

SPEED DATING

Dylan Tan

dylantan@college.harvard.edu

Jack Wongtam

cwongtam@college.harvard.edu

Fanney Zhu

fanneyzhu@college.harvard.edu

CS171 Final Project

TF: Samuel Gratzl

Spring 2015

Background and Motivation

Much to our dismay, a large proportion of the general populace are single. Amidst our hapless and woebegone plight, we all yearn for the dream of a rose colored daily life replete with raven-haired maidens and starry-eyed gentlemen. In a desperate ploy to escape the lovelorn monotony of an unfulfilling personal life, we sought guidance from the perennial bachelor's haven that is speed dating.

We found this data set while browsing for interesting trends in dating. According to Fisman's paper *Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment*,

"Women put greater weight on the intelligence and the race of partner, while men respond more to physical attractiveness. Moreover, men do not value women's intelligence or ambition when it exceeds their own. Also, we find that women exhibit a preference for men who grew up in affluent neighborhoods. Finally, male selectivity is invariant to group size, while female selectivity is strongly increasing in group size."

Thus, we hope to explore these relationships through a visualization of the data in Fisman's study along with.



image from: <http://twenties.co/>

Data

Our data can be found at this link:

<http://www.stat.columbia.edu/~gelman/arm/examples/speed.dating/>

Some interesting data fields include:

- Participant's Preferences in Partner
- Participant's Self Ranking
- Participant's Rankings of Partners Met
- Participant's Goal in Attending Speed Dating
- Participant's Demographics (ethnicity, income, city, field of work, university, etc.)
- Participant's Matches with Partner

Data Clean-Up:

Our data came in as a csv file with each row representing one participant's meeting with another participant of the opposite sex. The columns contain the participant's personal information as well as their preferences and how they ranked their partner. In order to make our data easier to use, we filtered out the data to include unique information pertaining to that one person and a list of all the partners that person met and how he/she ranked his/her partner. Afterwards, we aggregated the data by waves for easier accessibility. All the formatting is done in our index file and passed to our visual objects.

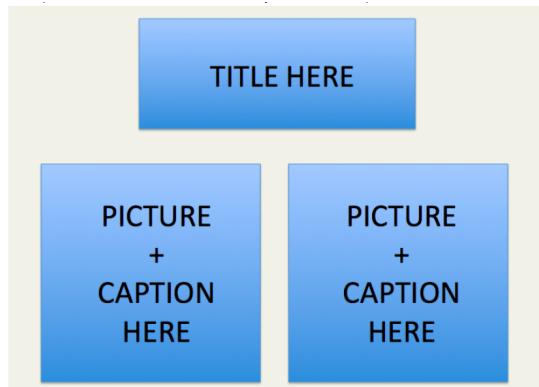
Initial Project Proposal

(April 3rd)

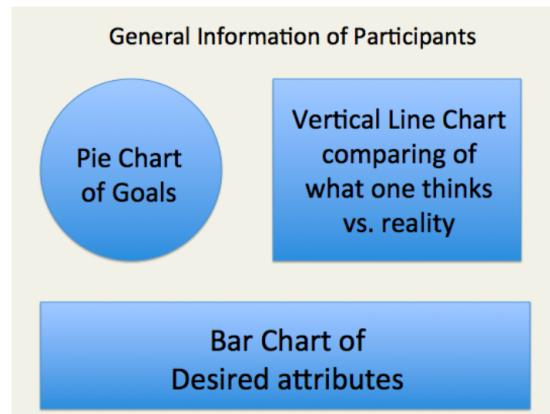
We begin our project by looking at the key questions that our data can answer. Our data contains 21 waves of speed dating sessions and the selections that each individual made in each session. Since demographical data of each participant was collected, this gives us an opportunity to explore selection preferences for different demographics (age group, ethnicity, income, location, etc.). Additionally, we have information on each participant's ratings of their partners in each wave. Lastly, the data contains information about how successful the speed dating session was for each participant after the event in the form of mutual matches. Thus, we want to explore the trends in participants' initial preferences, participants' actual preferences, and participants' success after the speed dating event by having three visuals.

Initial outline:

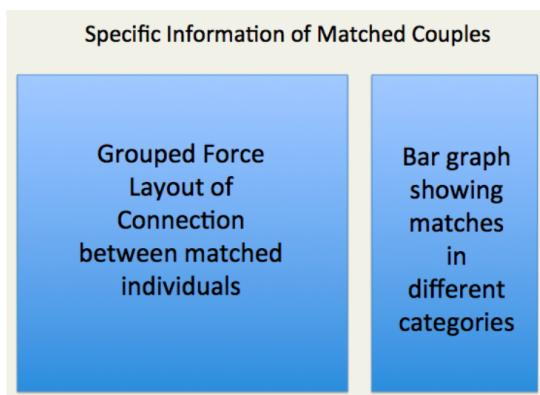
Visualization 1



Visualization 2



Visualization 3



Design Studio

(April 14th)

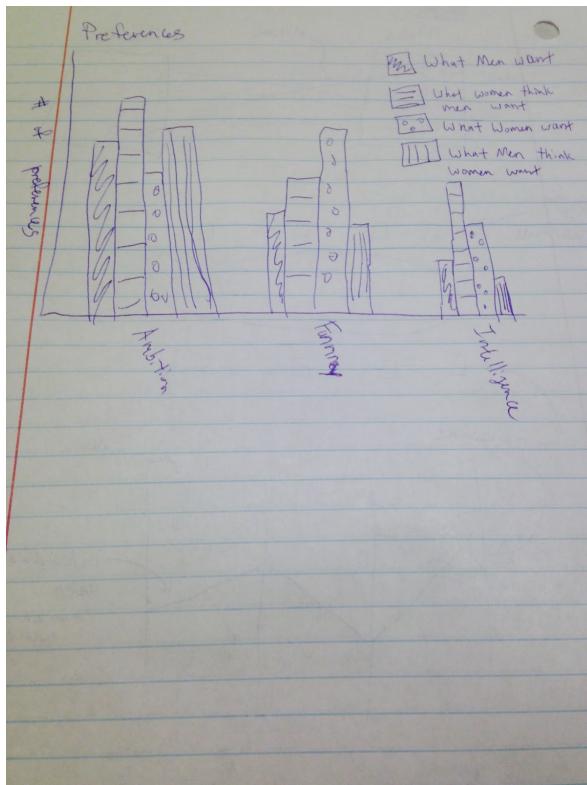
During the design studio, we met with another group to discuss our project proposal. After presenting our project, they had several suggestions. Many of the questions they had were what we were interested in:

- What was the success rate of speed dating?
- How did preferences differ between males and females?
- Were people likely to match with someone with similar qualities?
- What variables affected speed dating success? (i.e. position, race, etc.)

