Speed Dating: *Hobbies and Matching*

By Cate and Matt

Data Set

This dataset was compiled by Columbia Business School professors Ray Fisman and Sheena Iyengar.

Data was gathered from participants in experimental speed dating events from 2002-2004. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end, participants were asked if they would like to see their date again. They were also asked to rate their date on six attributes: **Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests.**

The dataset also includes questionnaire data gathered from: demographics, dating habits, self-perception across key attributes, beliefs on what others find valuable in a mate, and lifestyle information.

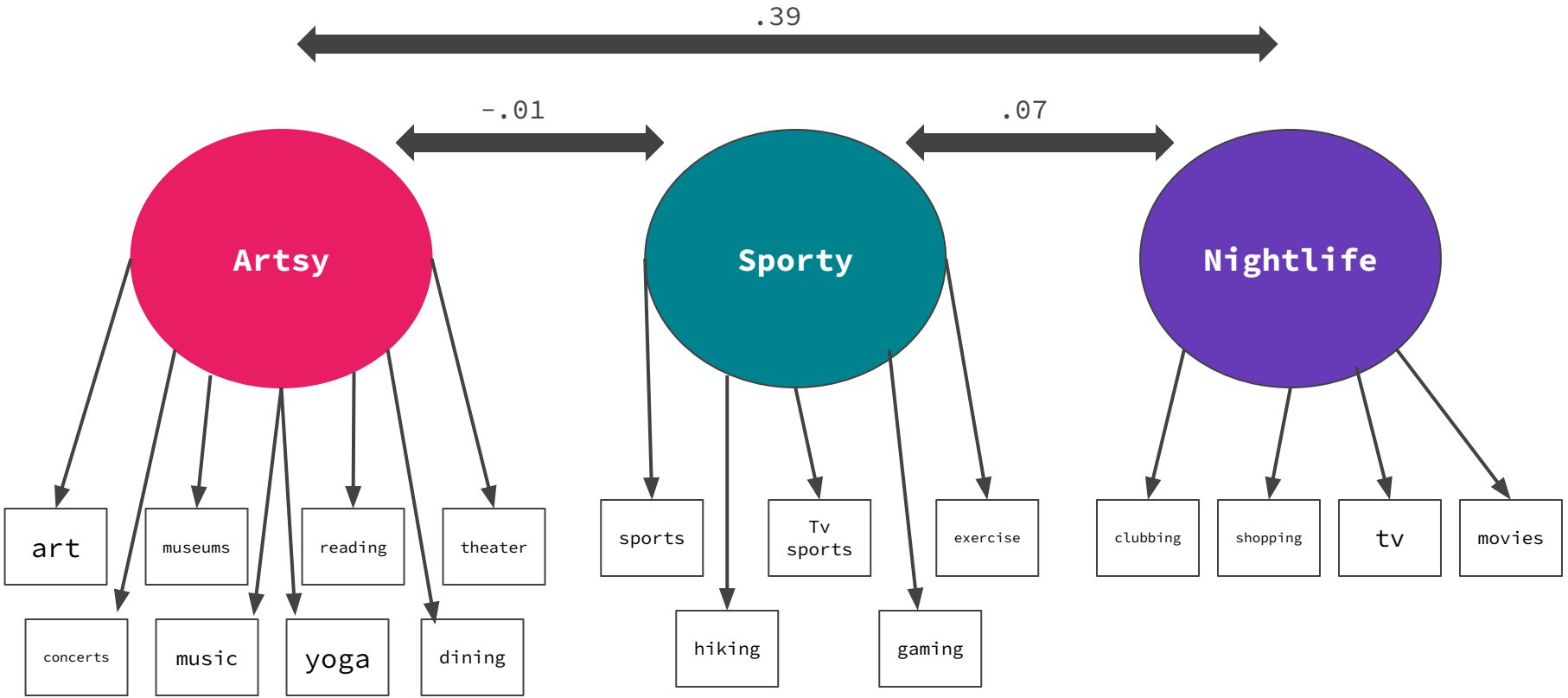
In total there was initially **8378 observations** of **195 variables**, but after we tidied the data there was **1562 observations**.

Main research questions

What attributes (as rated by your partner) are most predictive of matching? Do these differ between genders?

