Speed Dating Experiment: A Visual Analysis of Speed Dating Data

OMIS 473

Keith Earnest, Allison Marshall, Jeff Willer

Purpose

As the amount of data collected by businesses and organizations continues to grow, easy to understand analysis techniques are required to read, interpret, and share the most important factors within this massive amount of data. Over time, one technique has become the quintessential method for turning this data into easily understood, actionable information to be used in the decision-making process—data visualization. Throughout the course of the fall semester, the students in OMIS 473 have learned a variety of techniques for analyzing and presenting large datasets through the use of data visualizations. As discussed through this course, data cannot be simply thrown into a chart and given to an audience with the expectation of clarity; rather, data analysts must consider the context and story to be told prior to creating any visualization.

Through this project, our group will explore the full story to be told by an extensive data set—the Speed Dating Experiment—and show how this story can be communicated to an audience through a variety of visualizations. We will begin by providing an understanding of the dataset in terms of background and variables. Then we will explore nine of the common visualization types in the context of our story to highlight both important factors within the dataset, as well as the key benefits of the individual chart type. Each visualization will present valuable information from the dataset in a way that is impossible with a simple spreadsheet and will be created with considerations to the design guidelines as instructed in the course. Finally, we will use these visualizations to gain intelligence from the dataset and present these conclusions.

Dataset

Background: What is Speed Dating?

Before the era of online dating sites and dating mobile applications, Speed Dating was a highly popular method of finding a partner or just meeting new people. This craze is thought to be introduced Antony Beilinsohn, who was encouraged by his Rabbi to create new events for young, eligible Jewish adults to meet after a declining trend in marriages within the faith. These Jewish-only events began in Los Angeles, California in early 1998 and were seen as successful by those involved. (Brown, 2003) The news of this success began to spread, and the first open, non-faith-based event was held in 1998 in Beverly Hills, California. These events soon began to spread to other demographic groups across California, until eventually becoming a well-known event across the globe. Although this event has become less popular since its initial craze, speed dating events can still be found in most areas throughout the course of the year.

Speed dating generally involves an equal or near equal number of men and women who are rotated around to meet every other participant over the course of the event. The dates themselves vary from two minutes to ten minutes depending on the organizer and the overall goal of the event. At the end of each date, each participant indicates which partners with whom they wish to exchange contact information. The organizers will then send a list of matches to every participant following the event.

Experimental Method

In early 2002, Columbia Business School professors Ray Fisman and Sheena Iyengar devised an experiment to gauge how men and women perceive the opposite gender during the dating process. (Montoya, 2016) To do so, the experimenters held twenty-one speed dating sessions from 2002 to 2004. These sessions were comprised of four-minute speed dates with all the participants of the opposite sex meeting each other only once, which varied between nine and twenty-two dates depending on the session. . (Montoya, 2016) Prior to the first date, each participant filled out a survey including demographics, attribute preferences, self-assessments, interests, goals, and other identifying information relevant to the dating process. Following each individual date, the both participants filled out an additional survey scorecard. In this scorecard, each participant ranked their partner on attractiveness, sincerity, intelligence, fun, ambitiousness, shared interests/hobbies, how well liked they were, partner's probability of saying yes, and their overall match decision. The day after participating in the speed dating event, the experimenters sent the participants another survey, requesting additional information on the overall perceptions of the event. One final follow up survey was sent three to four weeks after receiving a list of their matches, which included questions about any contacts or dates caused by the event.

Types of Variables

Our group obtained the dataset from Kaggle.com, which has all of the data compiled into a .csv file with a single sheet spanning 119 total variables and 8,379 observations. After an initial exploration of the spreadsheet, we noticed that the dataset included zip code information, but did not have any associated geography hierarchy. This would have prevented us from doing any useful geography-based analysis, so we downloaded a supplemental United States Zip Code excel file, which mapped zip codes to their associated city, county, state, and country. After importing both these files into Tableau, we used an inner join to combine the relevant information from the two files. As the dataset has an extensive number of variables, we focused our analysis on a few key variables, which we will detail below based on variable type.

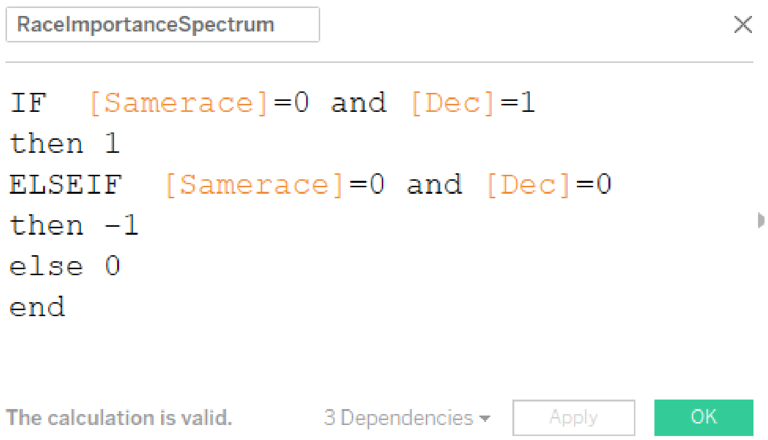
In our visualizations, we analyzed a variety of categorical variables. While the dataset includes more variables of this type that could be analyzed, doing so went beyond the scope of this project. The categorical variables and definitions used in the analysis are as follows: gender, which indicates the participants gender as male (1) or female (0); match, which indicates if both participants indicated yes (1) or one or both of the participants indicated no (0); race, which indicated the race of the participants; dec and dec_o, which indicates a yes (1) or no (0) for the participant and the participant's partner respectively; field_cd, which provides a coded field of study for each of the participants; and career_c, which indicates the career of each of the participants as it falls in one of the major categories. For the sake of the visualizations, we created aliases for these categorical variables, as they were coded with dummy values for the sake of statistical analysis by the experimenters. We used the data key provided with the dataset to accomplish this manually.

For our analysis, we use position as a time-based variable. The purpose of position was to indicate the order in the overall date lineup in which they met their partner. As speed dating involves the participants cycling from position 1 to the final position of the session, this can be used in place of a time variable in Tableau. Doing so allowed us to analyze any trends in position to see any inherent bias caused by being the first or last date of the night. We accomplished this by creating a custom time field using Day() and position, which will be seen in the visualizations below.

We also employed a variety of continuous variables in our visualizations. While the dataset includes more variables of this type that could be analyzed, doing so went beyond the scope of this project. The continuous variables and definitions used in the analysis are as follows: attr_o, amb_o, fun_o, intel_o, shar_o, and sinc_o, which indicate how the participant was ranked in the attributes of attractive, ambition, fun, intelligent, shared interests/hobbies, and sincere respectively by each respective partner; attr1_1, amb1_1, fun1_1, intel1_1, shar1_1, and sinc1_1, which have a combined total of 100 and indicate the importance of attractive, ambition, fun, intelligent, shared interests/hobbies, and sincere respectively for the participant; attr2_1, amb2_1, fun2_1, intel2_1, shar2_1, and sinc2_1, which have a combined total of 100 and indicates what the participant thinks the opposite gender rate attractive, ambition, fun, intelligent, shared interests/hobbies, and sincere respectively for the participant; imprace, which indicates on a scale of 1-7 how important race is in the opposite sex; numdat_3, which indicates the number of dates the participant has gone on with a match in the time after the speed dating event; and like_o, which indicates how much the partner liked the participant overall. For each of these variables, we changed the aggregation method to average rather than sum. This allowed us to perform a more accurate analysis that would not be impacted by a high variation in the number of categorical variables for which it was aggregated. For example, the speed dating dataset included a larger number of Caucasians than any other race. Any visualization measuring the count of matches by race would be skewed by the sheer number of Caucasians in the data.