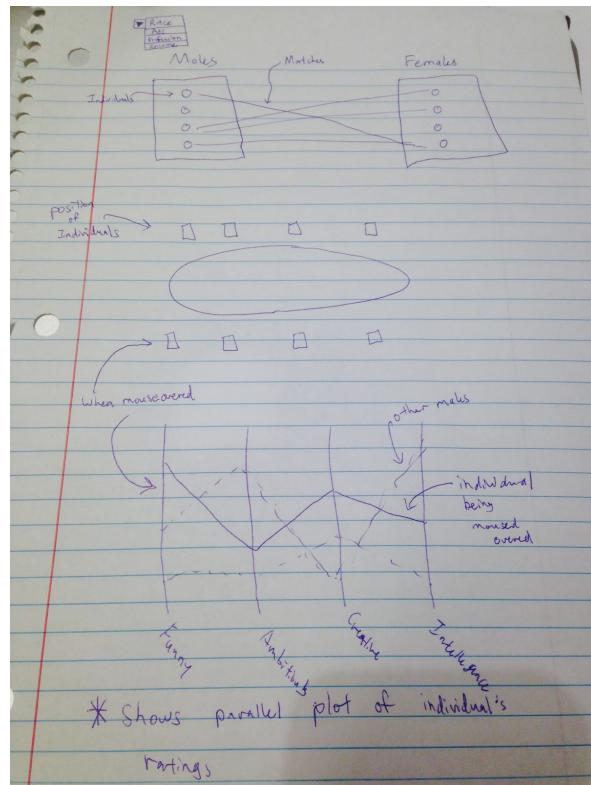
They had some initial reservations concerning our visualizations. They were unsure about the sheer number of visualizations we had proposed and how we would incorporate and interweave our visualizations. They thought the side-by-side bar chart was an effective way to visualize comparisons between two subsections of our data. However, they thought having three visualization would be a pretty big burden to the project. They also suggested that our third visualization may be unnecessary as the information it conveys (whether people follow-up on their matches) was fairly one dimensional and didn't tell us anything new or offer good options for interactivity. Lastly, they brought up position of participants as another interesting point to explore. In other words, does the position of the participants affect success or impression? This could potentially tie in to the force layout as nodes can also convey locational information.

Visualizations after Design Studio

Visualization 1



Visualization 2



We decided to simplify our visualization into two parts, looking at two different narratives of our data set: What the data says about speed dating trends in general (1) and what it can tell us about odd individuals (2).

- 1.) We thought a bar chart would be the best way to depict comparisons in general trends in speed dating like what men want in women and what women think men want. We would allow users to filter the data to allow for interactivity as well as switch between two views depicting preferences. A slider will show how preferences changed over time.
- 2.) The second visualization will have nodes and links in a diagram of all the individuals at one event. This allows us to explore any interesting outliers.
- 3.) The second visualization will be tied to a parallel coordinates chart which will plot how individuals rated themselves compared to the other individuals at the speed dating event.

Milestones

(April 3rd) - Initial project proposal. We found a data set we were all interested in (RIP Chess data set) and tried to come up with good visualization ideas. We decided to include a variety of bar charts, pie charts, force layouts, line charts, parallel coordinate charts, etc. In hindsight we realize that we were very unsure of what we wanted from the visualization

(April 7th) - Received initial feedback from TF. Some of his suggestions include:

- All the visualizations should be linked with each other. For example, highlighting/selecting in one chart will highlight/filter in all other charts
- Avoid pie charts
- Use a static layout instead of force layout because it is harder to scale with force layout
- Indicate position when partners met in order to see if position influences dating success

(April 14th) - Meeting after design studio. We discussed our goals with the visualization in more detail. We ultimately settled on focusing mainly on speed daters preferences and questions pertaining to that.

- What do men want in women?
- What do women want in men?
- What do men think women want?
- What do women think men want?
- Is there a difference in preferences if we filter the data by race, profession, intelligence level, etc.?
- Did individual's preferences change over the course of the events?

(April 15th) - We reformatted the data to better suit our visualizations. The code for reformatting the data is in the *index.html* in the *dataLoaded* function.

(April 23rd) - We met with our TF and discussed our progress. Some of his suggestions include:

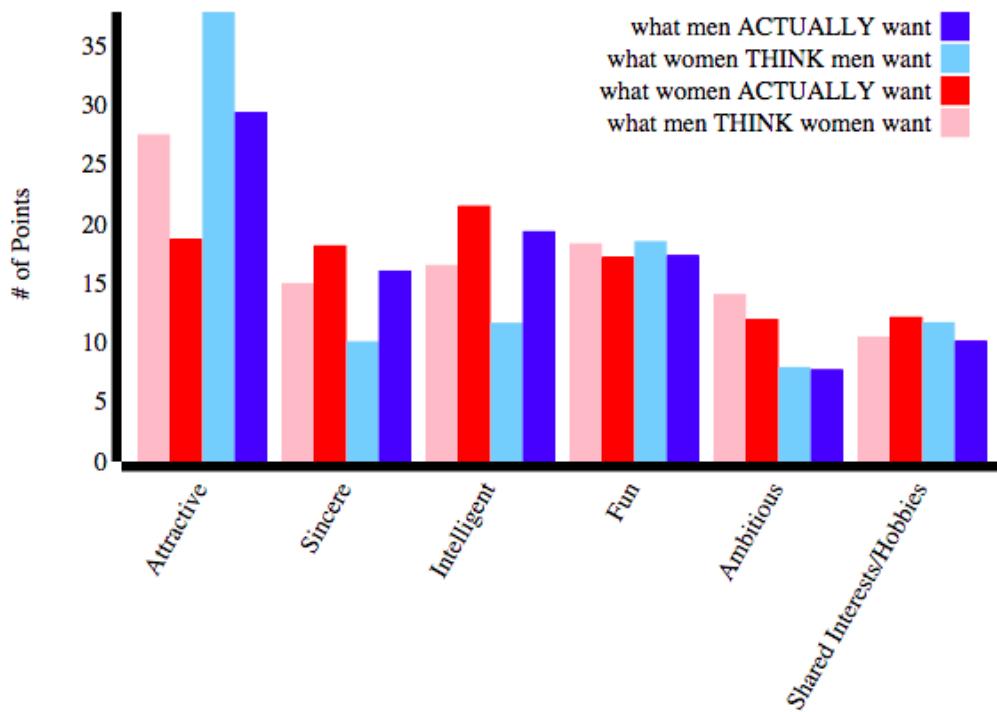
- Stacked bar chart is not displaying meaningful information
- Color links to show matches between individuals in node layout
- Make the parallel coordinates connect to node layout → if you click on a specific path, highlight the node
- Make sure all visualizations are connected to each other

