Data Visualization: Milestone 2

Hongyi Shi, Ajkuna Seipi, Nataša Krčo

7th May 2023

1 Introduction

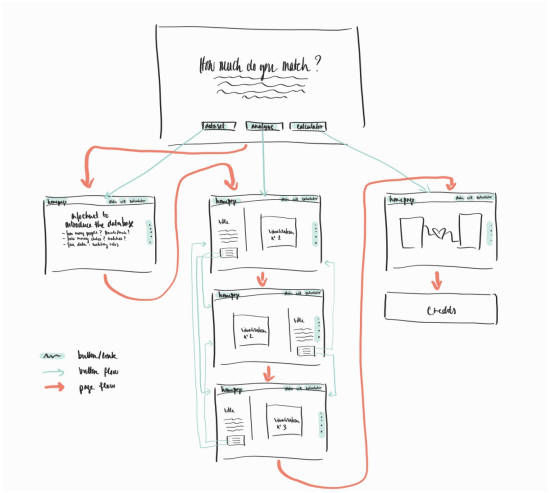
The final project will be composed of four interactive visualizations, where three of them are basic visualizations that require the D3 JavaScript library. The fourth visualization will be a matching predictor of a date between two people. It will require a Machine Learning model that predicts the percentage of matching, and a JavaScript implementation of the interactive user interface.

To be able to use the D3 library and to understand the basic implementations, we use the first lectures about *Data-Driven Documents* and *JavaScript* for all the visualizations.

2 Visualizations

2.1 Overall layout

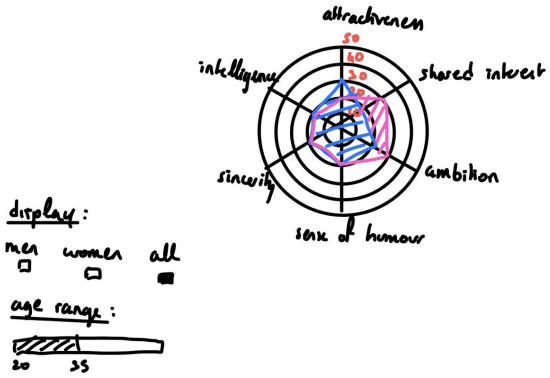
The website is designed as a one page website divided into 3 sections in addition to the main page and the footer credits: introductory information, advanced visualizations, match predictor. The orange arrows shows the natural flow by scrolling through the website and the green ones are the flow introduced by the links and the buttons.
In the first section, we will give basic information about the dataset such has how many participants, how many registered dates, women and men ratio in participants, and so on. Then in the second part, we will display three visualization mentioned in section 2.2 to 2.4. Finally in the last part, there will be the match predictor app.



2.2 Spider Chart: Preferences

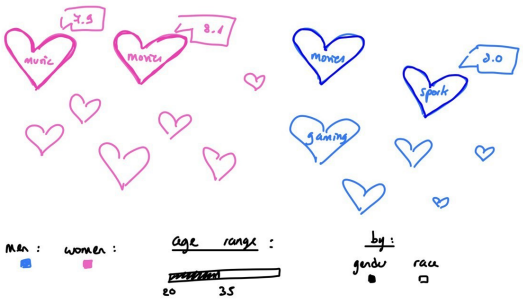
Our first visualization is a **Spider Chart** that showcases the preferences of participants in the event. At the beginning of each event, each participant was asked to rate the importance of a potential partner's aspect, ranging from 0 to 100, with the total scores adding up to 100. To create this chart, we utilized the D3 JavaScript library (Mike Bostock 2021), customizing a spider chart from the D3 Graph Gallery (*The D3.js Graph Gallery* 2018) to fit our project's requirements. The chart presently displays the preferences of both men and women separately, with the added functionality of selecting the age range of the participants via a range slider, implemented through the jQuery UI JS library (*jQuery UI library* 2023). To improve

the interactivity of this visualization, we intend to add an option that allows users to choose to display preferences of men, women, or all participants, as well as the option to select the general field of study (Science or Humanities) that the participants belong to.
We use and take inspiration from the lecture 5 on *Interactions and Views*, the lecture 6 on *Perception and Colors*, and lecture 7 on *Design for Data viz.*



2.3 Bubble chart: Dates and peoples' interests

Our second visualization shows the most relevant interests of the participants.They gave a rating between 0 to 10 to the interests, where 0 means they have no interest in it, and 10 means that they are fully interested in it. To show the average rating of the interests according to the gender and the age, we decided to use a **Bubble Chart** inspired by the D3 graph gallery (*The D3.js Graph Gallery* 2018).
Our initial visualization is a plot of circles of different sizes for each interest. To compare the genders, we color the interests of men and women in different colors. To add interactivity to the visualization, we display the average rating per interests in a box. We will also add a slider to see the ratings according to the age. To enhance the interactivity, we could add buttons showing the average rating per gender, per race or/and per field.
We use and take inspiration from the lecture 5 on *Interactions and Views*, the lecture 6 on *Perception and Colors*, and lecture 7 on *Design for Data viz.*

