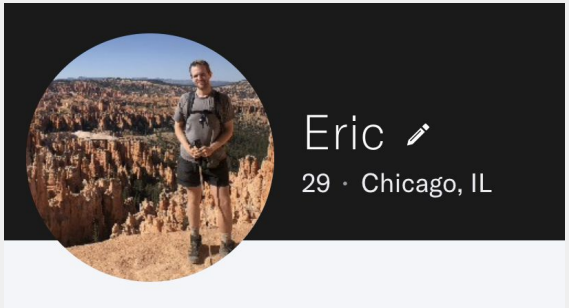
Trends in Online Dating

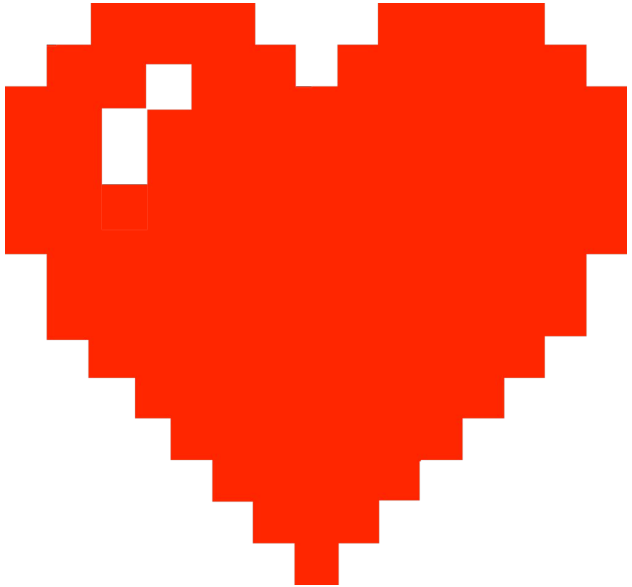
By



The Question

How do dating profiles and dating goals change with age?

- Make more targeted dating products and services



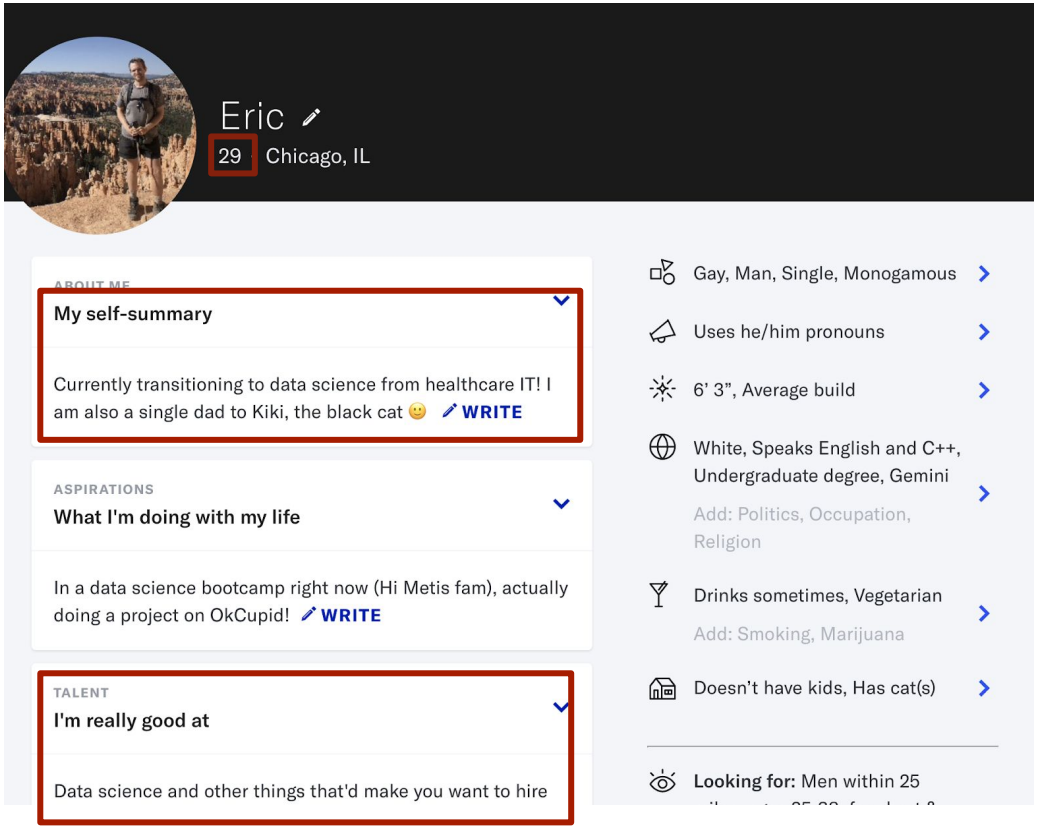
Methodology

Data

- 60,000+ OkCupid profiles
  - Data from the Journal of Statistical Education
- Profiles from June 2012
- Users located in San Francisco Bay Area

