SPEED DATING

Witty Subtitle (work in progress)

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.

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:

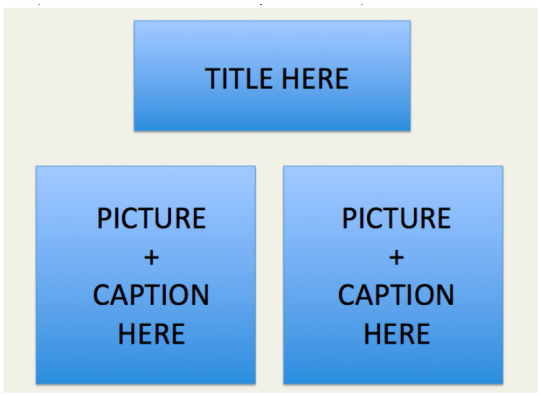
Our data came in as a csv file with each row representing one participant's meeting with another participant of the opposite sex. The columns contain the participant's personal information as well as their preferences and how they ranked their partner. In order to make our data easier to use, we filtered out the data to only include iid, age, gender, race, goal, career, wave number, partner preferences at beginning of event, preferences halfway through event, preferences a day after event, preferences a week after event, 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

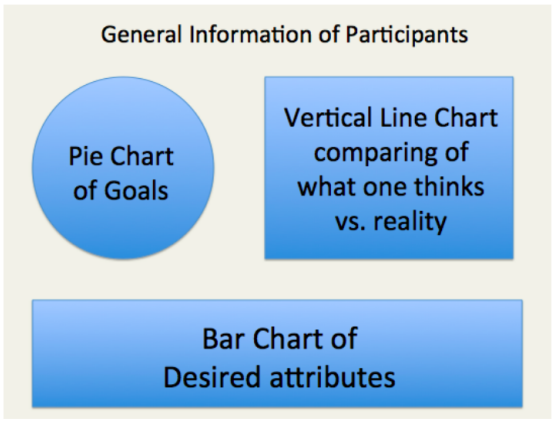
(April 3rd)

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. Thus, we want 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:

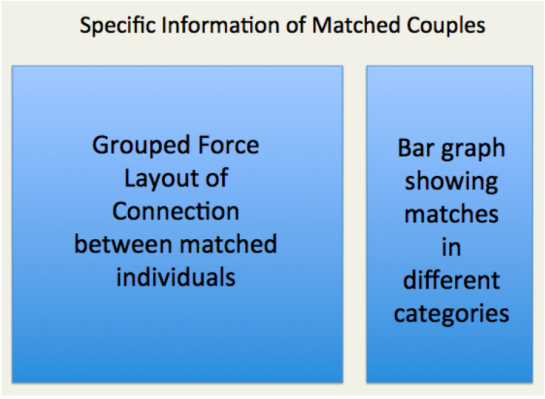
Visualization 1



Visualization 2



Visualization 3



Design Studio

(April 14th)

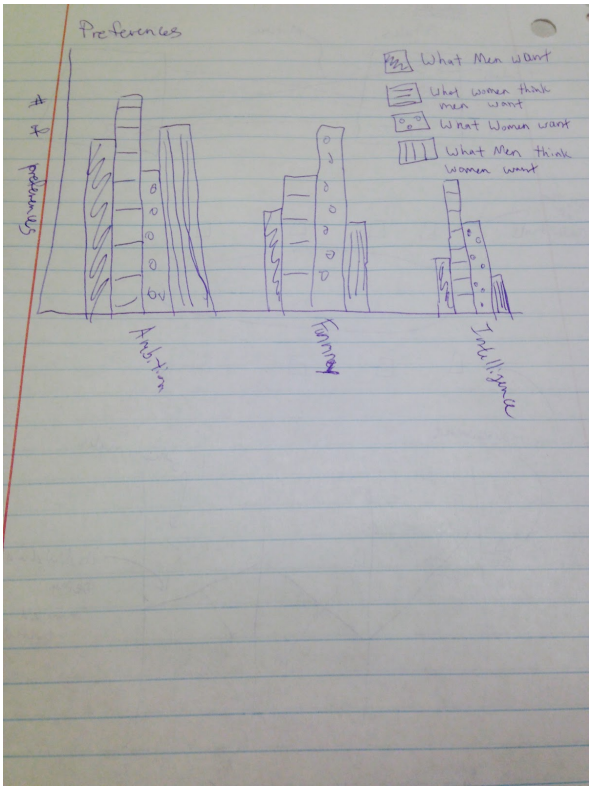
During the design studio, we met with another group to discuss our project proposal. After presenting our project, they had several suggestions. Many of the questions they had were what we were interested in:

- 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?
- Was speed dating an accurate representation of an individual?
- What variables affected speed dating success? (i.e. position, race, etc.)