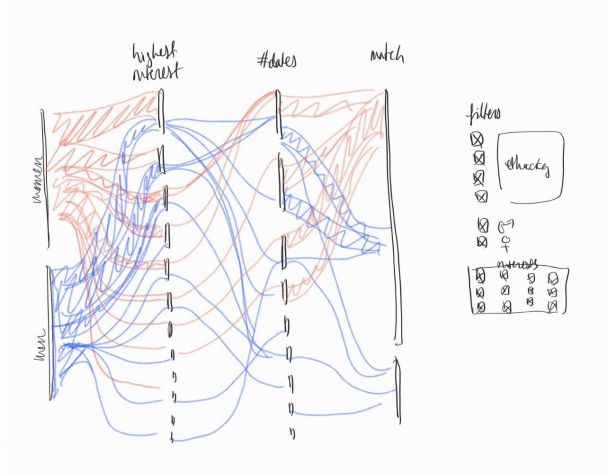


2.4 Sankey diagram: profiles and matches

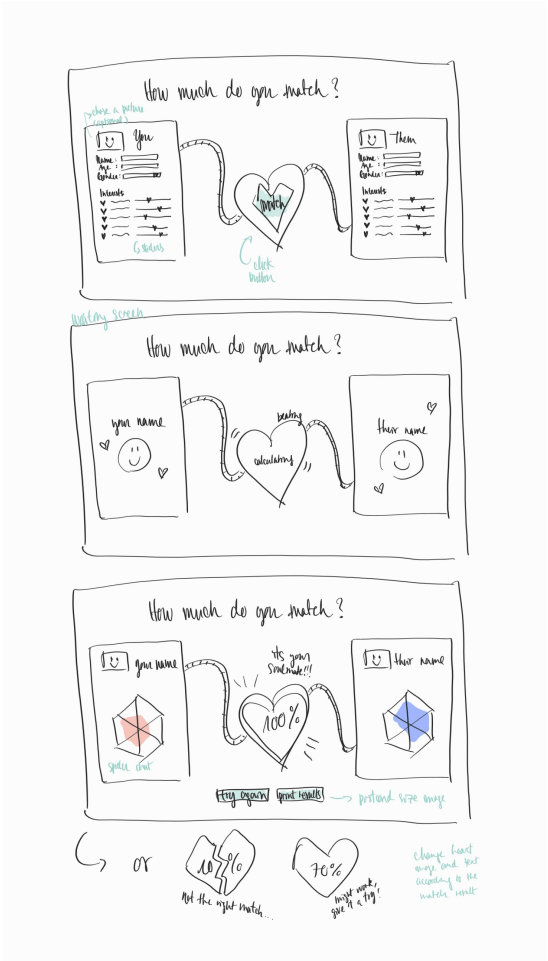
Finally, the main topic of our dataset is how likely a profile could find a match through speed dating. To show it, we plan on using a Sankey Diagram from the D3 graph gallery (*The D3.js Graph Gallery* 2018). The nodes would be divided into the following categories: gender, highest interest, number of dates attended and if the participant got a match or not. Filters would be available in order to allow more

interactivity, such as showing the visualization for specific ethnicity, interest or gender. We could also add data related to rejection after a date i.e. if only a single party would have liked to match with their date partner, or change the match nodes to the total number of match the participant got.

We take inspiration from lecture 11 *Tabular data*.



2.5 MATCH PREDICTOR



Our original idea is inspired by the various online love calculators we could find in the year 2000. Here, we are trying to make something which could give more accurate predictions. We plan on training a machine learning model using our dataset in order to predict if two persons are likely to match together or not, users could then use this model on our website to predict the matching rate given two personalized profiles. This would present as the following.

First, the user needs to fill a profile form for each of the two party in the matching which requires information such as age and their interest. Then the calculator starts the computation once the button at the center is pushed and, after some time, it outputs the results displayed in the middle of the two parties profile which is showed as a spider chart, similar to the one mentioned in section 2.2.

To do so, we will host our model on a web application hosting platform such as Fly.io. In case it is too complex, we could also limit the number of possible inputs and compute beforehand all the possible output, such that we already have a database of with all predictions. We could then display the matching score easily without having to use an external hosting platform.

REFERENCES

Mike Bostock (2021). *Data-Driven Documents*. URL: https://d3js.org/ (visited on 5th May 2023).

The D3.js Graph Gallery (2018). URL: https://d3-graph-gallery.com/ (visited on 5th May 2023).

jQuery UI library (2023). URL: https://jqueryui.com/ (visited on 6th May 2023).