Do ratings of important attributes in a match fall into broader categories of prioritization?

Do hobby ratings fall into broader categories of interest areas?


.39
-.01
.07
Artsy
Sporty
Nightlife
art
museums
reading
theater
concerts
music
yoga
dining
sports
Tv
sports
exercise
hiking
gaming
clubbing
shopping
tv
movies

Group Comparison: Attribute Ratings

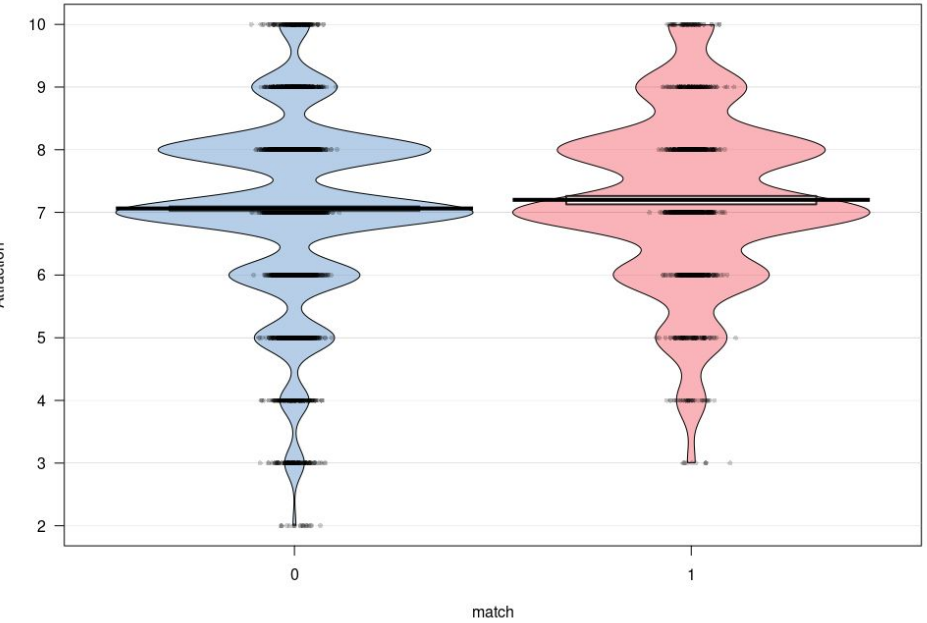
---

5 Attributes: Intelligence, Fun, Sincerity, Ambition, and Attraction

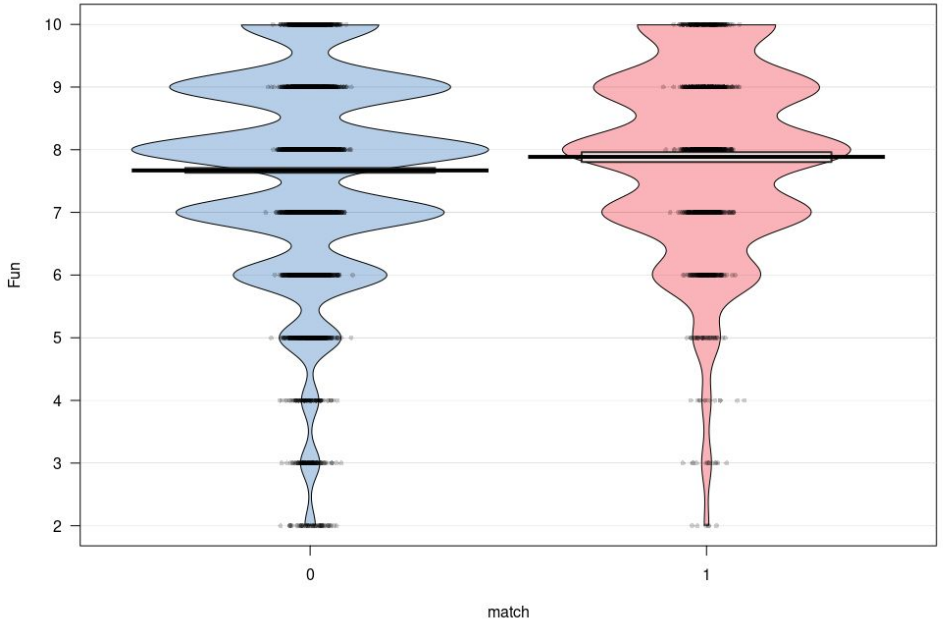
Which Attribute Rating is most important for getting a match?