Natural Language Processing

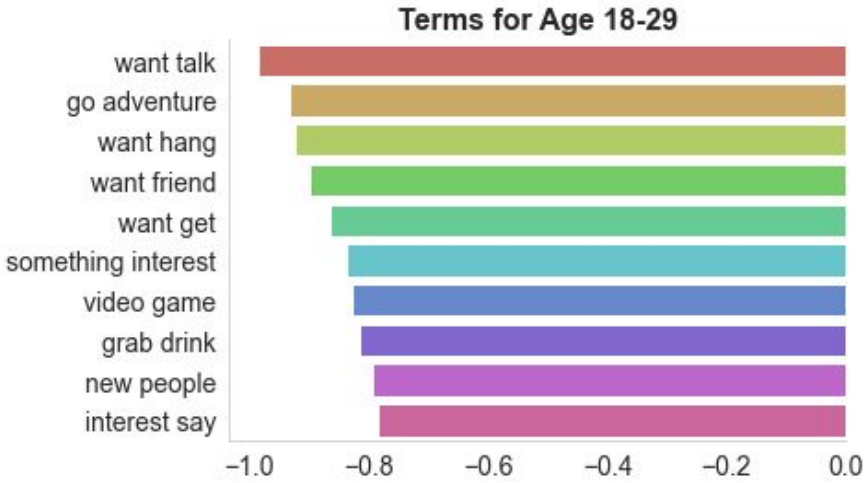
- Topic Modeling (NMF)
- Scaled F-Score (via ScatterText)
  - How strongly is a word associated with a category
  - Score between -1 and 1