As detailed in a prior section, the dataset originally included zip code information, but it did not have any associated geography variables that could be use for analysis. We joined an additional dataset in Tableau to add state, county, and country to the zip code data. Thankfully, the formatting between these two data points was identical, so we did not have to perform any transformation to join the datasets.

To supplement our analysis, we created a few calculated fields within the dataset. First, we created a RaceImportanceSpectrum variable, which was created to measure any relationship and bias in individuals who indicated they did not place importance on race, but consistently said no on partners of the same race. This field was calculated as follows:



We assigned a positive weight to saying yes to someone of a different race, and a negative weight to saying no to an individual of a different race. We then aggregated this based on the average to indicate age brackets, races, and locations that had a higher than average amount of declining dates with someone of a different race.

We also created a calculated field to measure the average overall MatchRating for each individual. The initial data set was prone to being slightly skewed in the average match and average decision based on the tendencies of each individual. Some participants may say yes to everyone and get more matches as a result, and vise versa. This was calculated as follows:

The MatchRating was used to even out the ratings for people who may have only said yes to a partner a few times and had less matches as a result.

Our analysis also included a couple of calculated fields to highlight different aspects of the dataset, including attractiveness and intelligence. These calculated fields are explained more in depth in the analysis section below.

Descriptive Statistics

We used Tableau's table function to gain a few basic details on the overall Speed Dating Experiment Dataset. Below is a breakdown of the participants average age, median age, minimum age, maximum age, and standard deviation of age by gender.

gender	Avg. Age	Median Age	Min. Age	Max. Age	Std. dev. of Age
Female	25.99	25.00	19.00	55.00	3.68
Male	26.44	26.00	18.00	42.00	3.64

These values give help provide a basic understanding of the participants and who may be considered an outlier when doing later analysis. As seen in the table above, both Males and Females had outlier participants of an age at least four standard deviations away from the mean. This method was also used to explore the datasets racial breakdown, as seen in the table below:

gender	African American	Caucasian	Latino	Asian/Pacific Islander	Other
Female	202	1,979	287	701	217
Male	158	1,757	147	642	116

These values helped provide some of the overall context of the dataset, so we understood later in our analysis how the disproportionate number of Caucasians may impact some of our visualizations.

Relationships Among Variables

Explicit Relationships

While not discussed in the variables above as we do not use them for analysis, there is an explicit relationship between igd (the unique id within the session) and id (the unique subject number for the experiement) and as igd is determined by the id. Other explicit relationships may exist in the 119 variables within the dataset, but a full exploratory analysis of this data was beyond the scope of the project.

Implicit Relationships

In order to investigate the types of visualizations we wished to create for this dataset, our group had to determine some of the implicit relationships within the dataset. These relationships are not as simple to determine, and they often require statistical or visual analysis to understand and interpret the relationship. For example, our data set has a variable for the importance of race as well as whether or not the participant and the specific partner were of the same race. We can analyze the data in more depth to determine the extent of the implicit relationship between the individual's importance of race and whether or not they matched with a partner of a different race. In addition, the dataset includes a large number of variables on the different rankings for the different attributes as detailed above. A relationship between attributes and gender, attributes and race, or attributes and age can tell us the different preferences and trends amongst different demographics. Furthermore, the relationship between position and decision can tell us if there is a better chance of getting a date at the beginning of and event, in the middle of an event, or at the end of an event. Our data has a significant number of implicit relationships as it has a large number of columns and observations, but we will explore a few of the relationships listed above and more in our visualizations below.

Justifications for Visualizations

In our analysis, we used the following visualizations to explore the Speed Dating Experiment data: radar chart, box and whisker plot, geomap, treemap, lollipop chart, motion chart, bar/column charts, donut chart, and stacked column chart. Each of these visualizations were selected for the specific advantages they bring in displaying complex data. Below, we detail the purpose of each of the visualizations and why it was selected for use.

One disadvantage of Tableau is that it does not have the capability of creating simple radar charts, so for this chart type, we used SAP Predictive Analytics to create the visualizations. This chart type allows an individual to clearly plot comparison information that may be too cluttered or complex to be read in a bar chart. Therefore, we used this external software to plot each of the attributes that share the same scale (ambition, attractive, intelligent, shared interests/hobbies, fun, sincerity) on their own individual axis to be displayed by gender. This chart gives the audience a clear understanding on where the genders differ in their desire for each of the six individual attributes, as well as how each gender perceives what the opposite gender desires on the same scale.

box and whisker plot

The box and whisker plots were used to display the interquartile range for the attractiveness and intelligence for each gender. In our initial radar chart, we saw that intelligence was ranked the most important to women and attractiveness was ranked the most important to men, so we wanted to further analyze this trend across all ages. The box and whisker plot can provide the audience a clear understanding on the average perception of males and females across all ages, as well as see any key outliers within this scale. We used a calculated field to create additional analysis in how one's own level of attractiveness and intelligence changes his or her desires in others. Color coding this makes this aspect clearer to the audience.

It is well known that the views on race, gender, and politics differ greatly across the United States. We wanted to use a GeoMap to explore if any of the inherent regional bias were indicated in the dataset in regard to race. By color coding this map both by the average importance of race (imprace) and adding a circle to show the average RaceImportanceSpectrum field, the audience can clearly see any state or region that rated race as highly important and frequently said no to partners of the same race.

For the treemap, we wanted to visualize any differences in identified race and importance of dating someone in the same race. We used the average importance of race (imprace) as the size factor and color coded by the average overall decision factor (dec). This provides a clear understanding of which races emphasize the same race the most based on clear size differences. The color coding adds an additional level of analysis by exploring if the individual races said yes or no to the overall partners more or less than the average.

In our Lollipop chart, our group wanted to explore the racial importance an additional time broken down by different age groups. This chart provides a unique view of the overall racial importance by having a full negative and positive spectrum centered around the zero line for the race important bar chart aspect. We added additional depth to this chart by adding the overall emphasis placed on attractiveness to display any correlation between racial preferences and attractiveness preferences. The lollipop chart, tree map, and geomap were tied together in a dashboard to allow exploration of each of these factors by geography, age, and race independently.

Our group employed a motion chart to visualize any trends in how the order of the date impacted the decision-making process. We wanted to visualize the differences in being the first date of the night versus the last date of the night. Each circle in the motion chart represents the average decision made by the partner at that phase, and the graph is highlight by the average decision by the participant at that stage. This visualization can clearly demonstrate to the audience which positions are best to improve the success of a date.

In our visual analysis, our group used a few bar charts to display categorical variables along a specific measure. One visualization used a bar chart and the MatchRating calculated field to highlight the overall success of the individual by age. This chart is tied to the position motion chart in a dashboard to allow further exploration in the success rating for each individual position in the speed dating session by gender. Alone, this chart can easily indicate to the audience the age groups that were the most successful overall in the speed dating event or break it down further by each individual gender. We also used bar charts to visualize the overall matches and match rating by career field. This provided a clear snapshot to the audience which careers were the most successful in the speed dating event by matches on average, or fields having a large bar, and color coded by the MatchRating calculated field indicating success rates.

Our analysis also used a stacked column chart to display a clear comparative exploration of how different career fields value the different attributes. As each of these attributes together were designed to add up to 100, this visualization clearly shows the audience the major differences between the top career fields. A radar chart could have also been used for this purpose, but we wanted to demonstrate how to satisfy a similar purpose using the tools integrated within Tableau.