Evolution of the Bar Charts

We decided to create a vertically grouped bar chart to compare the preferences for what one looks for in the opposite gender and what one thinks the opposite gender looks for. We chose to use a grouped bar chart because it allows for easy comparison across the different categories.

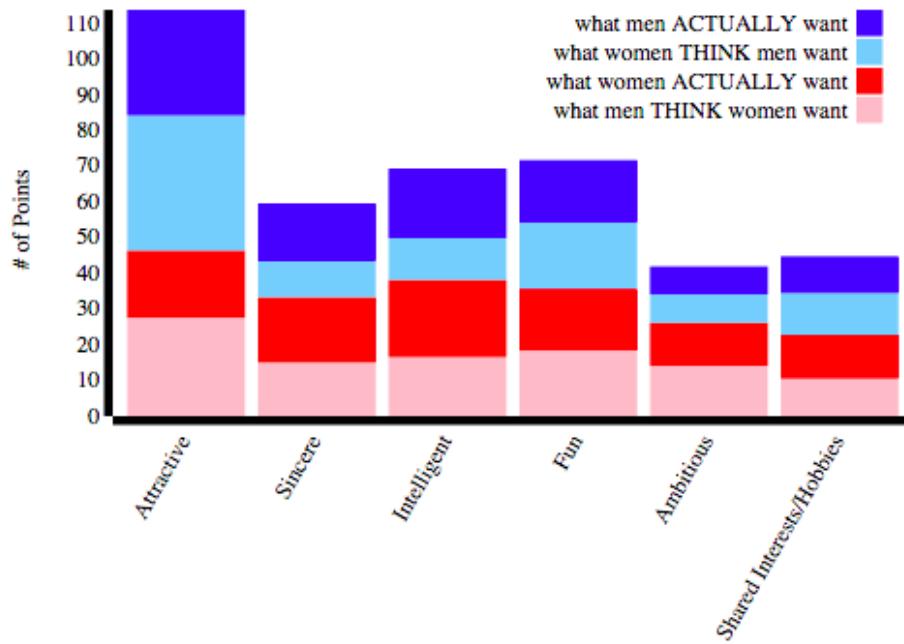
What data are we presenting in the grouped bar chart?

Participants were asked to distribute 100 points among 6 attributes, giving more points to those attributes that were more important to them in a potential date. On the x-axis are the 6 attributes, and on the y-axis are the number of points. There are four bars for each attribute: (1) what women want, (2) what men think women want, (3) what men want, and (4) what women think men want. We averaged the points for each attribute and presented that in the bar chart. There is a legend denoting the different categories with different colors.



What about a stacked bar chart?

Initially, we had included a layout for a stacked bar chart because we wanted to see which attribute was given the most weight.



However, after discussing about the stacked bar chart layout with our TF, we decided that the data shown in the stacked bar chart was not meaningful. As a result, we decided to remove the stacked bar chart layout from our visualization.

What do we want to filter?

We wanted to filter our data because we wanted to see if there were any trends in subgroups of the participants. We decided to filter our data by race, because we thought it would be interesting to see how the preferences for one racial group differed from the average or from another racial group. (We noticed that even though there was a category for Native Americans in the metadata, there weren't any participants who were Native American so we removed the filtering option for Native Americans.) We also wanted to filter our data by occupation because we thought it would be interesting to see how preferences for people in one profession differed from another. And finally, we also wanted to filter our data by goal. We thought it would be interesting to see if there were any differences in preferences depending on what people wanted to get out of the speed dating event. By including this filtering functionality, users can explore more trends and see if differences arise in subgroups.

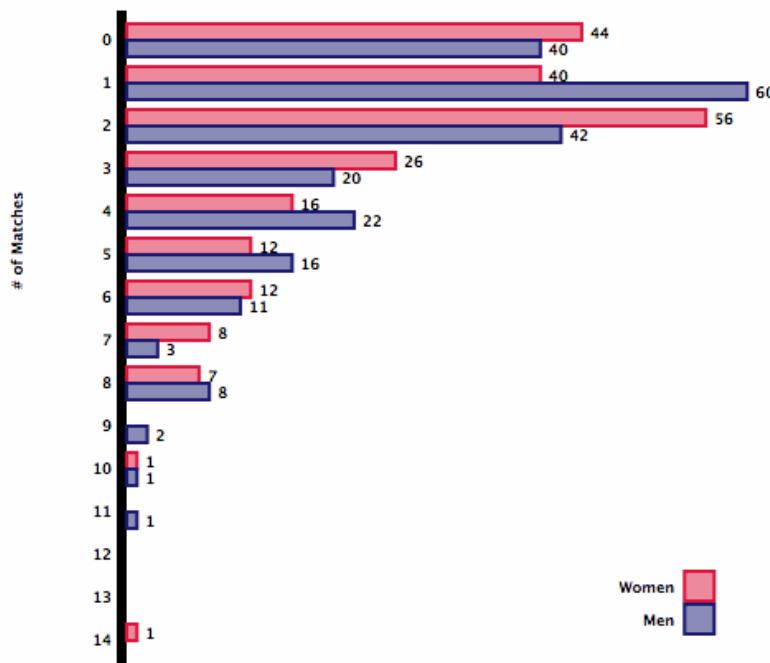
How do we know how many people is represented in the visualization?