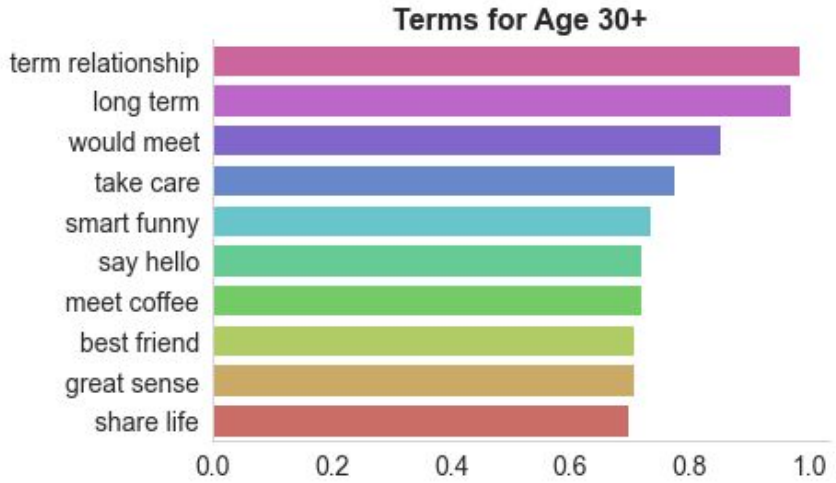


- Age
- Descriptions from selected sections

Findings: Age

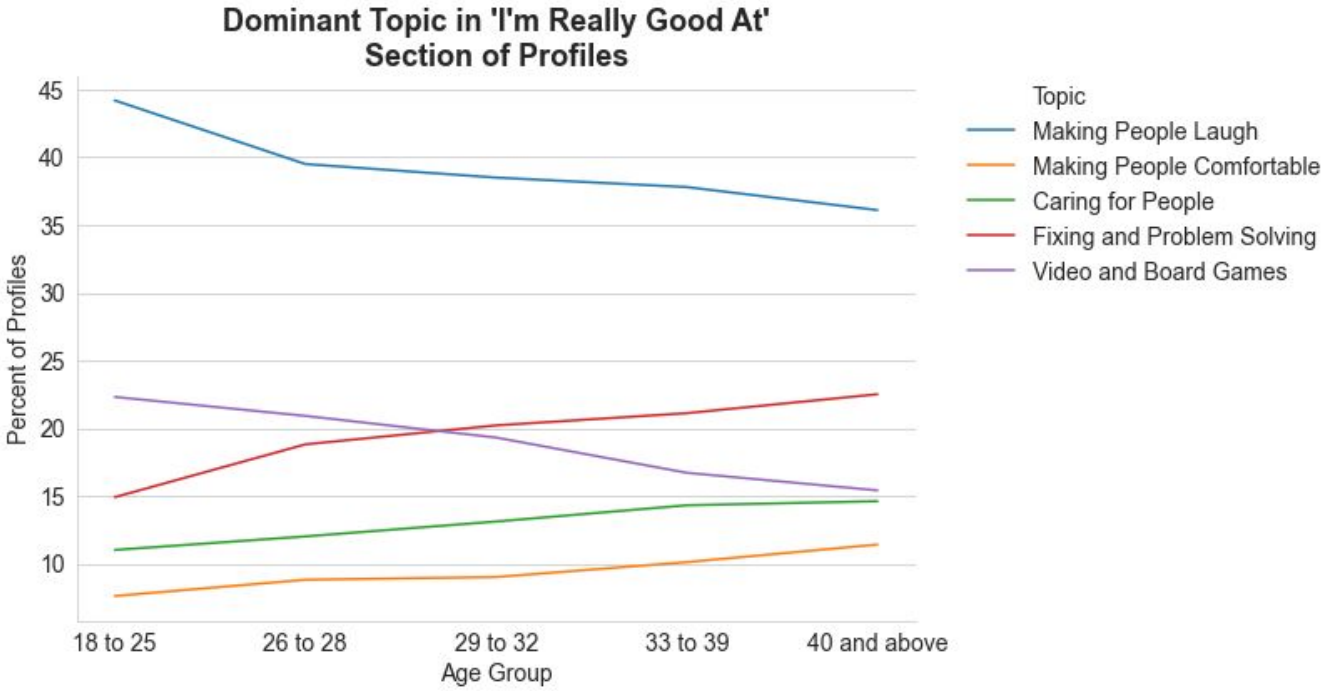
Terms strongly associated with one age group in **‘Message Me If’** section:





Older users are looking for more serious relationships.