Finally, we used a donut chart to aid in our exploration and presentation of the dataset. The donut chart shows the number of dates went on by some of the top fields within the dataset. While the other analysis focused on the short-term success of the event (matches and decisions) this exploration looked at what happened after the event itself. The inside of the donut shows a clear count of the number of dates, while the angle of the donut shows the number of dates gone on by individuals within this field

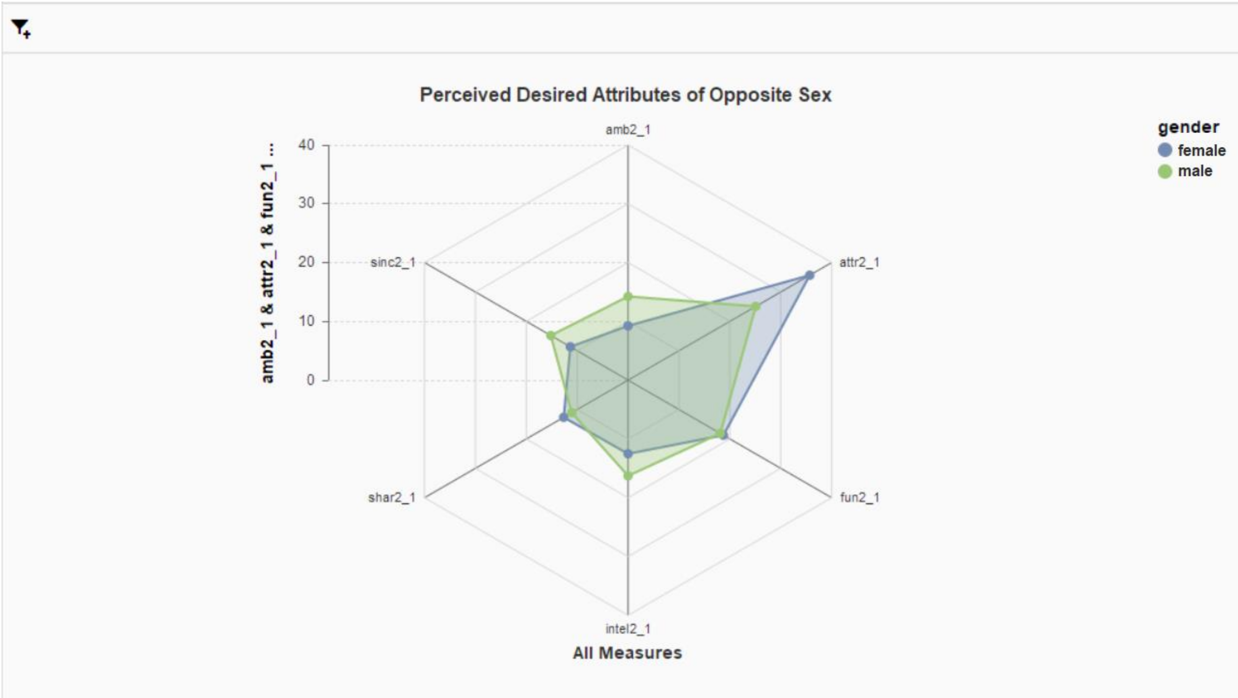
Visualization Findings

Radar Chart: Participants Desired Attributes

As detailed in the method and variable section, each of the participants was given a survey prior to the first date requesting him or her to allocate 100 points across each of the six attributes to answer the question, “What do you look for in the opposite sex?”. A higher value indicates a higher importance to that individual. In the visualization below, we can clearly see that there is a large disparity in the importance of attractiveness for men versus women. Men view attractiveness as the most important attribute by a large margin over the other five attributes. Conversely, women view intelligence as the most important attribute, but there is less variation in the overall ranking for the other five attributes.

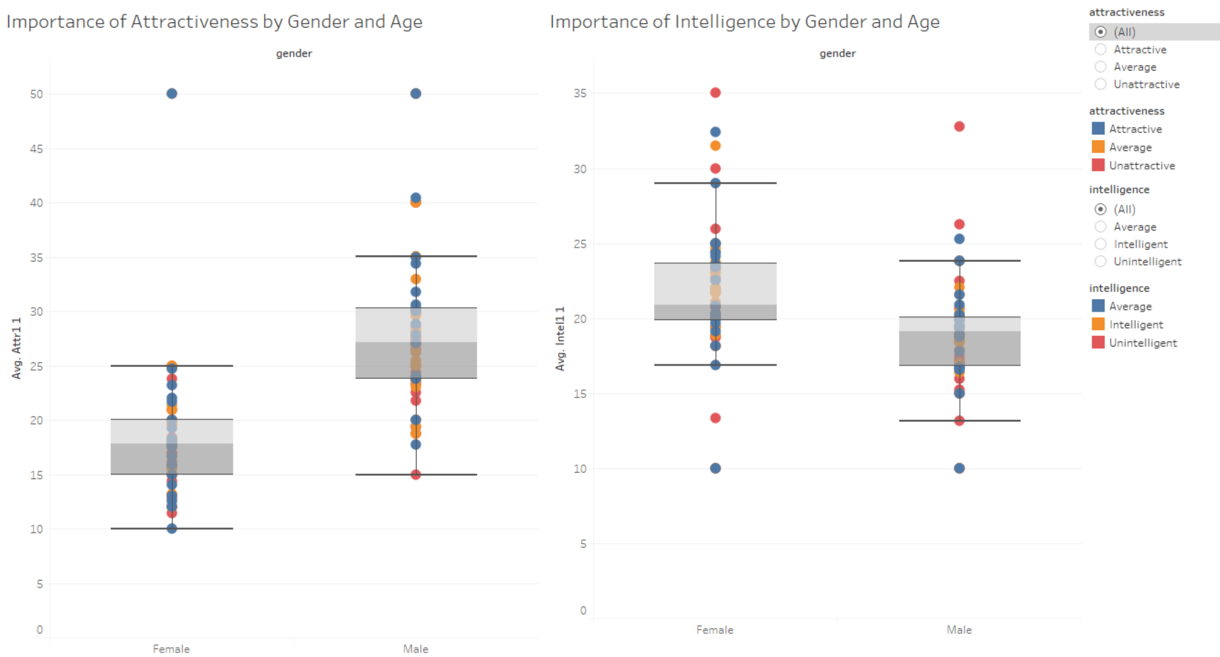
Radar Chart: Participants Perceived Desired Attributes of Opposite Sex

In addition to answering questions about what the participant themselves were looking for, each of the participants were asked to allocate 100 points across each of the six attributes to answer the question, “What do you think the opposite sex looks for in a date?”. Both men and women believed that the opposite sex valued attractiveness as the most important attribute in a partner. Men rated their perception of women’s interests at the same value on average as their own preference for that attribute. Across both the radar charts, overall shared interests/hobbies was ranked significantly lower than the other attributes. This comes as a surprise, as this would be one of the major conversation points for a speed dating event. Both of these charts indicate that personality does not play as great of a factor in male or female desires in the opposite sex, as attractiveness and intelligence were ranked highly for both genders.



Box and Whisker Plot (Dashboard): Attractiveness and Intelligence – All

From the visualization below, we can confirm the overall perceptions of males and females in the dataset in relation to intelligence and attractiveness in a partner. As discussed in the radar charts, men value attractiveness the highest while women value intelligence the highest. We can see in the visualization below that the interquartile range for men is significantly higher in the attractiveness attribute, while the interquartile range for women is higher in the intelligence attribute. What this visualization adds is the overall spread of this ranking across the dataset, as well as pointing out a few key outliers. First and foremost, far fewer women fell out of this average range when it comes to ranking attractiveness. This means there was very little variation in the responses for women when adding overall value to this attribute. Conversely, there was a large amount of variation when women ranked intelligence. We can see both positive and negative outliers from women, who fell outside the interquartile range for this attribute.

