

INFO 4410

Final Report

Noorejehan Umar, Sydney Bednar, Victoria Eshun

**Project Topic:** Using a speed dating dataset, we explored the differences in what people say they prefer in potential partners versus what actually led them to like their partner. We compared this data across demographics such as gender and race.

**Data Source:** [Speed Dating Dataset](#)

### Data Description

This dataset contains information about participants in speed dating events from 2002-2004. Participants were asked to answer a series of questions throughout the events about their demographics, preferences in a partner, self-perception, beliefs on others' values in a partner and interests. The data was compiled by professors Ray Fishman and Sheena Iyengar at Columbia Business School and was used in their paper, "Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment." For our project, we are interested in investigating the following questions/use cases from this data:

- What attributes do people think they value in a partner?
  - To investigate this question, we will look at how highly or lowly people rate the importance of attributes in potential partners. The attributes we will focus on are ambition, humor, attractiveness, intelligence, sincerity and the presence of shared interests.
- What attributes do people *actually* value in a partner?
  - To investigate this question, we will look at which attributes actually led participants to like their partner. To do this, we will compare how people rated their partner on all six attributes after speaking with them and how this correlated to how much they ended up liking their partner.
- How do preferences vary based on race and gender?
  - To investigate this question, we will include information about each person's race and gender.

We have subsetting the data, narrowing down the original 123 column dataset to include the following 24 columns:

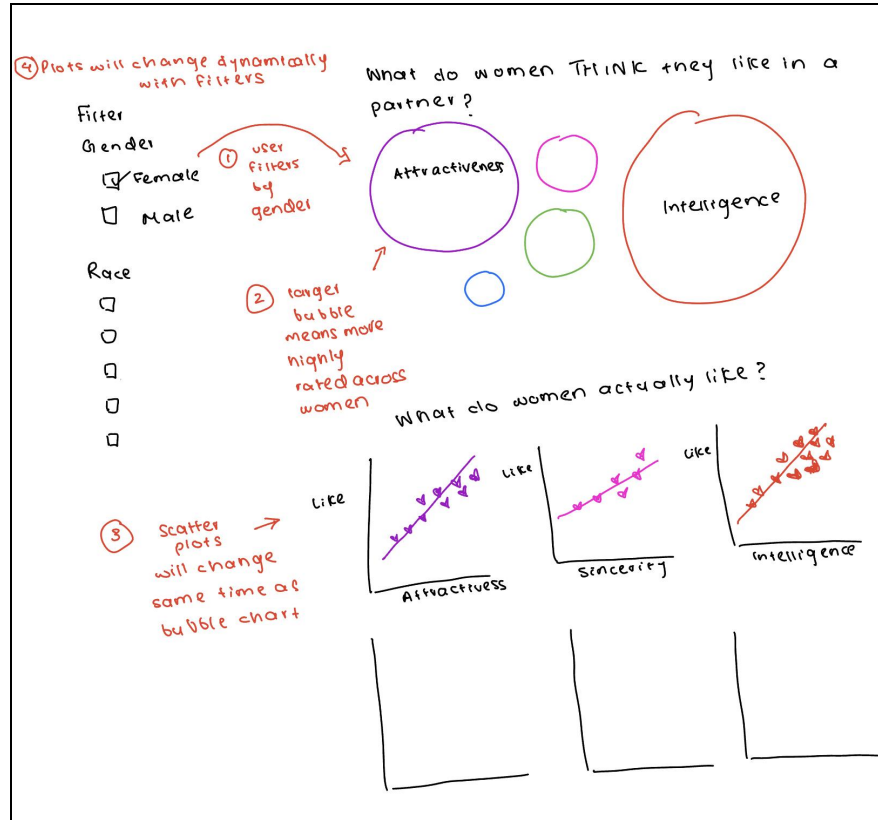
variable	description
age	Age of participant
age_o	Age of partner
ambition_partner	Rating of partner's ambition
ambition_important	How important does participant rate ambition
attractive_important	How important does participant rate attractiveness
attractive_partner	Rating of partner's attractiveness
decision	Decision to see date again (0: no, 1: yes)
field	Field of study
funny_important	How important does participant rate humor
funny_partner	Rating of partner's humor
gender	Participant's gender
importance_same_race	How important does participant rate being of the same race (1-10)
importance_same_religion	How important does participant rate having same religion (1-10)
intelligence_important	How important does participant rate intelligence
intelligence_partner	Rating of partner's intelligence
like	How much did the participant like their partner (1-10)
match	If participant and partner matched (0: no, 1: yes)
race	Race of participant
race_o	Race of partner