Group Comparison: Visualizations

Attraction



Fun



Group Comparison Results

Fun Probability of Superiority: 54%

Attraction Probability of Superiority: 53%

Ambition Probability of Superiority: 51%

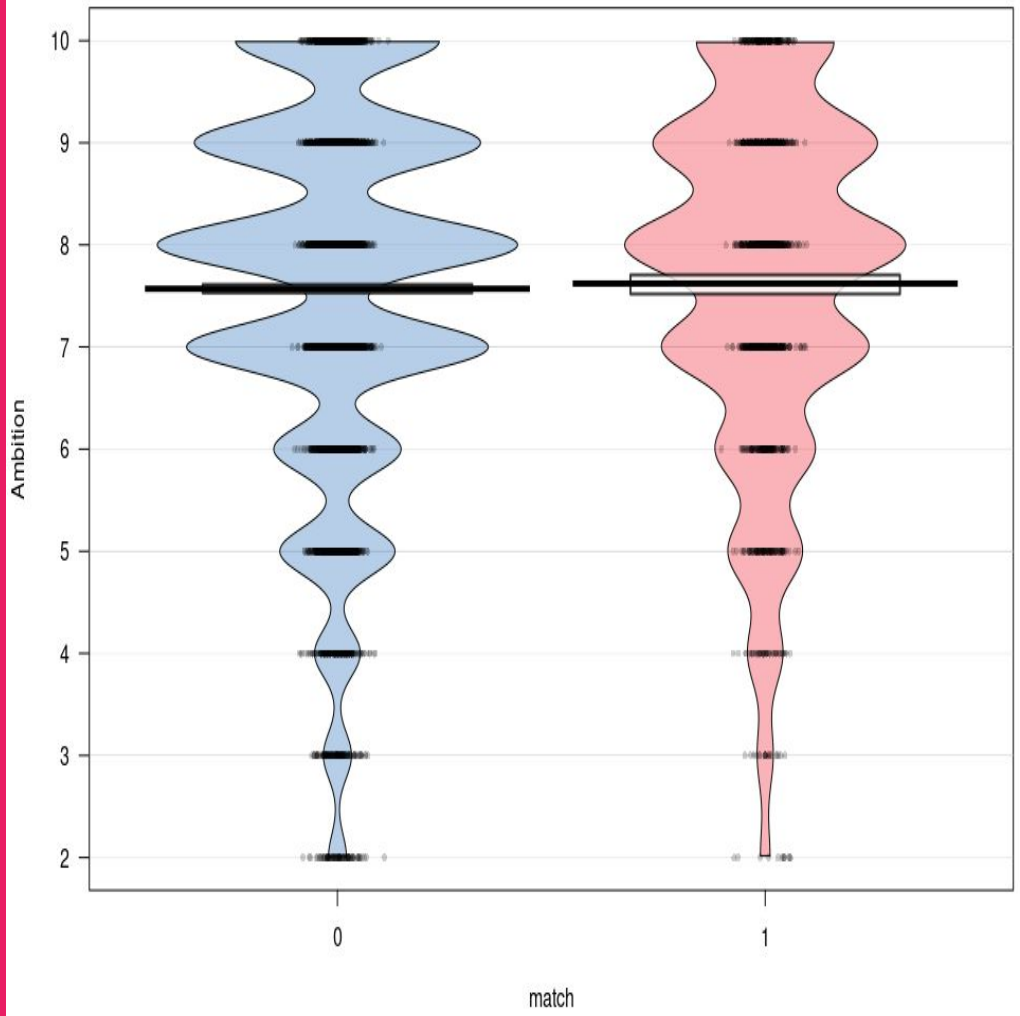
Intelligence Probability of Superiority: 50%