We thought that the number of people represented in the visualization was important because it tells us if the data is representative for the particular subgroup. For example, if there was only one African American person working in Academia, the data would not show any trend. Rather it would simply show the preferences for that one person, which is not as interesting. However, if there were 100 Asians working in Business, the data would be more representative of what Asian Business workers look for in the opposite gender, and this is interesting. To inform users to what they are looking at, we added a “current selection” to tell users how many people were represented in the visualization.

Filter by Race: African American Caucasian-American Latino/Hispanic Asian Other
Filter by Occupation: Lawyer Academia Medicine Entertainment Business Humanitarian Affairs
Filter by Goal: For Fun Meet new people Get a date Find serious relationship To say I did it Other
current selection: 116/449 people

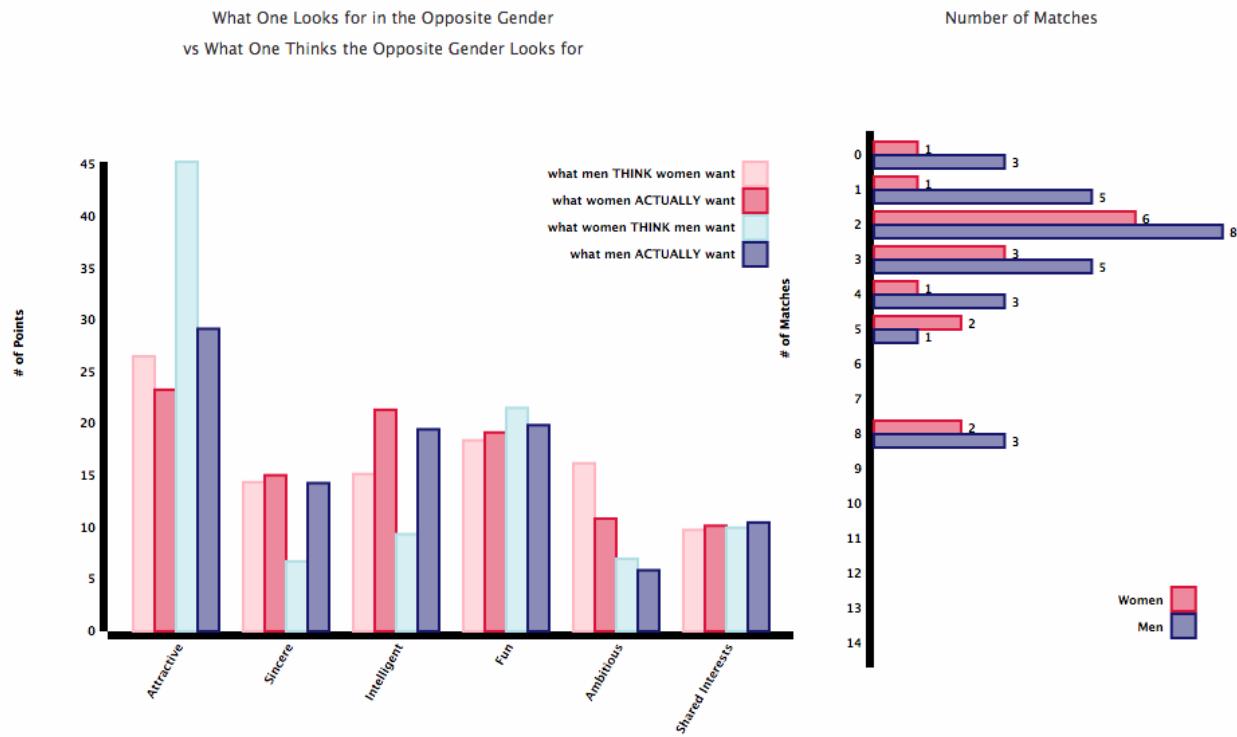
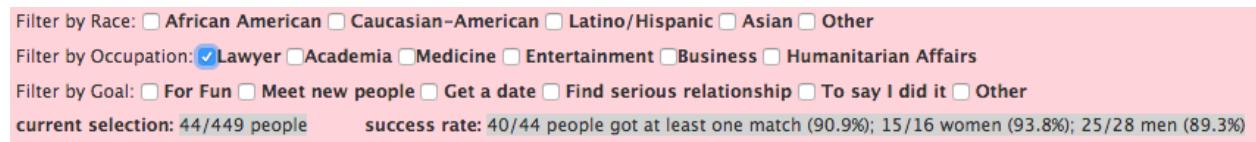
How do we visualize the number of matches and how successful speed dating is?

To answer the most interesting question, how successful is speed dating, we decided to create a second bar chart to visualize the number of matches for males and females. We chose to make a horizontally grouped bar chart because it allowed for easy comparison of the number of matches people got. On the y-axis are number of matches and on the x-axis is the number of people, so the width of the bar would represent the number of people who got “x” number of matches. In addition to the bar charts, we added a text element to show the number of people.



How do we know how successful speed dating is?

We thought it would be interesting to know the percentage of people who got a match, and the percentage of women and men who got a match. Therefore, we added “success rate” to tell users the three percentages.



How do the bar charts connect to other visualizations?

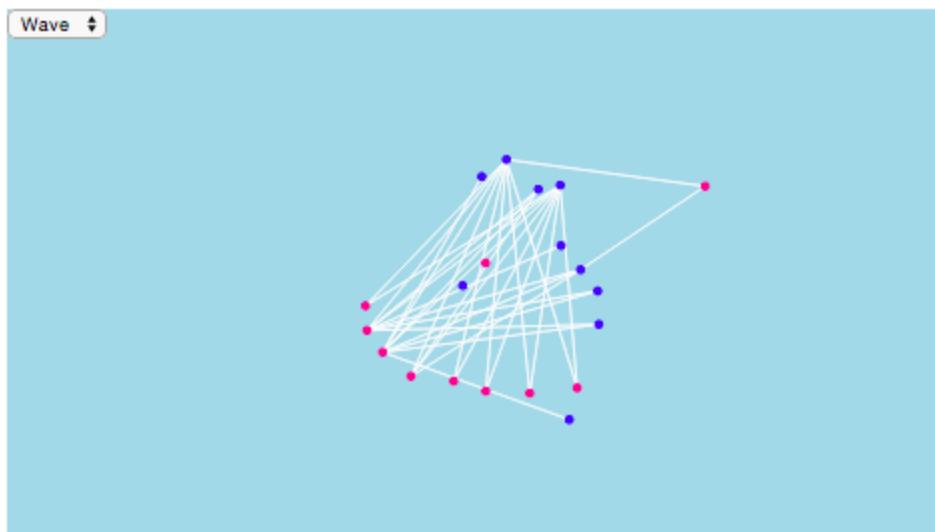
Because the node layout represents data for one wave at a time, there is a dropdown menu for the user to select which wave they want to see. When a wave is selected, the data for both bar charts are filtered so that only data for the selected wave is shown.