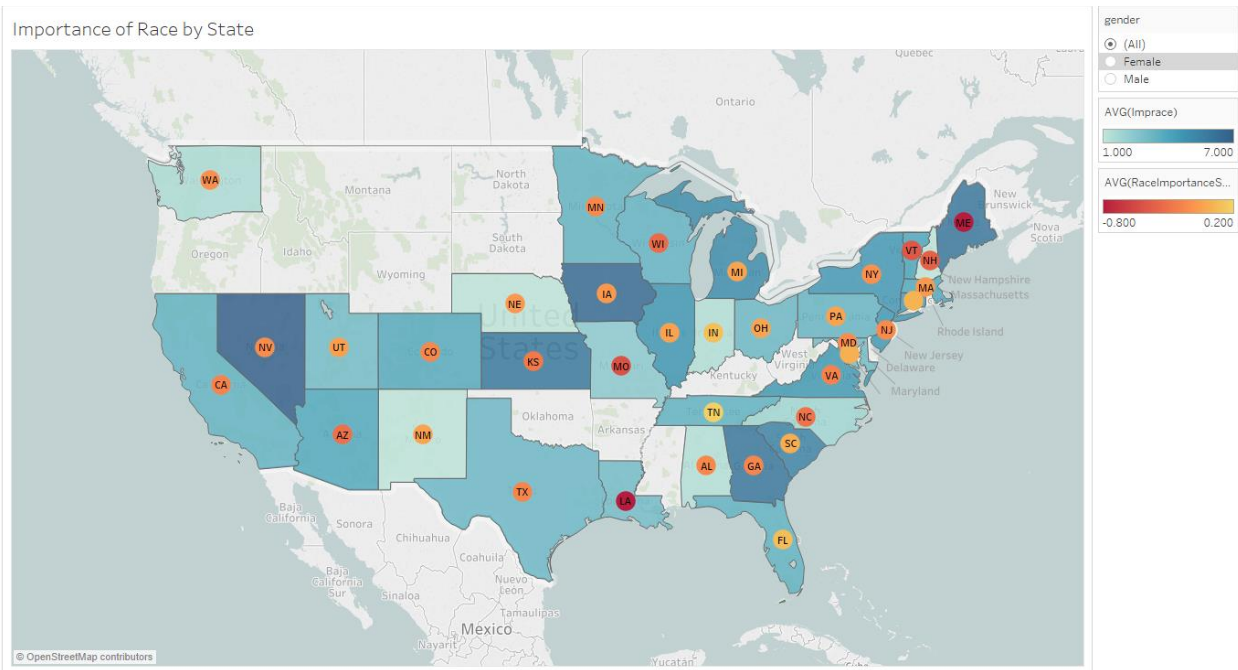


Box and Whisker Plot (Dashboard): Attractiveness and Intelligence – Filtered

While the unfiltered dashboard paints a picture of the overall dataset's perceptions, we were interested in investigating if there was any relationship between how an individual was perceived by others in attractiveness and intelligence and what they preferred for each trait respectively. To aid in our analysis, we created a calculated field to divide up how the opposite sex ranked the individual in attractiveness and intelligence into three categories. A rating of >8 in either attribute was rated as attractive or intelligent; a rating between 4 and 8 was rated average, and a rating below 4 was rated unattractive or unintelligent. When compared to the unfiltered dataset, individuals who are viewed in the bottom ranking for both attractiveness and intelligence tend to have lower expectations for the same trait in others. Even when factoring this in, we can still see the same preferences explored above for each attribute, with men valuing attractiveness more and women valuing intelligence more. The charts also highlight some of the interesting outliers within the dataset. As seen in the visualization below, some males who were viewed as unattractive by females desired extreme attractiveness (50 of 100 points delegated to this attribute) in women. A near identical outlier existed within the intelligence data, were a female viewed as unintelligence by males desired a high level of intelligence (35 of 100 points delegated to intelligence) in men.

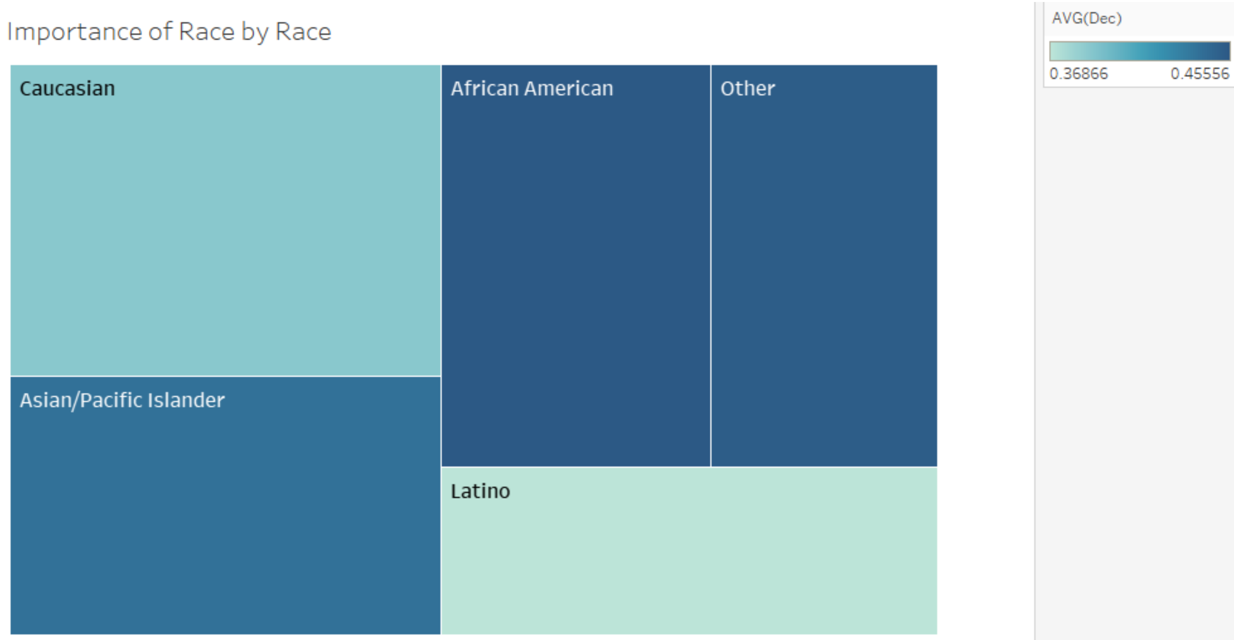
GeoMap: Importance of Race by State

It is a common perception that the Southern states in the United States hold more antiquated values towards race and religion when compared to the rest of the United States. We created the GeoMap below to explore any such bias within the dataset by comparing the importance of religion on average for each state. This alone was not a good indicator, as an individual may believe they are uncaring about religion, but they consistently pass on partners who are a different religion than their own. By adding the calculated field RaceImportanceSpectrum as detailed in the variable section above, we were able to highlight areas that both had a high level of importance placed on race as well as actively declined dates with those of a different race. We can see locations like Louisiana, who had a relatively low importance placed on race, that have an extremely high tendency to decline dates with individuals of the opposite race, contrary to their self-assignment. Other locations, such as Maine, ranked race as a highly important factor and this is also reflected in the RaceImportanceSpectrum.