Sincerity Probability of Superiority: 50%

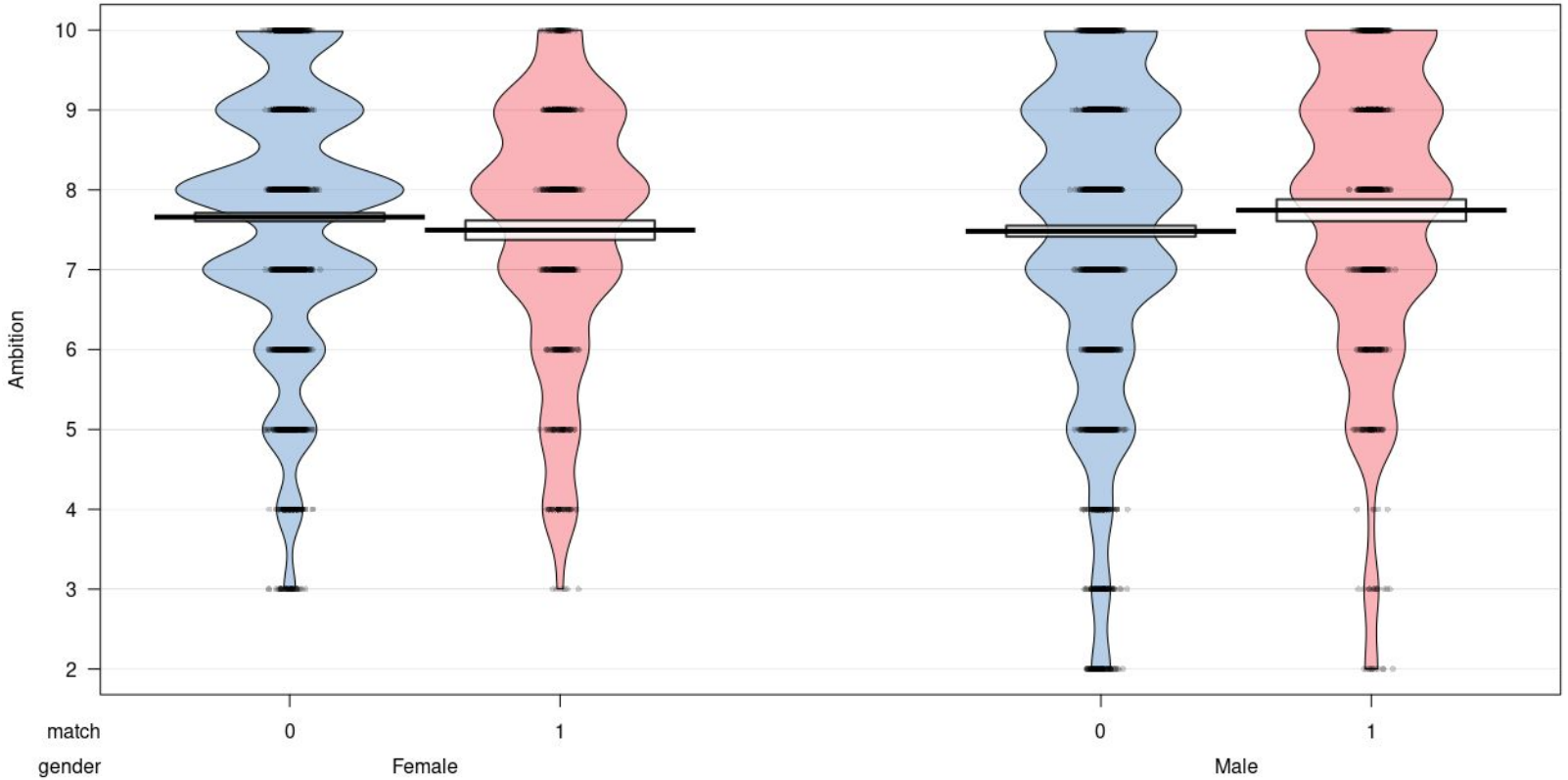
```
> cles(Attraction ~ match, data = all)
Parameter        | Coefficient |        95% CI
-----------------------------------------------
Pr(superiority)  |        0.47 | [0.46, 0.49]
Cohen's U3       |        0.46 | [0.44, 0.48]
Overlap          |        0.96 | [0.94, 0.98]
>
> cles(Fun ~ match, data = all)
Parameter        | Coefficient |        95% CI
-----------------------------------------------
Pr(superiority)  |        0.46 | [0.44, 0.48]
Cohen's U3       |        0.44 | [0.42, 0.47]
Overlap          |        0.94 | [0.92, 0.97]
>
> cles(Intelligence ~ match, data = all)
Parameter        | Coefficient |        95% CI
-----------------------------------------------
Pr(superiority)  |        0.50 | [0.48, 0.51]
Cohen's U3       |        0.50 | [0.47, 0.52]
Overlap          |        1.00 | [0.97, 1.00]
>
> cles(Ambition ~ match, data = all)
Parameter        | Coefficient |        95% CI
-----------------------------------------------
Pr(superiority)  |        0.49 | [0.48, 0.51]
Cohen's U3       |        0.49 | [0.47, 0.51]
Overlap          |        0.99 | [0.97, 1.00]
>
> cles(Sincerity ~ match, data = all)
Parameter        | Coefficient |        95% CI
-----------------------------------------------
Pr(superiority)  |        0.50 | [0.49, 0.52]
Cohen's U3       |        0.50 | [0.48, 0.53]
Overlap          |        1.00 | [0.97, 1.00]
```