Evolution of the Nodes

The concept of using nodes began when we decided that it would be interesting to look at how specific people matched with each other in each speed dating session. This could allow us to spy outliers and see what characteristics allowed them to get many matches or no matches at all.

How do we visualize each individual?

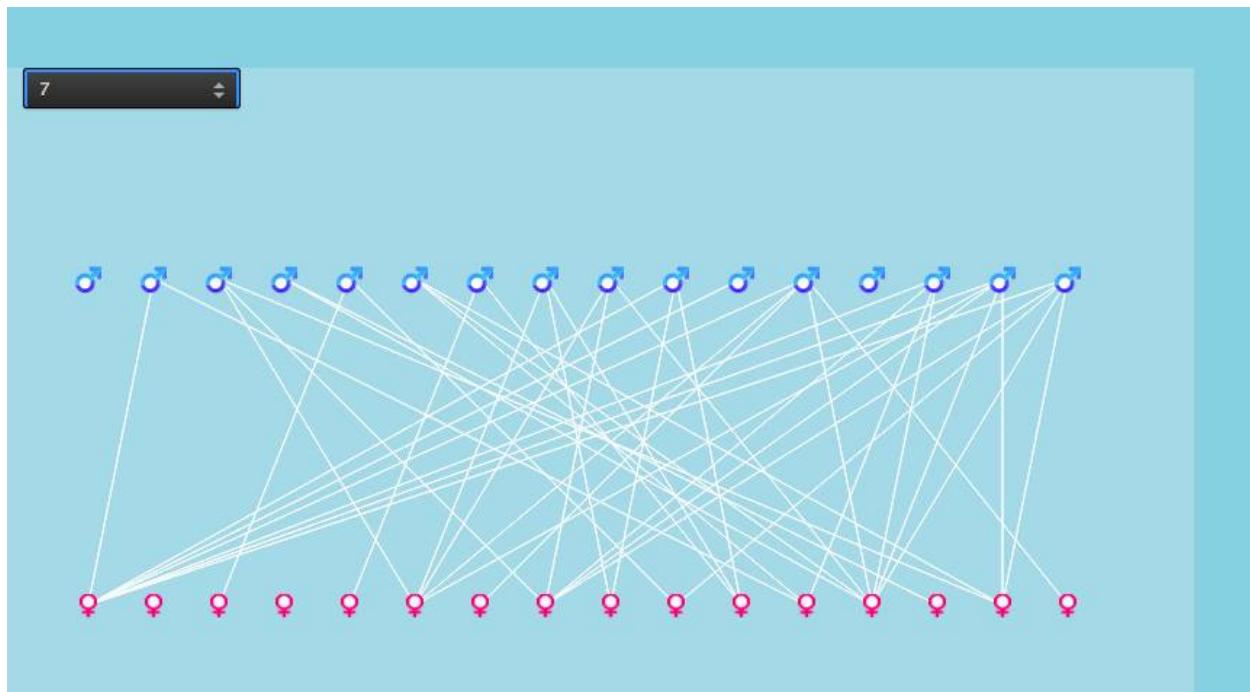
In the beginning, we decided to use a force-layout to visualize individuals in each speed dating session. Each blue node represents a male participant and each pink node represents a female participant. Links between nodes represent a mutual “match” between the participants. There is also a drop-down menu on the top-left corner where the user can select different speed dating sessions to visualize.



What is the significance of each node's position?

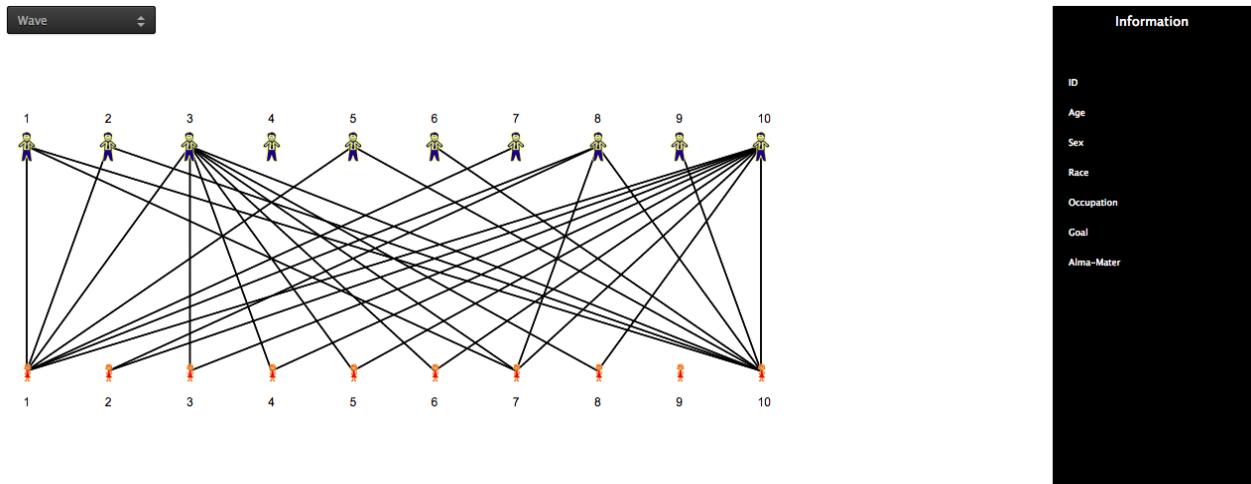
Unfortunately, having randomly generated positions for the nodes do not hold any meaning, nor does the interactivity of the force layout impart any new meaning. Thus, we decided to use the positions of the nodes to signify starting order of the speed dating session. This is done by switching to a static layout instead of a force-layout. In this case, all female nodes will be aligned to the bottom, and all male nodes will be aligned to the top. To show starting positions, all nodes are ordered based on their initial seating arrangement, with the first position on the left and the last position on the right. This grew complicated as some waves had uneven males and females.