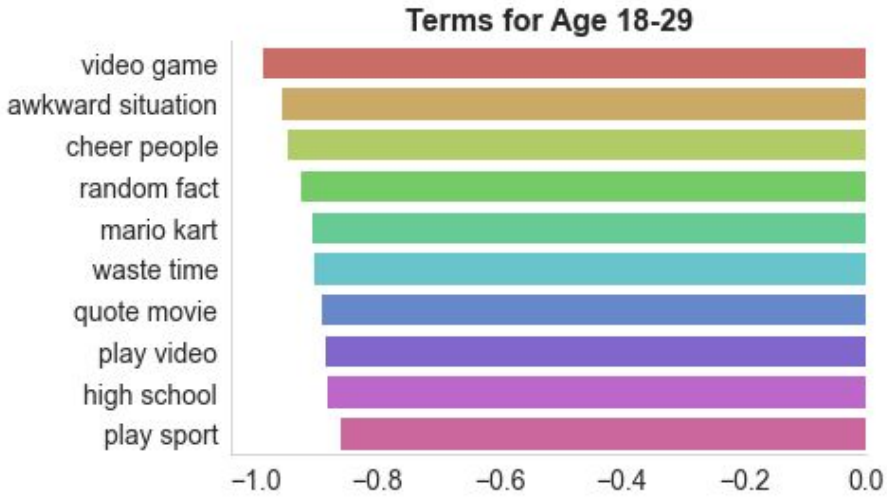
Findings: Age

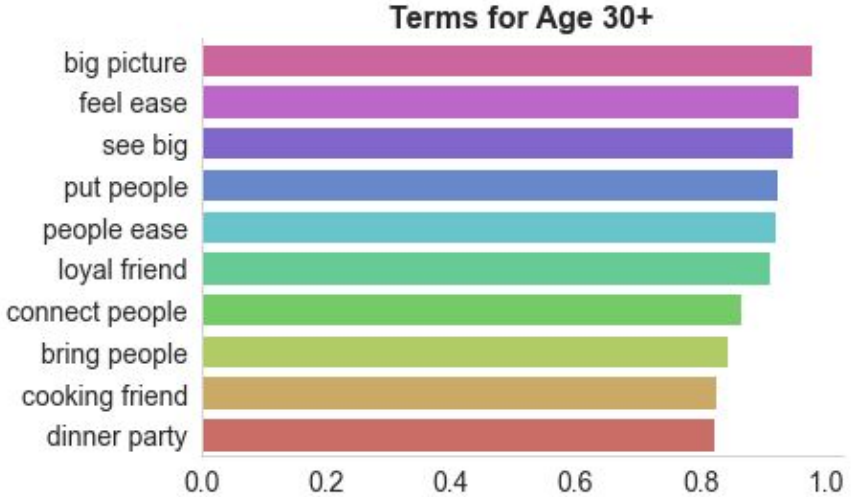


Note transition from fun topics to more practical and nurture-driven topics with age.

Findings: Age

Terms strongly associated with one age group in **‘I’m Really Good At’** section:

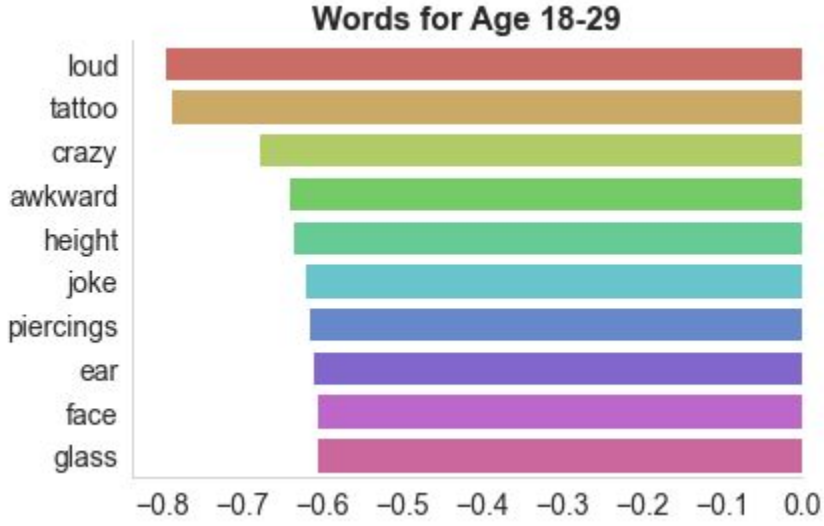


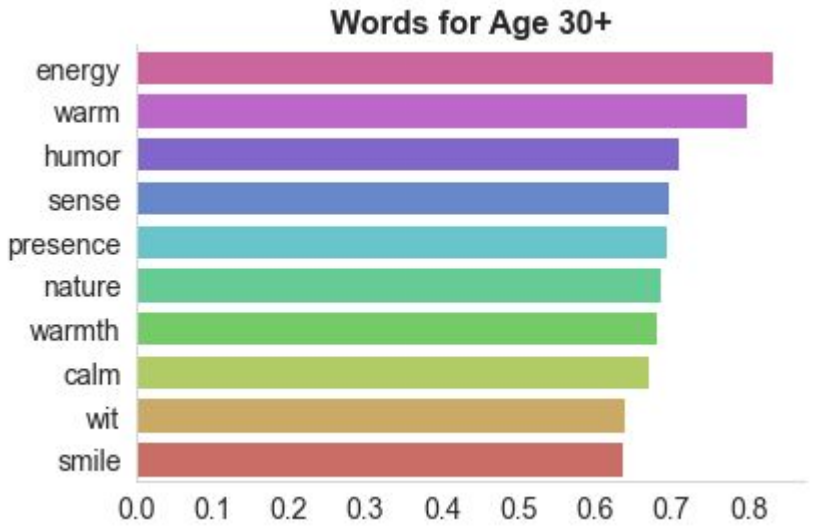


Note similar transition from fun and hobbies to social and practical skills.

Findings: Age

Terms strongly associated with one age group in **‘What People First Notice About Me’** section:





Note transition from physical to more universally positive personality attributes with age.