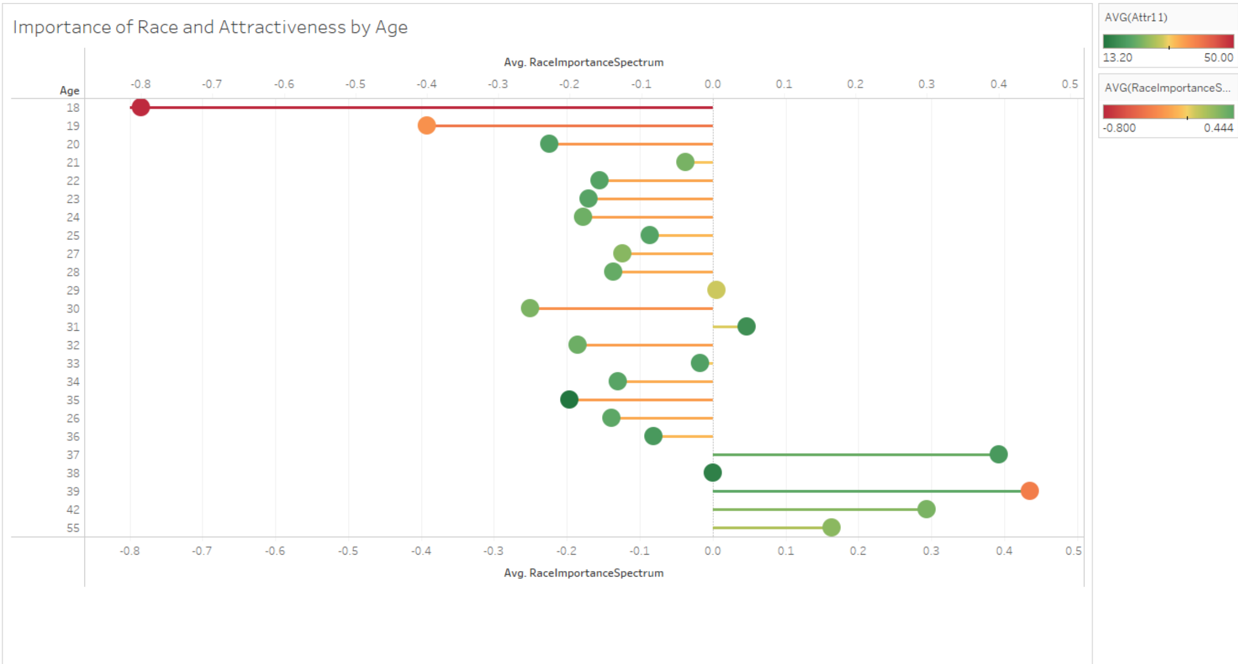
Treemap: Importance of Race by Race

Location alone is not enough of a factor to display racial tendencies within the dataset. We supplemented this with a treemap, which indicates via the size of the box the overall level of importance placed on race by each race. In the visualization below, we can see that Caucasians on average place the highest level of importance on race, and individuals who identified as a race other than those listed (other) had the lowest level of importance placed on race. This map was also colored by the average decision made on average, which indicates that Caucasians and Latinos said yes to a date the least on average.



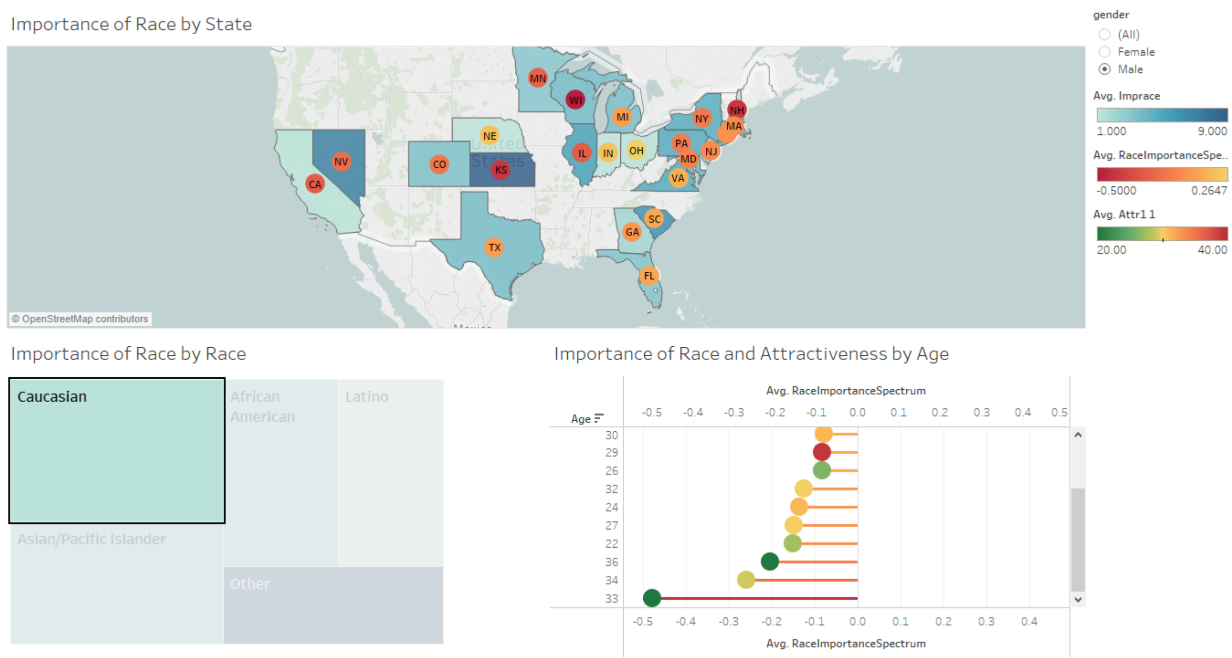
Lollipop Chart: Importance of Race and Attractiveness by Age

We wanted to add one more supplement to our analysis on racial bias within the dataset, so we created a lollipop chart to display the importance of race across the different age groups within the dataset. On average, younger individuals placed the most emphasis on race based on our calculated RaceImportanceSpectrum. This means they said no to a date more frequently when the partner was of another race. By also including the level of importance placed on attractiveness as the color of the circle, we can see a trend in the overall level of racial importance and the importance of appearance. The individuals with the greatest emphasis placed on attractiveness were also those who had the most racial bias.



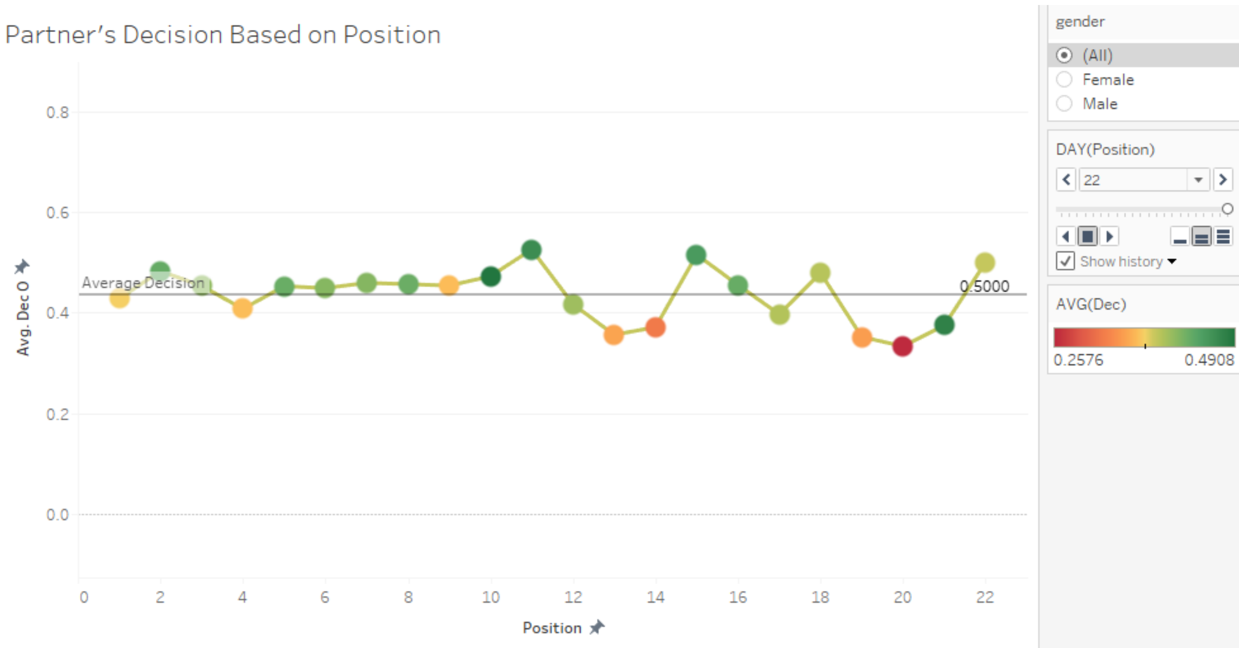
Dashboard: Racial Preferences for Dataset