To overcome this, we had to positioned some of the nodes such that they do not have a complementary partner. However, since each speed dating session functions by rotating the male partners while the female partners are stationary, the order in which they meet is still preserved in the nodes' positions.



How do we display individual information?

Since the point of having a static layout with nodes was to visualize individuals, we also needed a way to display the specific information represented by each node. Initially, we added a separate individual information box to the right of the node layout that holds certain fields: individual ID, age, sex, occupation, goal, and alma mater. When a node is selected, this box will be updated to contain the specific information of the individual. As for how the node is selected, we have two options: selection by clicking or selection by hovering.



What happens when a node is clicked and when a node is hovered?

Since these two actions introduce interactivity into the visualization, it is important to separate the functions of each. Thus, we decided to have the action of hovering over a node to update the individual information box. This is in part because displaying individual information is one of the more important parts of this visualization and because it is easier to hover over a node rather than clicking on a node. Since clicking on a node is still free for further functionality, we decided to implement an ordering function. As a result, clicking on a node now reorders the nodes of the opposite gender in order of when they met the node that was clicked.

node info	
ID:	14
Age:	23
Sex:	Male
Race:	Caucasian
Occupation:	Lawyer
Goal:	To meet new people
Alma Mater:	Undisclosed

How does the node layout connect with the other visualizations?

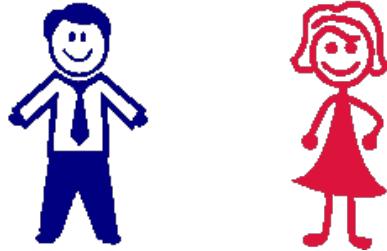
In total, there are three visualizations in this project: the bar graph which displays the general summary of the dating data, the node layout which reveals individual matches in each speed dating session, and the parallel coordinates graph which explores individual preferences compared to his or her peers in each dating session. In the beginning, selection of the speed dating session was isolated to just the node layout. After pondering on how to link the visualizations, selection of speed dating session was extended to the bar graph as well. When the visualizations are initialized, the node layout will begin by visualizing the first session. When a different session is selected, the node layout will update accordingly. On the other hand, the bar graph initializes with data from the entire dataset, but will update accordingly when a session is selected.

How does color signify information?

Color was initially used to differentiate male nodes from female nodes. After switching from svg circles to images, color still represented the binary sex code, but with redundancy since the images for each sex are distinct.



However, to make each node more appealing, we decided to switch to cartoon images



Nevertheless, we realized that other colors can be used to impart more information. With the realization that the bar chart's filters can also be applied to the nodes, we decided to use a new color for the node to indicate that those nodes fit the filtered criteria.

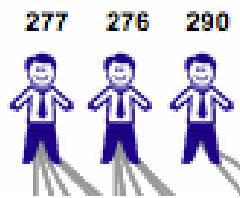


Additionally, when a line in the parallel coordinates graph is selected through clicking, that same individual in the node layout is also highlighted. This allows users interested in trends in the parallel coordinates graph to see which individual it is in the node layout. This also works in the opposite direction, where clicking on a node will display that individual's preferences and how others rated him or her in the parallel coordinates graph.

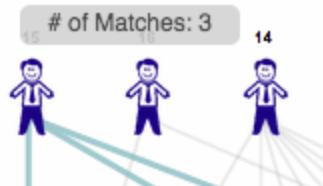


How do we keep track of the nodes?

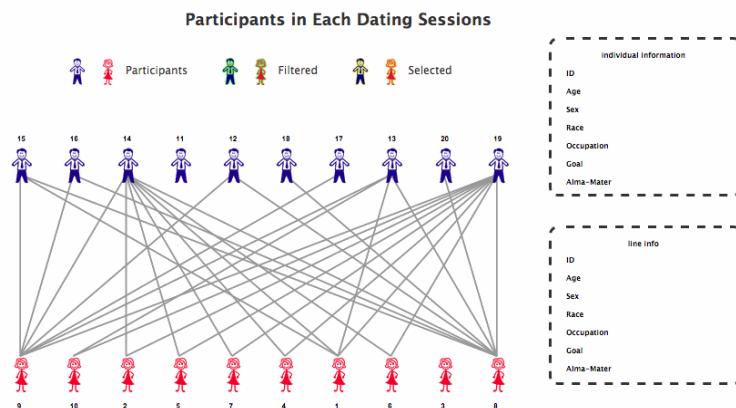
At the start, each node had no other identifying feature besides its male/female image. Ergo, we needed to add some identifying feature in order to recognize who's who in each session. We began by using a tooltip (which displays when hovered over node) to display each individual's ID. However, we realized that it would still be difficult to locate specific people if you have to hover over all the nodes. Thus, we agreed to display each individual's ID on top of each node.



This frees up the tooltip for further use. Since it was difficult to see the number of match links for individuals with many matches, we decided to use the tooltip to display the number of successful matches the individual has when hovered over.



Final Node Visualization



Evolution of the Parallel Coordinates

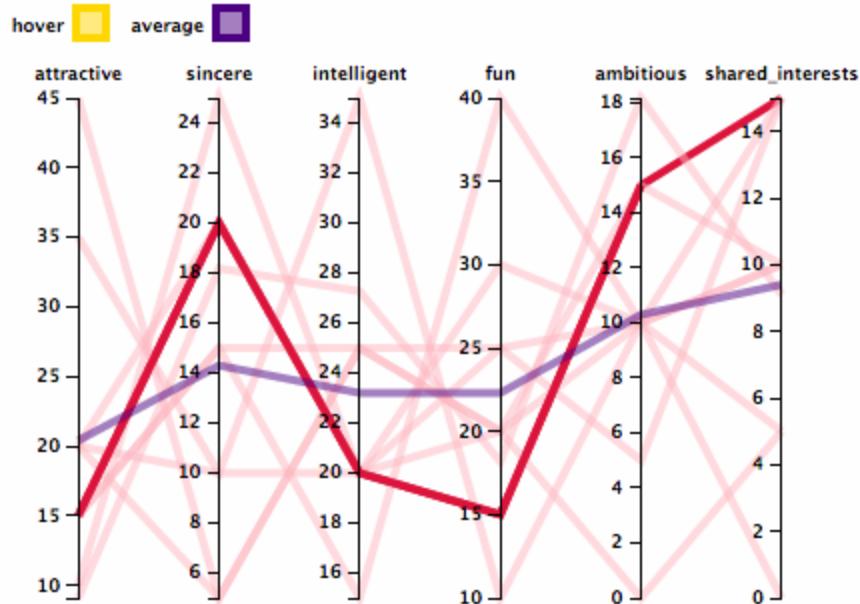
We decided to make two parallel coordinates to show (1) how the preferences of one individual compare to other individuals of the same gender and the average, and (2) how the individual's ratings of him/herself compared to how individuals of the opposite gender rated him/her. We thought that it would be interesting to see what makes an individual stand out from other people.