samerace	If participant and partner have the same race (0: no, 1: yes)
shared_interests_important	How important does participant rate having shared interests
shared_interests_partner	Rating of partner's shared interests
sincere_important	How important does participant rate sincerity
sincere_partner	Rating of partner's sincerity

## Storyboards

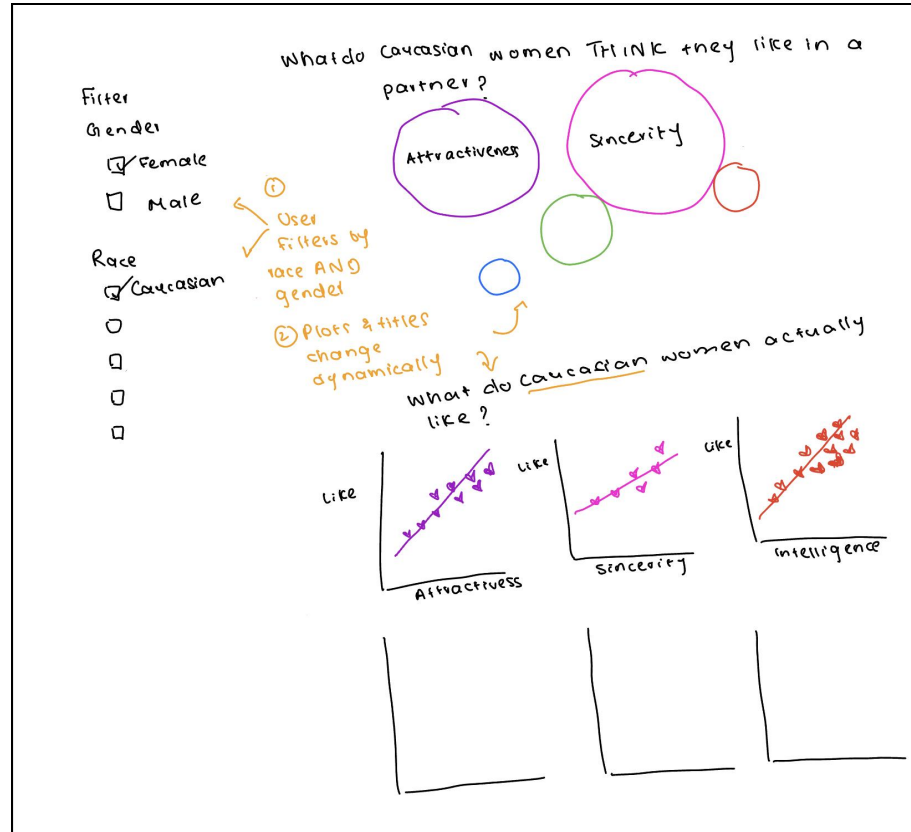
Our main goal is to allow users to compare what they think they value in a partner with what they actually like, while also allowing them to filter by gender and race. In the first image below, our storyboard highlights the steps the user will take to interact with our visualization:

1. First the user will choose a filter, e.g. the user below chooses to filter by gender and look at what women think they like in a partner vs. what they actually like
2. The bubble chart would change to reflect the attributes that all women value collectively, with bigger bubble sizes corresponding to higher rated attributes
3. The scatterplots would update at the same time as the bubble chart, and would show what women actually like; Each of the six attributes would be placed on the x axis, with 'like' on the y axis \*\*SEE NOTE BELOW re: changing design to heatmap
4. Lastly, if the user chooses a different filter, the plots would update dynamically



We also want to show how the attributes people value vary with gender AND race, i.e. do Caucasian women's valued attributes vary significantly from Latin American women? To allow users to see these differences, we allow them to filter by a specific gender and race. For example, in the image below:

1. The user filters by women and Caucasian
2. The plots update to reflect the attributes that Caucasian women think they like and what they actually like
3. The titles of the graphs also update to signal to the user what demographic group they are looking at

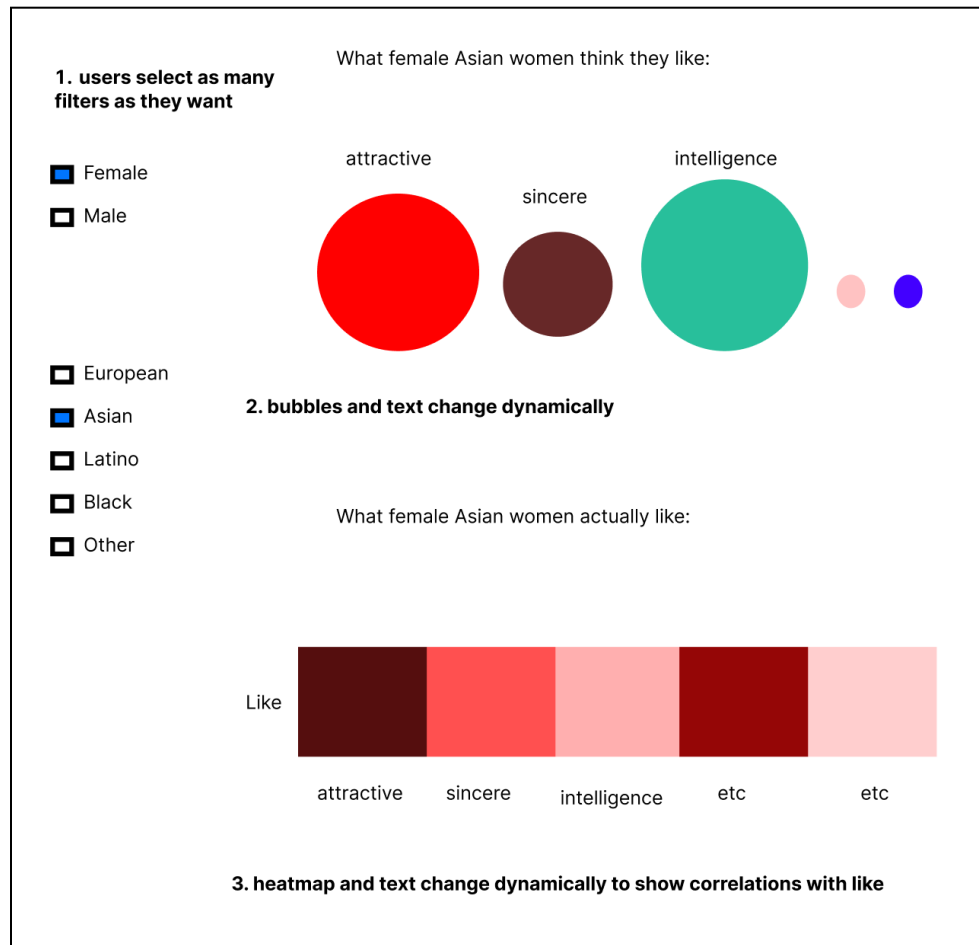


## Justification

The reason we chose to focus on these interactions is because we believe filter buttons would be the most intuitive way for users to view subsets of the data; Since the categories are all listed on the screen, the users will be able to easily switch between filters. We also believe that stacking the two plots on top of each other will make it easy for users to compare perceived and actual values easily, because the plots would be updating at the same time. We chose to focus on the bubble size dynamically changing because it would provide visual cues of which attributes are the most highly rated amongst that demographic and would not require a lot of work to interpret from the users' side.