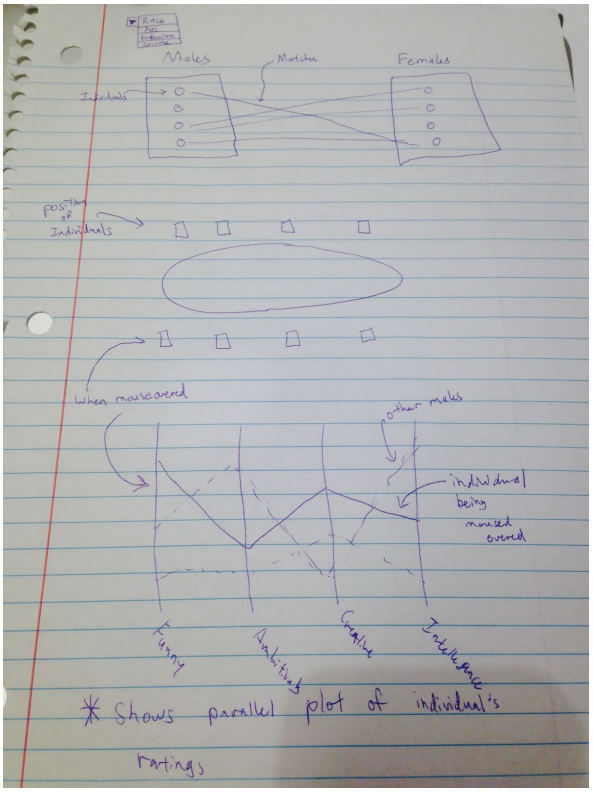
They had some initial reservations concerning our visualizations. They were unsure about the sheer number of visualizations we had proposed and how we would incorporate and interweave our visualizations. They thought the side-by-side bar chart was an effective way to visualize comparisons between two 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?

Visualizations after Design Studio

Visualization 1



Visualization 2



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

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.

The second visualization will be a nodes and links diagram of all the individuals at one event. This 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.

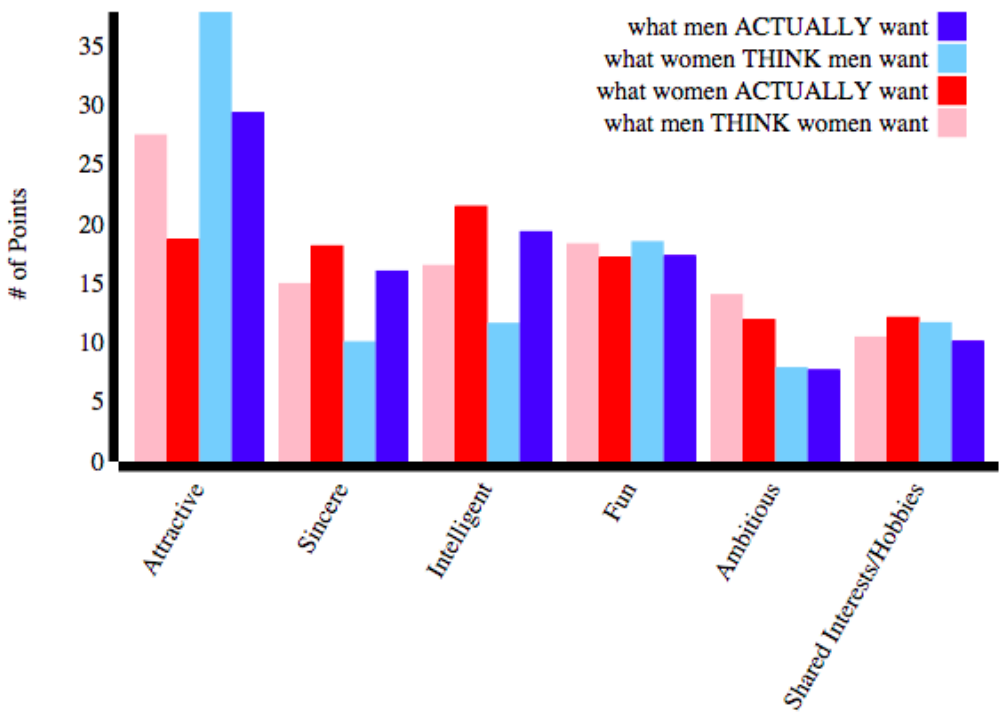
(April 14th) - Meeting after design studio. We discussed our goals with the visualization in more detail. We ultimately settled on focusing mainly on speed daters preferences and questions pertaining to that.

- 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?

