Wanna Date-a?

CSE 457 Final Project Proposal

Basic Info. The project title, your names, e-mail addresses, IDs, a link to the project repository.

- Title: Wanna Date-a?
- Names & IDs: Mariah Yelenick (458347), Emily Wilson (450399), Lydia Ho (457170)
- Emails: myelenick@wustl.edu, emwilson@wustl.edu, lydiaho@wustl.edu
- Repository: https://github.com/mariahyelenick/CSE457

Background and Motivation. Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

- As three beautiful, mostly single ladies, we wanted something relatable to our lives and something all three of us were interested in
- We wanted to look at Tinder data to see what kind of people are swiping which way but unfortunately, that data is not publicly available (for good reason)
- The next best thing was publicly available, anonymized OkCupid demographic data

Project Objectives. Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

- The goal of the visualization is to break down the demographics of a typical OkCupid users
- We would like to be able to profile an average user on the platform, and help the general public decide whether OkCupid suits their needs in finding a partner
- Users would get to "people-read" (a la people-watching) to get fun insights into what people put on dating profiles
- A user could put in their demographic information, see what people typically say in their open ended responses, and modify their responses to fit in with the crowd or stand out, whatever their preference
- We hope to produce an informative and entertaining final product

Data. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

- https://github.com/rudeboybert/JSE_OkCupid
- There's a zip file with a csv full of the data

- There are around 60k data points out of around 30M total OkCupid users

Data Processing. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

- The data is mostly clean already for us, in a csv file with ~60k entries
- Most of the data processing we want to do will be text analysis for the essay responses. This will allow us to display summaries or other analyses of the open ended questions after a user has filtered the selections they're interested in
- One thing we'll need to do is split up the columns that have multiple options, like ethnicity
  - I'll create boolean columns for each ethnicity and also a multiple ethnicities column so we can allow users to filter based on one or more ethnic classifications or just view people with one or more ethnicities
- I will also bucket the age data so that we can view groups of people at a time, rather than just 31 year olds (for example)
- Get longitude and latitude data for each city

Visualization Design. How will you display your data? Provide some general ideas that you have for the visualization design. Create **three alternative designs for your visualization**. Create **one final design that incorporates the best of your three designs**. Describe your designs and justify your choices of visual encodings. You use the [Five Design Sheet Methodology (Links to an external site.)Links to an external site.](#).

Must-Have Features. List the features without which you would consider your project to be a failure.

- Filtering by certain demographic data to see breakdown of users
- Profiling an average user

Optional Features. List the features which you consider to be nice to have, but not critical.

- Randomly select a user's profile for viewing in an interactive way
- Choose two categories and compare users who identify as such
- Map how many users are in each city in the US, and potentially display locations on a map

Project Schedule. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

- March 18: Project Proposal

- March 25: Data Wrangling
- April 1: Prototype 1
- April 8: Visualization Design
- April 15: Prototype 2
- April 22: Implementation, Presentation
- April 29: Final Version, Evaluation, Video