We created an interactive dashboard to help us explore each of these factors in more depth. By adding filtering actions across the dashboard, we can drill into an age, gender, state, or race to see the overall breakdown of racial importance, RaceImportanceSpectrum, the other variables dependent on the selection. Take for example the dashboard below. In selecting male filter and the Caucasian tree map selection, we can see a breakdown of the racial importance for each state for this demographic, as well as a breakdown by age. For Caucasian Males, ages 33, 34, and 36 placed the highest overall importance on race as ranked by the RaceImportanceSpectrum, but they also were below average on the desired level of attractiveness in their partner.



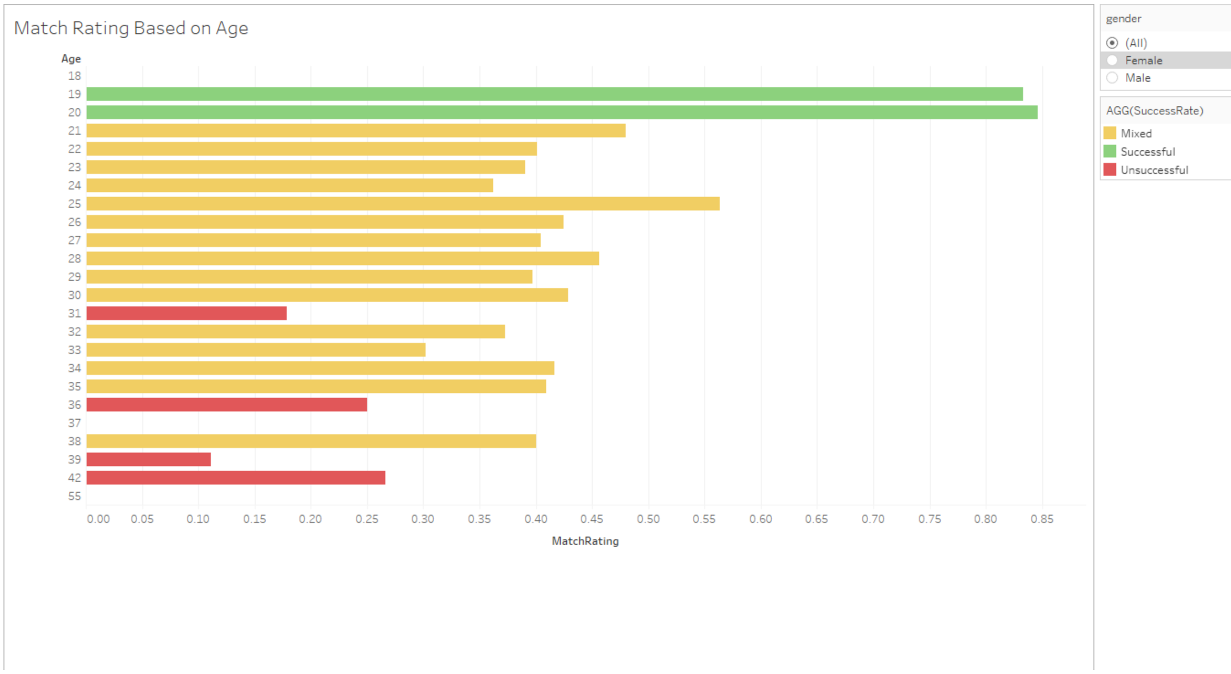
Motion Chart: Position and Partner's Decision – All

In order to analyze how the position in the session impacted the overall decision process, we created a motion chart incrementing by each individual date number in the larger order. We can see that for all participants, there was a slight decline in overall partner decisions (circle) after position 12, as well as an overall decrease in the participants decision (color) after position 13.



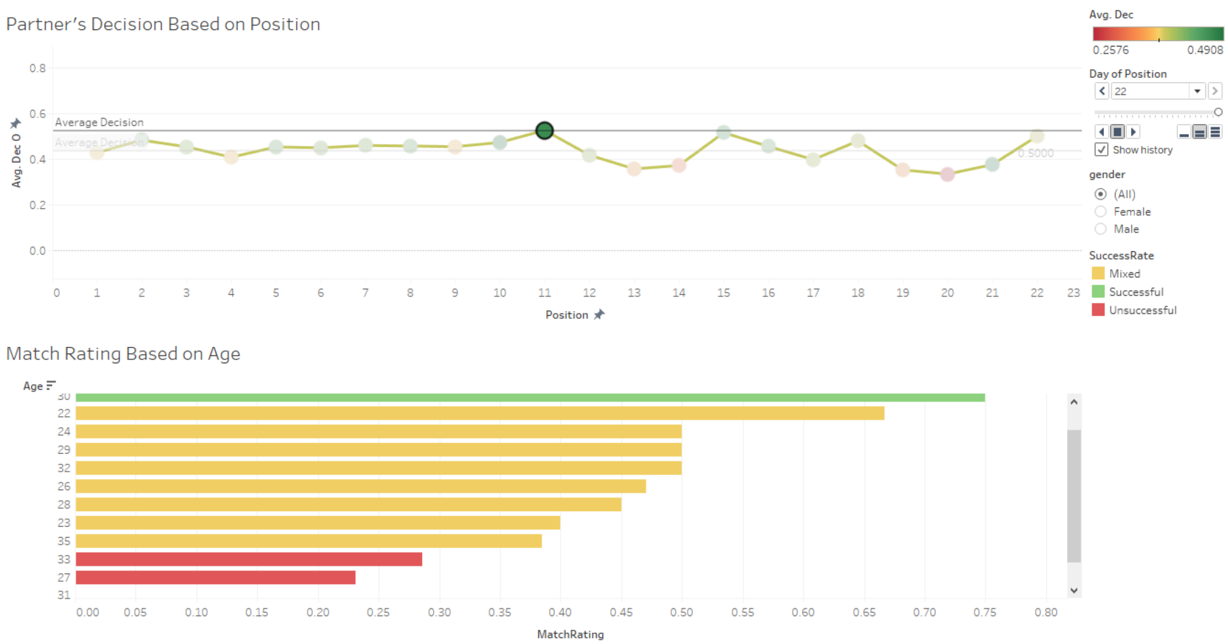
Bar Chart: Match Rating by Age – All

We also wanted to explore how age impacted the overall success of the dates. We created a SuccessRate calculated field using our calculated MatchRating to display the differences in success based on the following criteria: If [MatchRating]>.7 then "Successful" ELSEIF [MatchRating]>.3 then "Mixed" else "Unsuccessful". This was added to the color section of the visualization to highlight which ages showed the greatest percentage of matches when compared to the number of times they also said yes on their partner. The visualization below shows that the younger participants (ages 19 and 20) were disproportionally more successful than those on the highest end of the age spectrum.