Some questions this visualization addresses are: (1) Does this individual weight attributes differently compared to individuals of the same gender? (2) Does the individual rate him/herself relatively the same as how people of the opposite gender rate him/her? (3) How are the ratings correlated with how many matches the individuals get?

What does the first parallel coordinate show?

When the page is first loaded, the first parallel coordinate shows the preferences of the females in the first wave. The female with an id of 1 is represented with a dark red line, while all the other females are represented with pink lines. The purple line represents the average of the preferences for females.

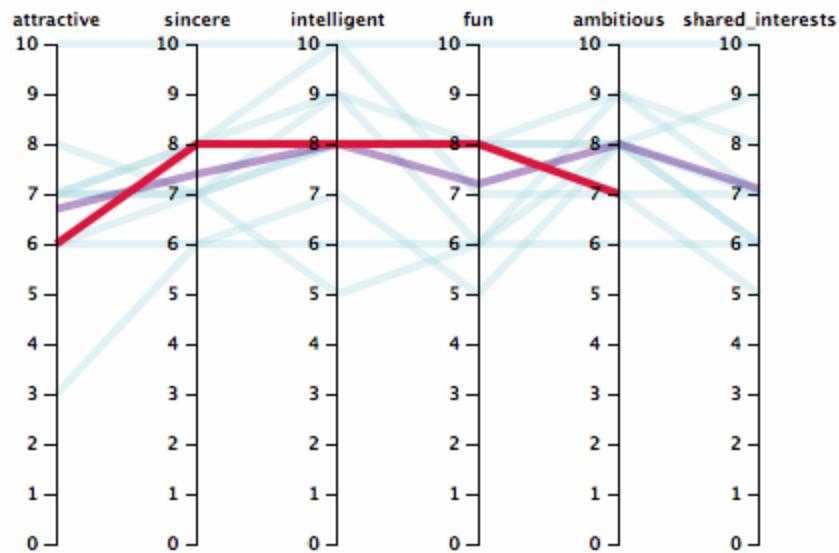
Personal Preferences vs Others of the Same Gender



What does the second parallel coordinate show?

When the page is first loaded, the second parallel coordinate shows the ratings (scale 1-10 with 1 being awful and 10 being great) of the 6 attributes for the female with id 1. The red line represents how the female rated herself, the blue lines represent how the males rated the female, and the purple line represents the average rating for the female.

Opinion of Self vs Impression Left on Others



How does this visualization interact with the node layout?

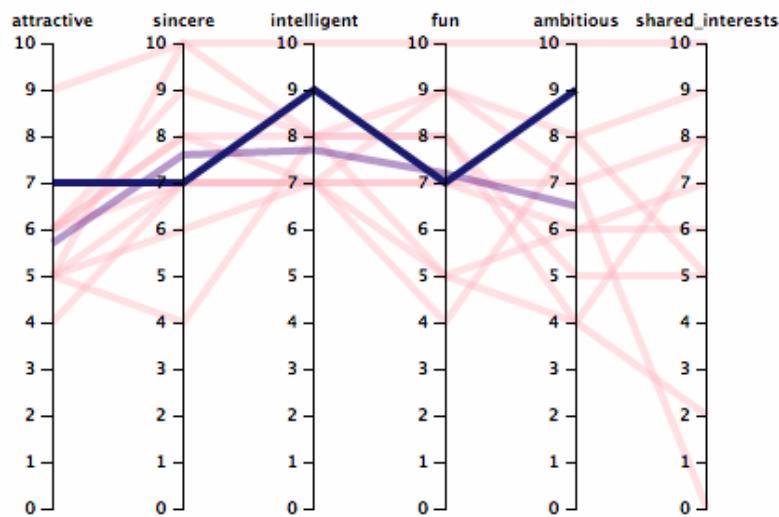
When a node is clicked, the parallel coordinates are updated to show the information for that specific node/individual. The first parallel coordinate will highlight the preference of that individual in dark blue if the individual is a male and in dark red if the individual is a female. The lighter lines of the corresponding color represent the preferences of those of the same gender as the individual. For example, if the individual clicked was a male, the individual would be represented in dark blue, and the other men would be represented in light blue.

Personal Preferences vs Others of the Same Gender



The second parallel coordinate will highlight how the selected individual rated him/herself in dark blue if the individual is male and dark red if the individual is female (same colors as the first parallel coordinate). The lighter lines of a different color represent how people of the opposite gender rated the individual. Continuing the example from above, if the selected individual is male, the dark blue line represents how the individual rated himself and the light pink lines represent how the females rated the individual.

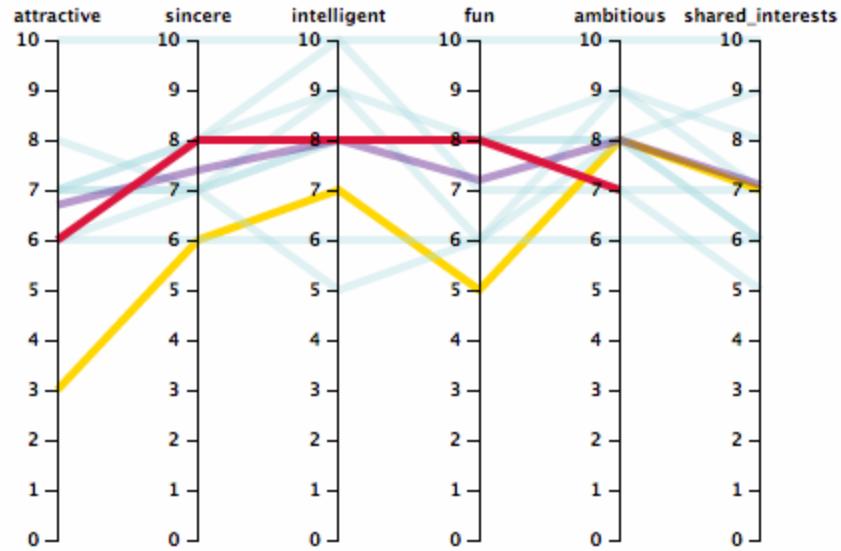
Opinion of Self vs Impression Left on Others



What happens when you hover over a line?