Group Comparison: Subgroups

Ambition: General

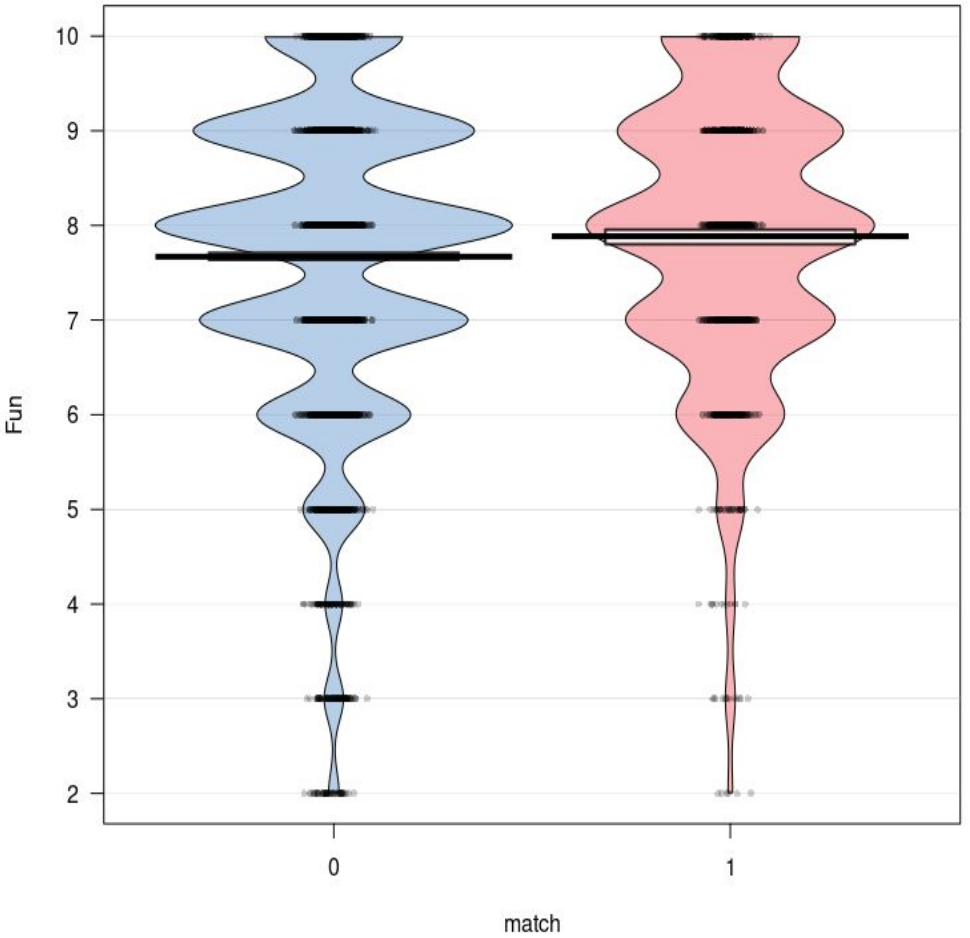


Subgroups: Ambition and Gender

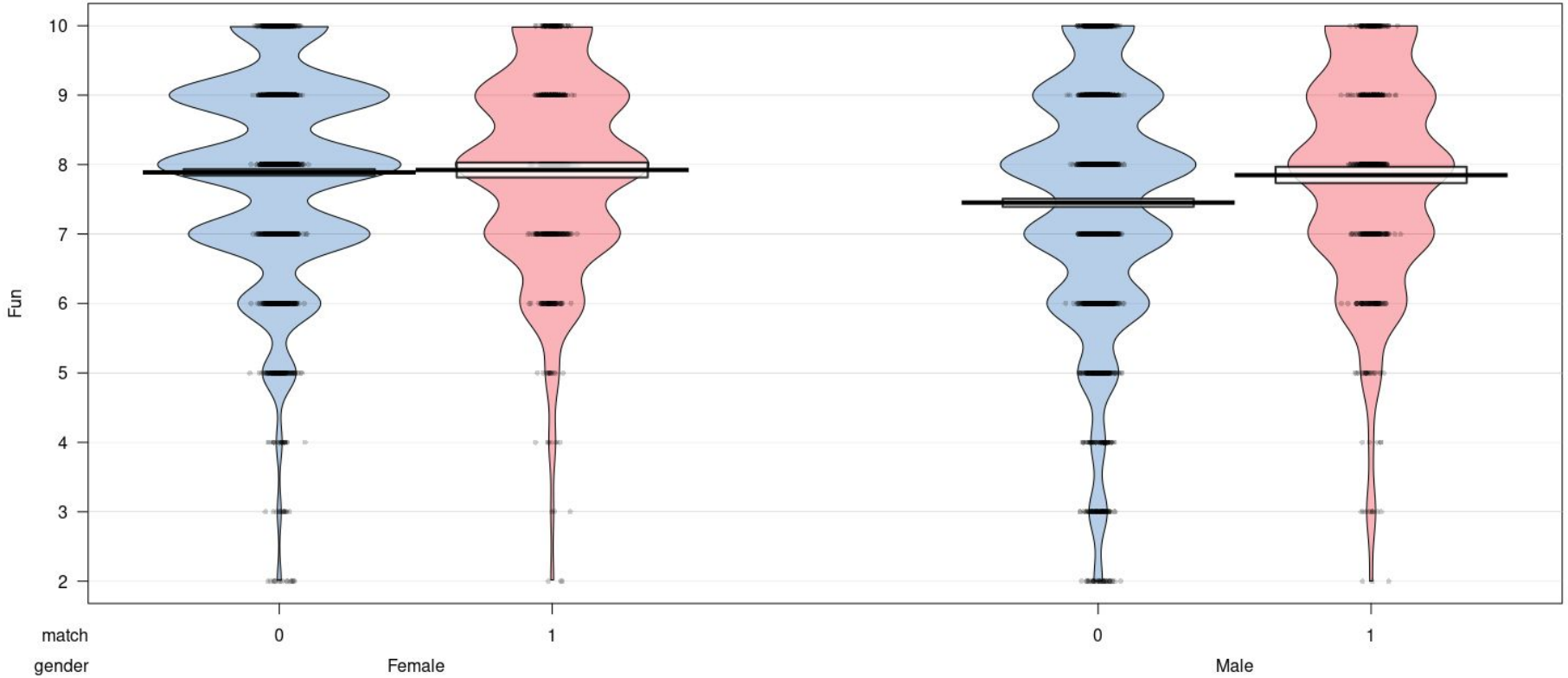


Group Comparison: Subgroups

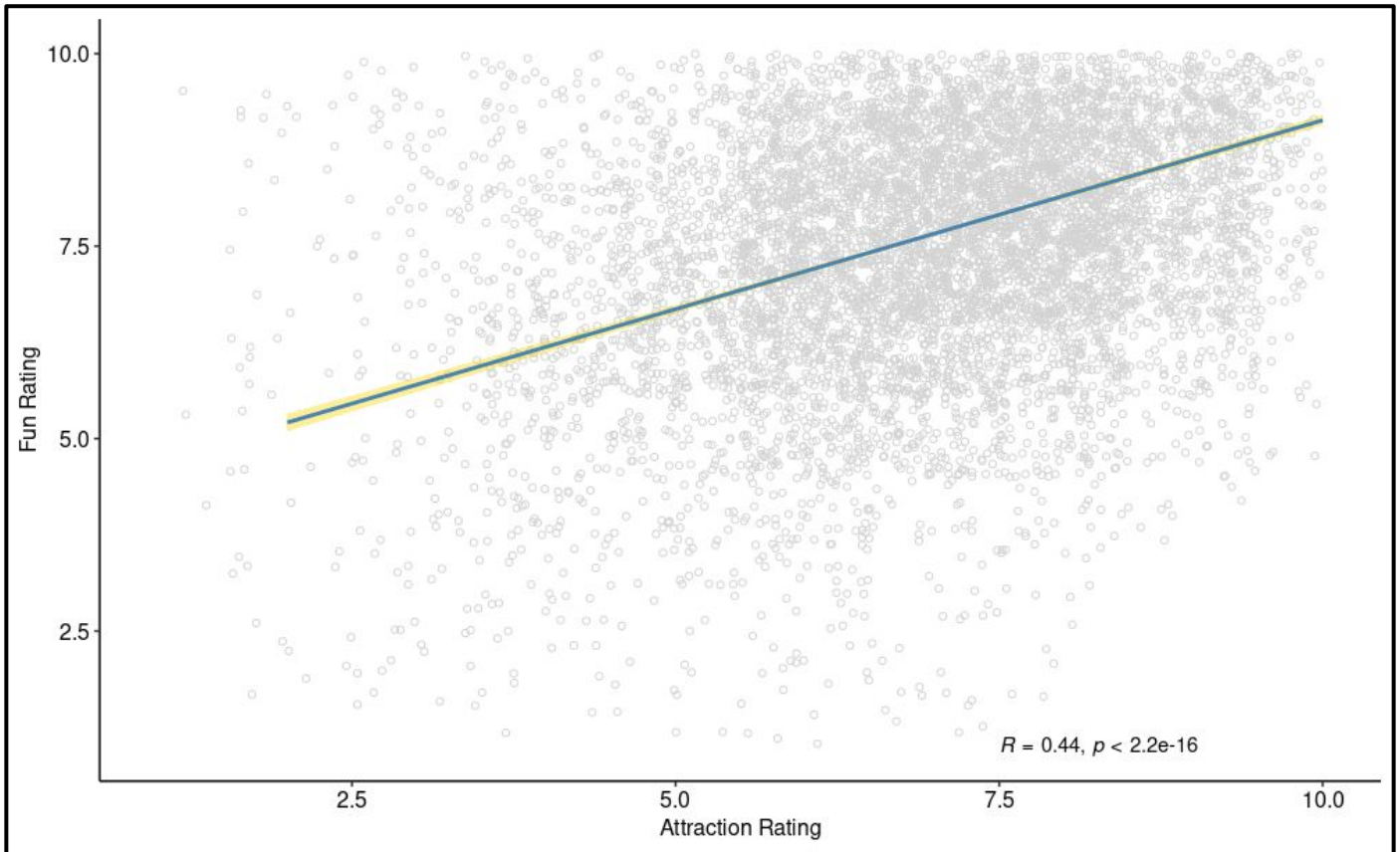
Fun: General



Subgroups: Fun and Gender



Regression Visualization: Attractiveness and Fun



Regression Results

Regression results using Attraction as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	4.07**	[3.94, 4.21]						
Fun	0.39**	[0.37, 0.41]	0.44	[0.42, 0.46]	.19	[.18, .21]	.44**	
								R^2 = .192**
								95% CI[.18,.21]

Ethical reflection

Were participants informed on how their data would be collected or used?

How can people use our data analysis in a harmful way?

Have we tested for fairness between different user groups?

What groups have been collecting this data?

Future directions

Does demographic background influence a person’s preferences on the six attributes and/or hobbies?

Is age a factor in a person’s preferences or eagerness for a match?

What might not be able to be captured by this data set that could play a role in getting a match?