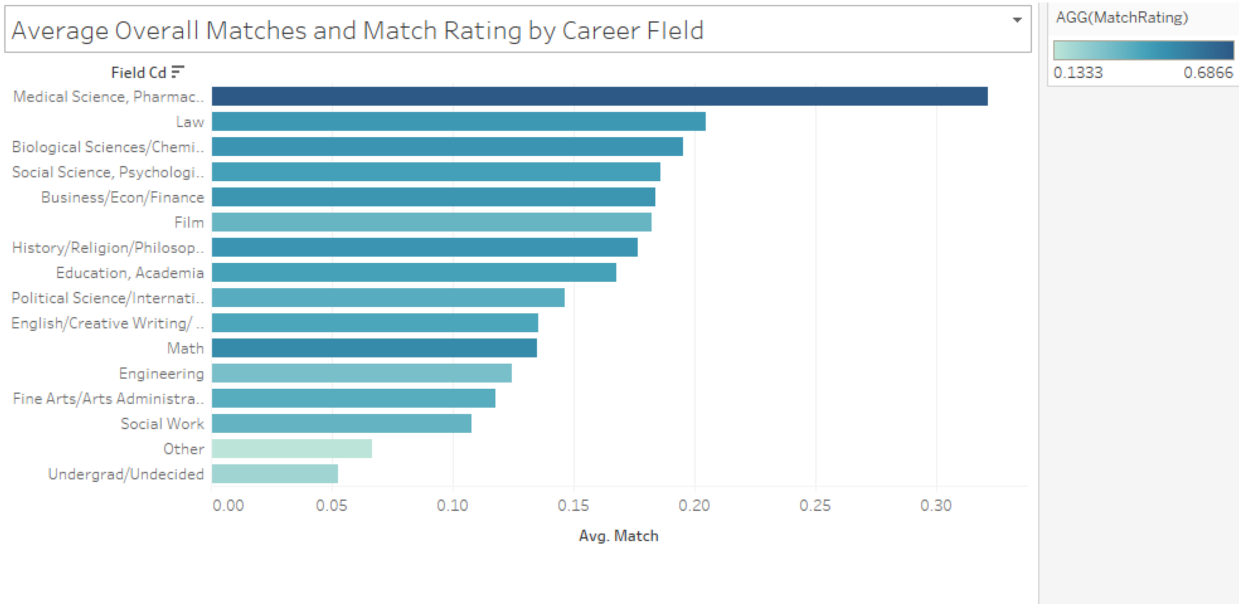
Dashboard: Position, Age, and Match Rating

While the visualizations alone told a story about the dataset, we wanted to combine them to make the analysis more robust. By creating a filtering action on the dashboard, we can drill into any position in the session to see the success of the participants at that specific point in time. At the highest point for all participants (round 11), many of the middle age range participants received mixed levels of matching, but 30-year-old participants were highly successful at this point.



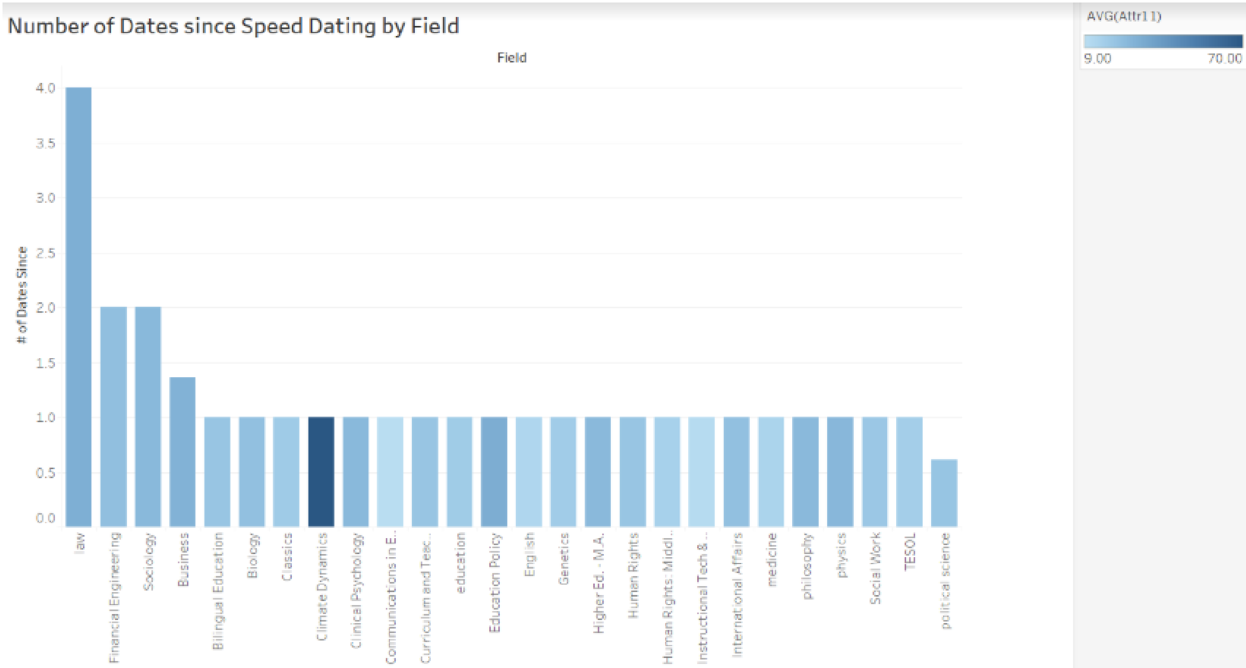
Bar Chart: Matches and Match Rating by Career Field

Up to this point, we focused on race, gender, age, and interests to visualize the important aspects of the dataset. Another major factor in the overall success of the dating event came from the different career fields. Below shows the average match rating for different participants in the individual fields. We can see that Medical Science, Pharmacy had the highest number of matches on average as well as the highest MatchRating. This means that nearly 70% of the time that an individual in this field said yes to a partner, the partner had also said yes on them. Conversely, undergraduates frequently said yes to a partner, but did not receive the yes to themselves in return (did not match). If we consider the average age and age distribution of the dataset, it is possible that the undergraduates were seen as too immature or not interesting enough for not having a desired career field.