Summary and Future Work

Summary:

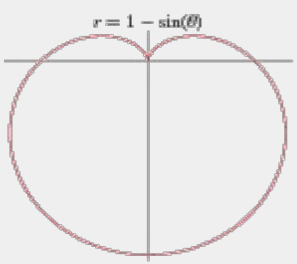
- With age, dating users look for more serious relationships and frame their profiles accordingly
- Age also brings more focus on nurturing, practical skills, and personality

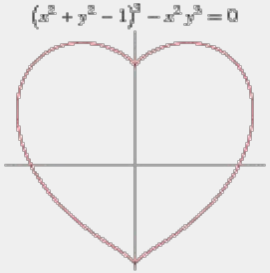
Future work:

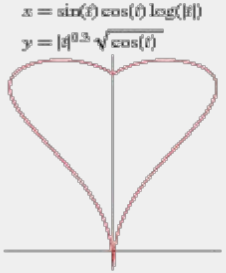
- Find and/or create more current data, if possible
- Recommendation system

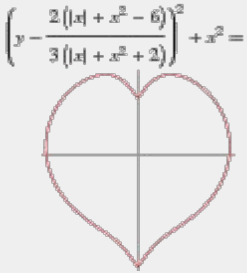
Questions?...

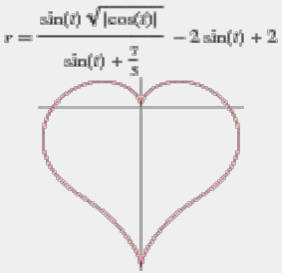
...or leads for eligible men in the Chicagoland area?

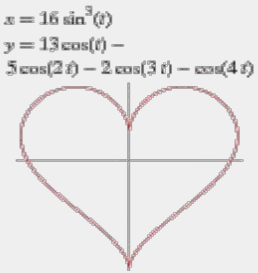






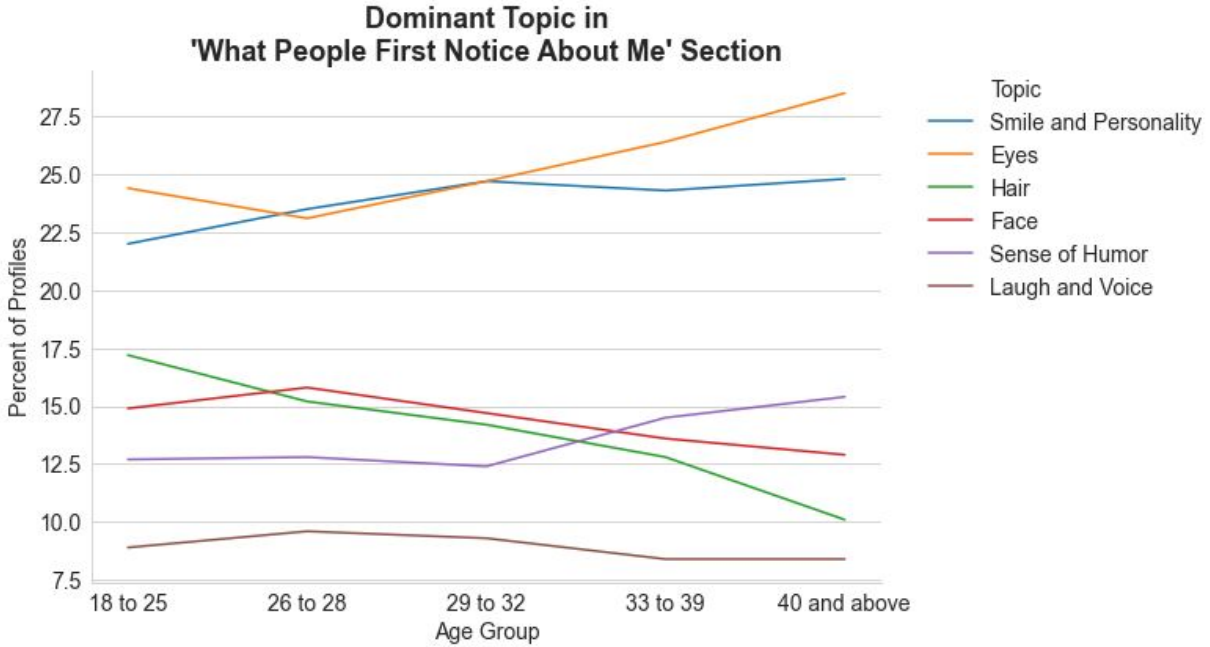






Appendix: First Notice and Age

Topic modeling results- note eyes become more popular with age, while other physical attributes become less common.



Appendix: Cat vs Dog People!