We also focused on dynamic scatterplots to show the actual like, instead of bubble charts, because we wanted to show more detail about how each attribute is correlated with like, which is something the user would not be able to see through just the bubble charts.

**[Note]:** After starting the design process and seeing the actual visualizations, we decided to change our second visualization to a heatmap. Here is a storyboard that we designed for it before implementing it. We describe why we chose it in the development process.



## Final Interactive Visualization

### Speed Dating

Do we have a good perception of what we desire in a partner? In these charts, we explore differences in what people say they prefer in a partner versus what actually led them to like their partner during a speed dating event. We examine 6 key attributes: attractiveness, sincerity, intelligence, humor, ambition, and the presence of shared interests.

**\*\*Select a filter to begin.\*\***

#### Gender

- ☒ Female  
☐ Male

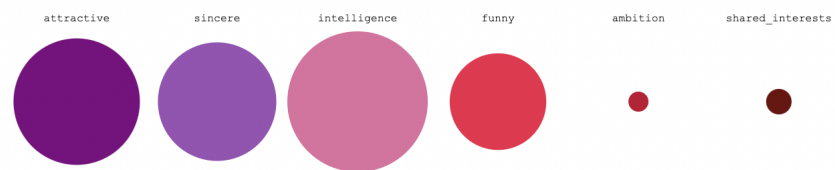
#### Race

- ☐ European/Caucasian-American  
☐ Asian/Pacific  
☒ Islander/Asian-American  
☐ Latino/Hispanic American  
☐ Black/African American  
☐ Other

### What do people *THINK* they prefer in a partner?

This bubble chart shows that people who are **Female, Asian**

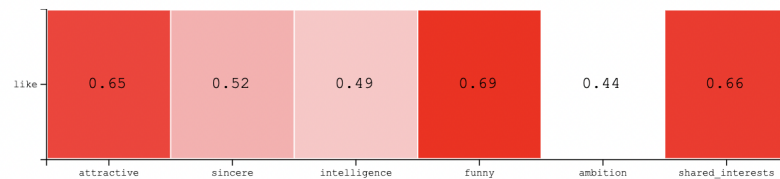
*THINK* that the most important attribute in a partner is **intelligence**



### What do people *ACTUALLY* prefer in a partner?

This correlation heatmap shows that people who are **Female, Asian**

*ACTUALLY* like a partner because of the attribute **funny**



Our final visualization allows users to compare what people from different demographics (ie race and gender) think they prefer in partners and what they actually like. It does so in the following ways:

- **Checkboxes:** The way that we allow our users to filter is by providing them with checkboxes on the left side of the screen. Once users check the filter that they want to view, the visualizations change accordingly. The user can filter by any combination of race and gender, including by both genders and all races.
- **Bubbles:** The bubbles change dynamically with the filters selected, i.e if a user selects female and Asian, the bubbles will show the attributes that all Asian women think they prefer in a partner. The text on top of the bubbles

also dynamically changes to reflect what demographics the user has selected, as well as what the top rated attribute of this demographic is. The bubbles are scaled so that their size is reflective of how highly rated the associated attribute was, eg. If Asian women think that intelligence is the most important attribute in a partner, the intelligence bubble will be the largest. Each attribute bubble also has its own unique color to make it easier for users to differentiate between them.

