfect mate. What are you waiting for? Get all dressed up and sign for the next Speed Dating event in your campus!