Cat people show more interest in science and culture, dog people seem more athletic, traditional, and social.

Appendix: Scaled F-Score, explained

Based on ScatterText.

For a given word and category (e.g. age), we calculate its precision, or how often the word appears in documents for the category vs how often the word appears in documents for all categories.

For a given word and category, we calculate its frequency, or how often the given word appears in documents for the category vs all words for documents in the category.

The harmonic mean of the precision and frequency result in the F-Score. To make sure precision and frequency contribute equally (without adjusting beta), we also take the CDF of the precision and frequency with respect to other values in the given category. Finally, we scale the values from -1 to 1.

In my implementation of the Scaled F-Score, I used a beta value of 0.5, which gives more weight to precision than frequency.

Appendix: Scaled F-Score, explained

Scaled F-Score

Associatied terms have a *relatively* high category-specific precision and category-specific term frequency (i.e., % of terms in category are term)

Take the harmonic mean of precision and frequency (both have to be high)

We will make two adjustments to this method in order to come up with the final formulation of Scaled F-Score

Given a word $w_i \in W$ and a category $c_j \in C$, define the precision of the word w_i wrt to a category as:

$$\text{prec}(i,j) = \frac{\#(w_i, c_j)}{\sum_{c \in C} \#(w_i, c)}.$$

The function $\#(w_i, c_j)$ represents either the number of times w_i occurs in a document labeled with the category c_j or the number of documents labeled c_j which contain w_i.

Similarly, define the frequency a word occurs in the category as:

$$\text{freq}(i,j) = \frac{\#(w_i, c_j)}{\sum_{w \in W} \#(w, c_j)}.$$

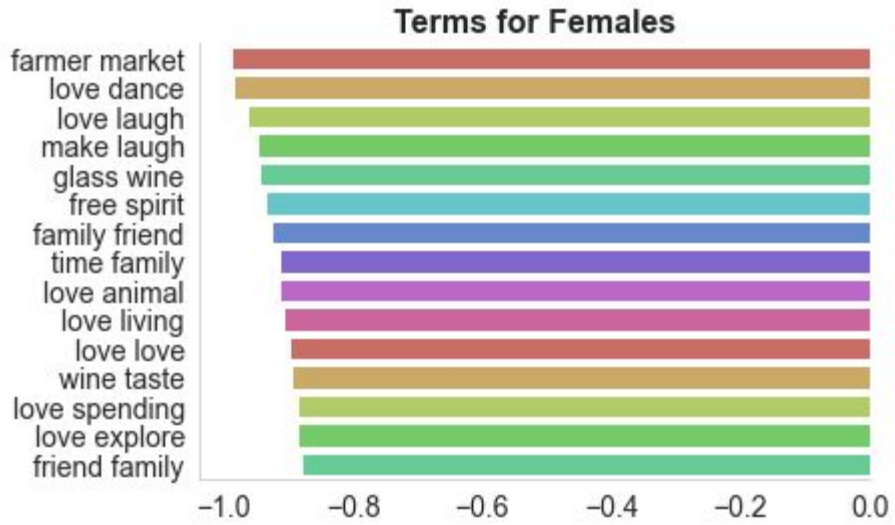
The harmonic mean of these two values of these two values is defined as:

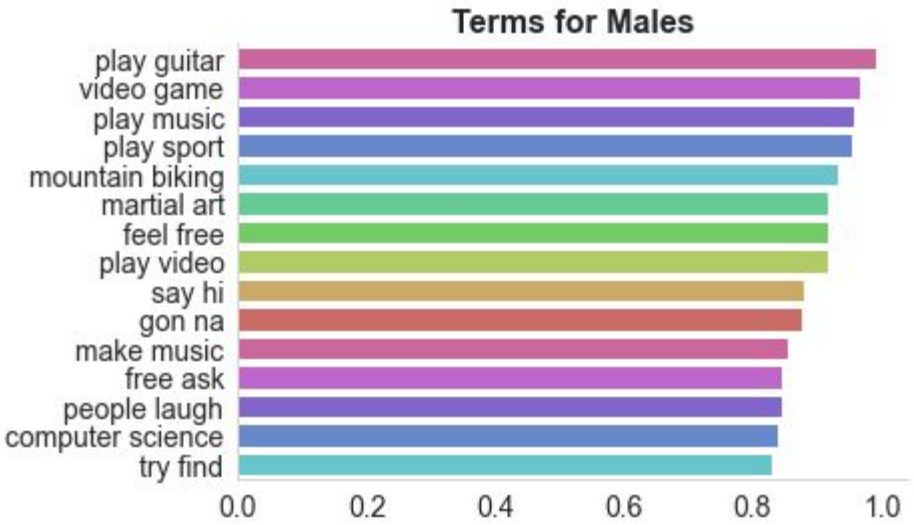
$$\mathcal{H}_\beta(i,j) = (1+\beta^2)\frac{\text{prec}(i,j) \cdot \text{freq}(i,j)}{\beta^2 \cdot \text{prec}(i,j) + \text{freq}(i,j)}.$$