When you hover over a line, the line is highlighted in yellow so users will have an easier time seeing the data for the line that they are hovering.

Opinion of Self vs Impression Left on Others



What happens when you click a line?

When you click a line, the parallel coordinates are updated to center around the selected individual that the line represents. This is connected to a line info box that shows the information of the selected individual.



This is also connected to the node layout. When the line is clicked, the node corresponding to the individual represented by the line is highlighted in yellow.



3

How is this visualization connect to all of the visualizations?

When the user selects a specific wave in the dropdown menu, the parallel coordinates updates to show information pertaining to the selected wave. Also, the lines who fit the criteria of the filtering options will be highlighted in green.

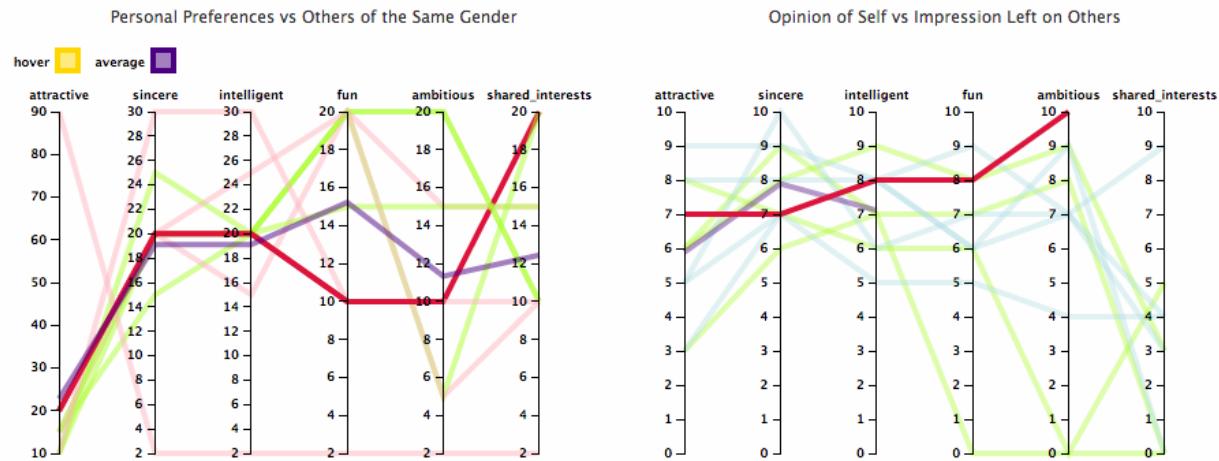
Filter by Race: African American Caucasian-American Latino/Hispanic Asian Other

Filter by Occupation: Lawyer Academia Medicine Entertainment Business Humanitarian Affairs

Filter by Goal: For Fun Meet new people Get a date Find serious relationship To say I did it Other

current selection: 8/449 people success rate: 8/8 people got at least one match (100%); 4/4 women (100%); 4/4 men (100%)

6



General Layout

Our website is divided into three main sections: (1) Header/Title, (2) Narrative, and (3) Visualizations, which is further divided in three sections (a) bar charts, (b) node layout, and (c) parallel coordinates.

The first section includes the title, a quick introduction to the website, and a short summary of what the visualizations are showing.

Speed Dating



Still single? Haven't been successful in getting a match at a speed dating event? Learn the secrets to success in speed dating here. Find out what the opposite gender is looking for and explore why some people get matches and others don't.

The first visualization shows you what people look for in the opposite gender and what people think the opposite gender looks for. (You can filter by race, occupation, and goal and the people who fit the criteria will be highlighted in the subsequent visualizations.) The second visualization shows you successful matches in a particular wave. (You can click on specific people and information about the person will appear on the side.) And the third visualization shows you (1) preferences of the selected person from the second visualization compared to people of the same gender and (2) how that person views him/herself compared to how people of the opposite gender view him/her. (Clicking on a line and will highlight the person in the second visualization.) Follow our narrative and click on the highlighted words to explore the relationships.

The second section includes the narrative. The user can click on the highlighted words and the visualizations will be updated to show the context of the narrative.

There were 17 total speed dating sessions beginning from October 2002 and ending on April 2004. In each session, males rotated through the stations while females stayed stationary. Some sessions were **highly successful** while in comparison, other waves were **total misses**.

Looking at **individual** preferences, there does not seem to be a discernible pattern to how individuals compare to others within their own speed dating session. While one person may value one trait higher than others, it forces them to weight other traits less. However, looking at an individual's opinions about themselves leads to some interesting results.

Individuals who attend speed dating events tend to have high **self esteem** rarely rating themselves below 5. People are also very **forgiving of others**, rarely rating others below 5 as well.

Also of note, is that individual's who receive no matches often rate themselves much higher than what their fellow speed daters rated him or her. Take for example the **one person** in session 10 who got no matches or the **only dater** in session 11 who didn't get a match. Unlucky or unlikable?

Of course, the biggest question we are asking is, who got the most matches? And the winner is **female #524** of session 17, who successfully matched with 14 other participants! What did she value the most? **Intelligence!** And what did other guys think of her? Most liked her, though 1 rated her 0 for every trait, while another didn't believe in her **sincerity** or **ambition**... I guess you getting 14 matches isn't ambitious enough.

The third section includes the three visualizations.



We chose to have three visualizations to give users the ability to see the general trends, in addition to information about specific individuals. The bar graph shows the general information of the participants and the success of the . The node layout shows information for specific individuals in each speed dating session. The parallel coordinates lets you further explore information about a selected individual. These three visualizations allows you to see general trends and explore any interesting specific individuals.