- *Heatmap*: The heatmap also changes according to the filters selected. Unlike the bubbles, the heatmap shows what people actually like in their partners by computing the correlations of each attribute with the 'like' variable. The darker colors on the heatmap show the attributes that had higher correlations with like (ie the attributes that they liked more), and lighter colored ones show the ones with lower correlations with like. We also added text onto each box that showed the correlation values to account for accessibility issues, especially because a lot of correlations were very close in value to each other and the color differences were less obvious. Similar to the bubble chart, the text on top of the heatmap also changes dynamically to show what demographics the user has selected, and what their top attribute is.

## **Issues and Trade-offs**

- *Bias*: The visualizations that we created are based on self-reported data, which could be biased and not reflective of what people actually desire in a partner. While we cannot avoid what people rate their preferences to be and make sure they're being honest, we could annotate our visualizations to make sure that viewers know that the information is based on people's personal preferences and ratings, and might not be completely accurate.
- *Representation*: We chose to limit the demographics that we focused on to race and gender because we didn't want to overwhelm our users with too much information. The trade-off associated with this is that other factors that we did not look at could also be affecting what people liked or thought they liked. For the sake of this project, we still wanted to limit our visualizations to the demographics we chose because those were the topics we were most interested in. However, a larger scale project could investigate every contributing factor in the dataset to see which one seemed to have the largest effect on people's preferences.



- *Complex Information:* Our interactions give users the opportunity to explore many different filters and presents them with two charts, which is a lot of information at once. We tried our best to make the visualizations as straight-forward and simple as possible by annotating them dynamically and giving the users as much context as possible. Some users may still find the information overwhelming, but we hope that by reading our descriptions, they eventually understand how the two charts connect.
- *Accessibility:* One issue that we really considered with our interactions is that a lot of it is dependent on visual cues such as changes in size and color, which could potentially lead to accessibility issues. To alleviate this, we did our best to add text to both of our visualization which provides users with the same information that they see in the visualization eg. our text changes dynamically to highlight the highest attributes, the demographics being shown, and the correlation values.

## **Development Process**

During the storyboarding process, we wanted to focus on a bubble chart that showed what people thought they liked and corresponding scatter plots that showed what people actually liked in their partners. Furthermore, we wanted to let users filter by a gender and a race. However, as we started developing we realized that some of these interactions were not the best way to answer the research questions we were focusing on.

First of all, the scatterplots were not good at showing any trends because we were focusing on two ordinal variables ie. type of attribute and like. We considered other possible visualizations such as stacked histograms, but in the end we decided to go with a heat map because we felt that it would be the more intuitive visualization for comparison between different attributes.

Another thing that we changed since the storyboarding is the dynamic filters. Initially we had thought we would allow users to filter by gender and by race, or by one gender and one race. However, after we started creating our visualizations, we felt that some users may want to explore more filters - they might want to see what both genders in a certain race liked or what males in all races liked.

Therefore, we changed our filtering so that users could select as many as they liked.

As mentioned earlier in the trade-offs sections, some concerns we had were that we were presenting a lot of information and focusing a lot on visual cues such as

color and size. However, we added dynamically changing text to ensure that people with accessibility concerns could also interact with our visualizations. Another trade-off we considered was whether we should show all the information on the screen before the users interacted with the checkbox filters. However, we thought this would be really confusing if the checkboxes were not checked yet - instead, we decided that nothing would show on the screen in the beginning and users would be prompted through text to start selecting filters to view visualizations.

### **Contributions:**

Stage of Project	Time Spent
Cleaning Data	1 day
Static Visualizations	1 - 2 days
Dynamic filtering	3 days
Styling	3-4 hours
Write-up	1 day

### **Individual Contributions:**

Sydney:

- Brainstorm and create initial sketches for charts, create heatmap, color scales and connect charts, create dynamic text, styling for webpage, help with report

Noorejehan:

- Brainstorm and create initial sketches, create bubble chart, created dynamic filters, create dynamic text, styling, wrote report

Victoria:

- Brainstorm dataset ideas, worked on bubble chart, created filters, worked on dynamic text, styling, worked on heatmap, edit report