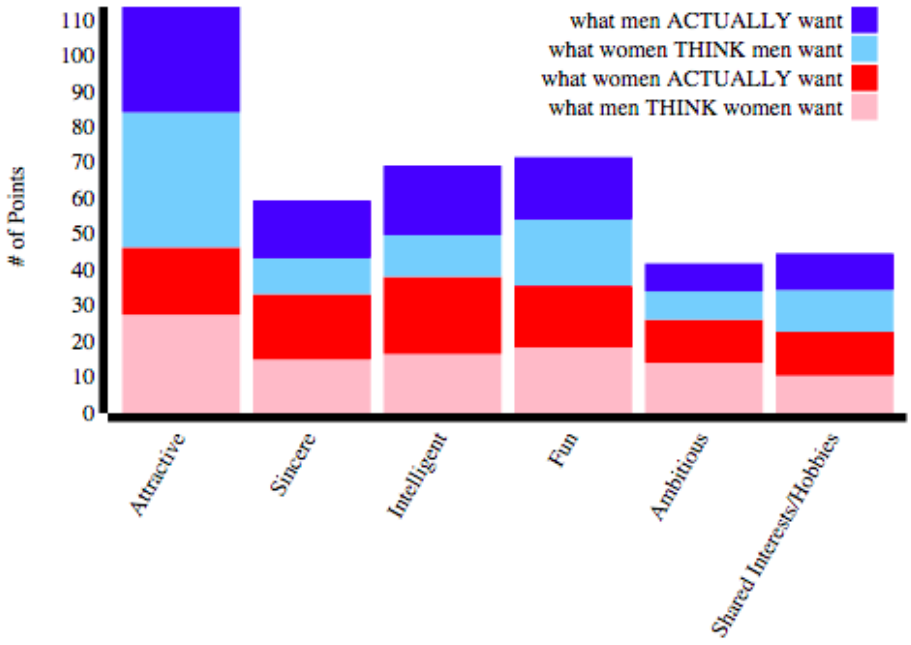
Unwilling to drop nodes and links we decided to

(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.

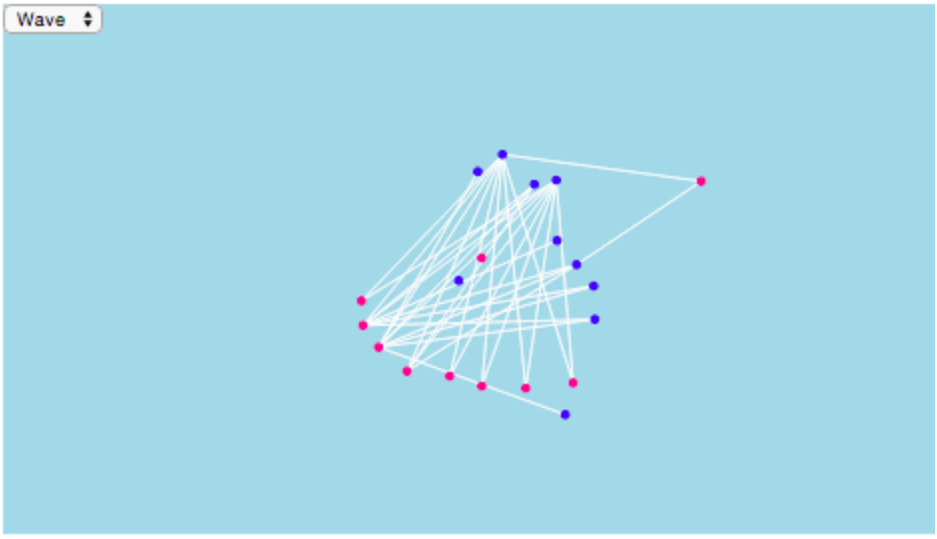
(April 16th-17th) - We started implementing the visualizations.
We worked on the grouped bar chart showing the preferences of men/women.



The next step in this visualization is adding an interactive functionality for the stacked layout. It will look like this.



We worked on the node visualization to see the matches between individuals in each wave (speed-dating session). There is a drop-down menu on the top-left corner where the user can select different waves to visualize.



We are in the process of aligning the nodes so that all the females will be on one side and all the males will be on the other. Additionally, we plan to use color codes to indicate who matches with who. For instance, when mouse-overed a node, a red link would mean the individual likes the partner, but not vice-versa, while a green link would mean mutual attraction. Lastly, this visualization will be linked to a third parallel coordinate chart visualization that is still under development. When a node is clicked, a parallel coordinate chart will appear and display the individual's self assessment as well as how the other partners rated him/her. In addition, a popup will appear to give the demographical information about the individual.