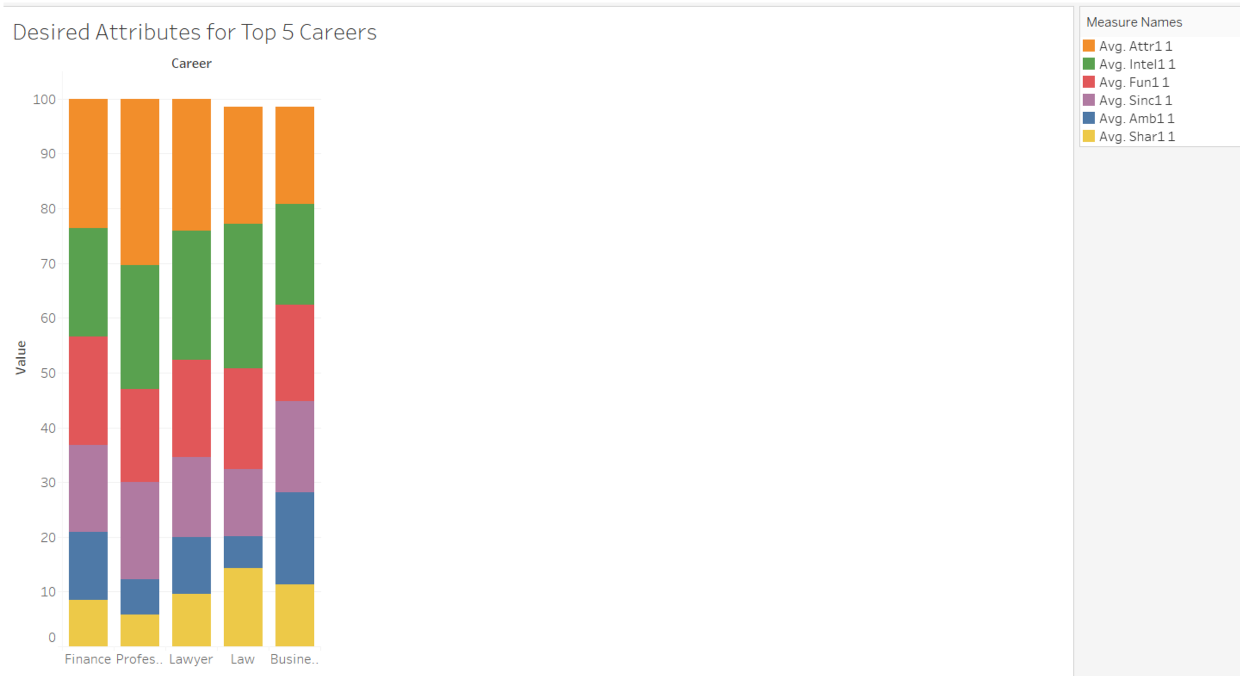
Column Chart: Number of Dates

In addition to analyzing the careers with the highest MatchRating and average match level, we wanted to determine which careers had the highest success rating in the long-term. We decided to create a visual that merely identifies the most successful careers by the number of dates gone on since the speed dating event. As seen below, the top career, Law, has almost double the number of dates than the other fields. The bars in the graph are highlighted by the rated level of importance for attractiveness, as rated by the participant. Participants in climate dynamics had a disproportionately high emphasis on attractiveness over other career fields but went on average only one date after the event. We used this visualization and the total number of overall participants to generate a top five most successful careers in the speed dating event, which will be analyzed in the following sections.



Stacked Column: Desired Attributes by Field

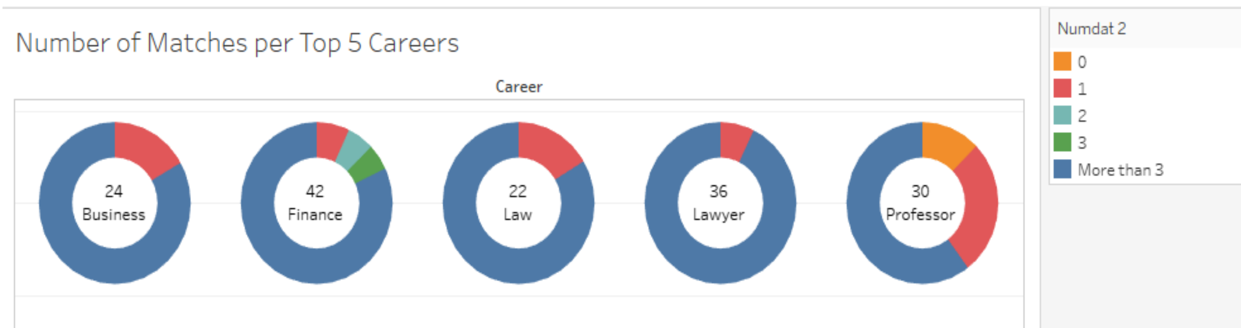
As shown previously, attractiveness plays a large part in the decision-making process for men and women in regard to finding a partner. We wanted to explore how the other attributes ranked for each of these top five fields. This stacked column chart clearly depicts the most desired traits in each career field by category of trait. As shown below, attractiveness and intelligence are the overwhelming favorites in the first four career fields, but business shows a much more balanced ranking for each of the attributes. What was most surprising is the low level of importance placed on ambition in each of these career fields. Each of the fields below require a significant amount of education and ambition to be successful, yet these individuals do not care to seek a partner with a high level of ambition. Only the business field had a more than 15% of the total 100 points allocated to ambition on average.



Donut Chart: Number of Dates and Matches

We wanted to analyze the top five fields from above further, so we created a donut chart to show multiple different relationships between career and speed dating success. The center of the donut chart displays the total number of matches for each field. More importantly, the angle on the donut chart is represented by the amount of dates that each participant went on with their match after the speed dating experience. As the legend indicates, of the matches made by the top 5 successful careers, the majority of the matches have been on "more than 3" dates with their match since speed dating. Speed dating is often stigmatized as non-effective and almost a comedic model used to obtain a partner, but this visualization portrays a different sentiment. The career fields with the most matches show a promising result for participants.

When examining the relationships between the most successful career fields in speed dating and other variables that can influence them, we see that there are similarities among the group. The top careers seek out attractive companions more so than almost any other category. Finally, these groups that have been successful in obtaining matches in speed dating, have proven to be successful post-speed dating. The majority of the matches in each group have gone on more than 3 dates since speed dating, which shows that the initial match and success of the entire group cannot be considered an anomaly.



Conclusion

While visualizations are an important aspect of the data analysis process, one must first explore the full context of the dataset and understand the goal of the visualization before attempting to tell a story through the data. To do this, our group first performed a thorough investigation on the background of the dataset, the types of variables, and the different relationships—both understood and theorized—within the data. Before creating our visualizations, we mapped the story we wanted to tell in each visualization based on the audience and our initial exploration. We then chose the best visualization to tell that specific portion of the story based on the strengths of the chart type, as this is the best way to communicate our message to our audience. Finally, through the use of visualizations, our group was able to thoroughly analyze a large dataset and gain a variety of valuable insights from this process.

While we have moved from the age of speed dating to the age of smartphones, various insights can still be gained from analyzing the perspectives, tendencies, and preferences of the thousands of participants in the dataset. Overall, most of the speed dating participants were Caucasian between 23 and 29 years of age, with outlier participants ranging up to 55 years of age. Despite this varied age group, the preferences for attributes remained largely consistent. Men consistently value attractiveness over any other trait, while women seek partners who are more intelligent, which remains consistent even across the different career fields. This is consistent with the current dating field, which uses applications such as Tinder and Bumble to "match" partners only off of a few pictures and a snippet of a biography. In addition, we explored the racial tendencies of the dataset. Despite the common conception that the Southern United States hold the most racial bias, we found a varied level of bias towards dating individuals of the opposite race all across the United States. By creating calculated fields, we were able to highlight areas were this bias was at its peak. We also determined that the overall position of the date can play an influence in how you are perceived by your partner. Early dates show a higher level of participant and partner "yes" decisions, but the later you get into the speed dating sessions, the more variation can be found in these decisions. Finally, we determined that certain career fields, such as Business and Law, had a higher overall success rate for long-term dating. This project required our team to carefully consider the various aspects of visualization design, including both the who, what, and why, as well as the affordances, aesthetics, accessibility, and acceptance of overall design. We could not just play around with the data set to find what visualizations looked interesting, rather, we had to work together to tell a complete story with our visuals.

We can use some of this information to make general recommendations to individuals who will be participating in short-term dates like a speed dating event. As most individuals valued ambition extremely low, it would be best to save conversations about long-term ambitions and goals for a later time, such as a third or fourth date. In addition, appearance-based first impressions do matter in how likeable you are, so put care into how you look before attending any meeting or first date. Your career can also play a key role in how you are perceived by the opposite gender as well, especially if you are in business or law. Finally, it is important to gain a true understanding of your own racial biases. You may believe that you do not value race in the dating setting, but if you find yourself saying no consistently to individuals of the opposite race, it may be time to reassess your priorities.

As our group focused only on a small set of variables, this data set can be used to perform additional analysis. While we focused our visualization on the importance of race, we could also have explored the importance of religion using the same methodology. Furthermore, the surveys also requested information on an individual's perceptions of their own gender and themselves, which could also be used to investigate how these impact the decision process.

Works Cited

Brown, Nell Porter. "The Road to Romance". Harvard Magazine (March–April 2003): 28–28A. https://web.archive.org/web/20061030035039/http://www.harvardmagazine.com/on-line/030382.html

Montoya, Anna. "Speed Dating Experiment." Kaggle, 9 May 2016, www.kaggle.com/annavictoria/speed-dating-experiment.