$\beta \in \mathcal{R}^+$ is a scaling factor where frequency is favored if $\beta < 1$, precision if $\beta > 1$, and both are equally weighted if $\beta = 1$. F-Score is equivalent to the harmonic mean where $\beta = 1$.

Appendix: Sex

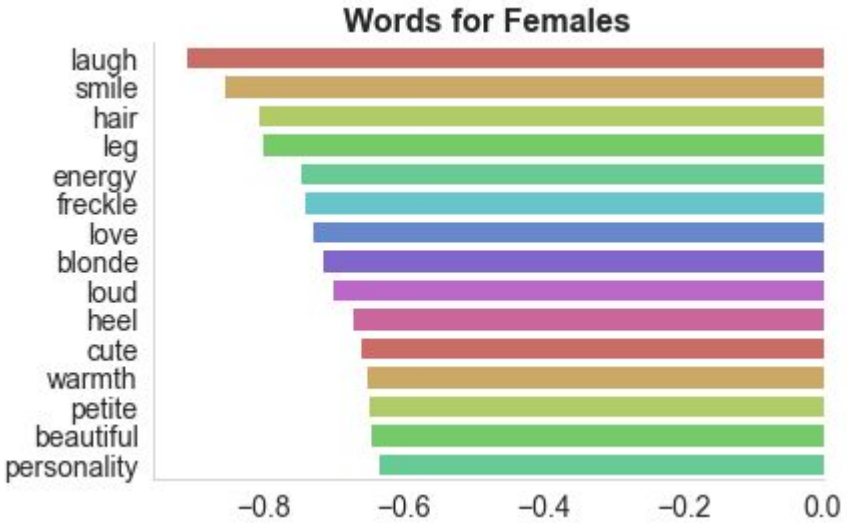
Terms strongly associated with one sex in '**Self Summary**' section:

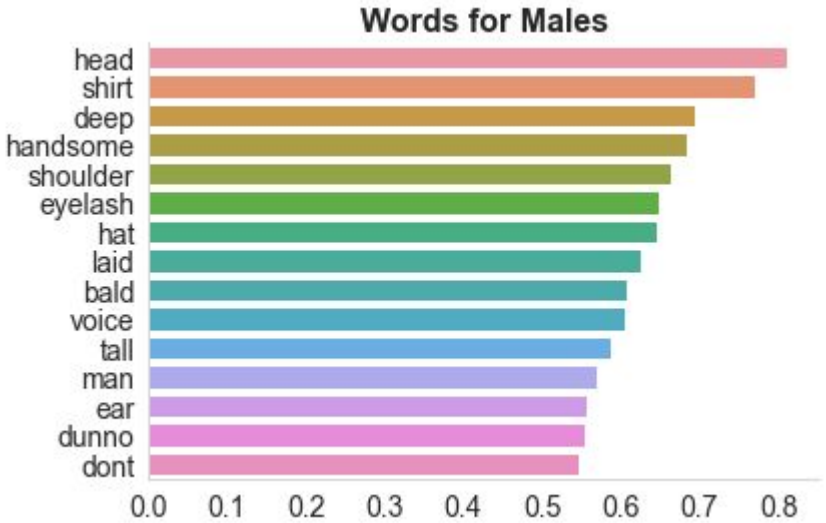




Appendix: Sex

Terms strongly associated with one sex in '**What People First Notice About Me**' section:

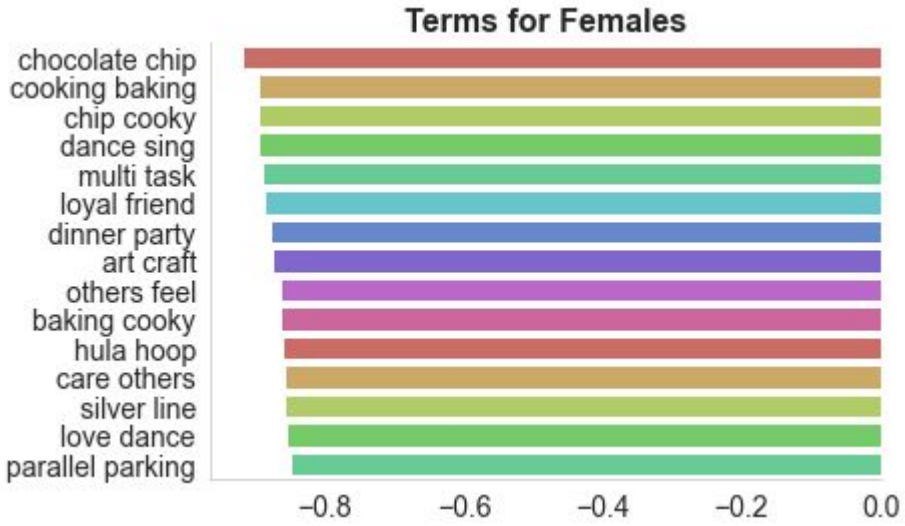


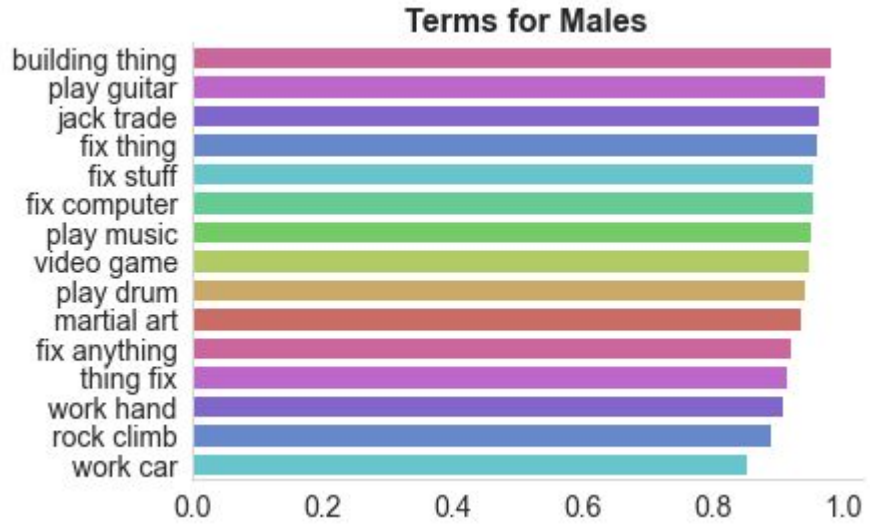


Note females tend to write about physical attributes and personality; males only write about physical features (including eyelashes?)

Appendix: Sex

Terms strongly associated with one sex in **‘I’m Really Good At’** section:

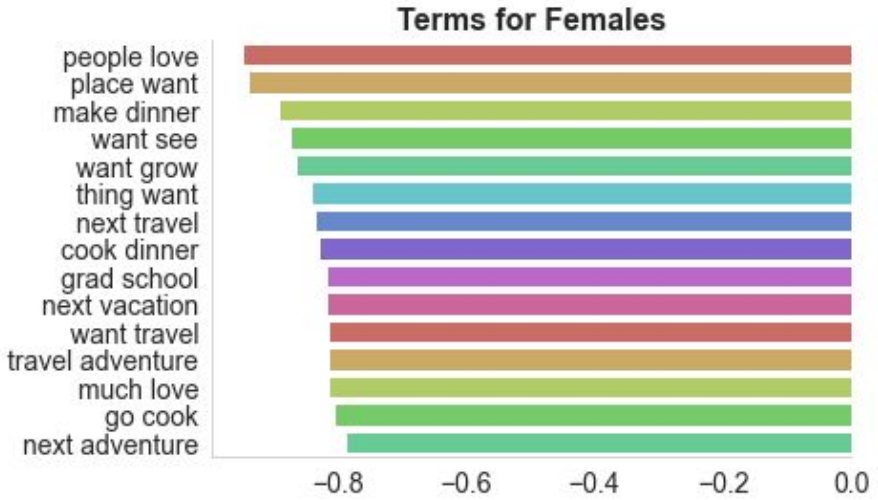


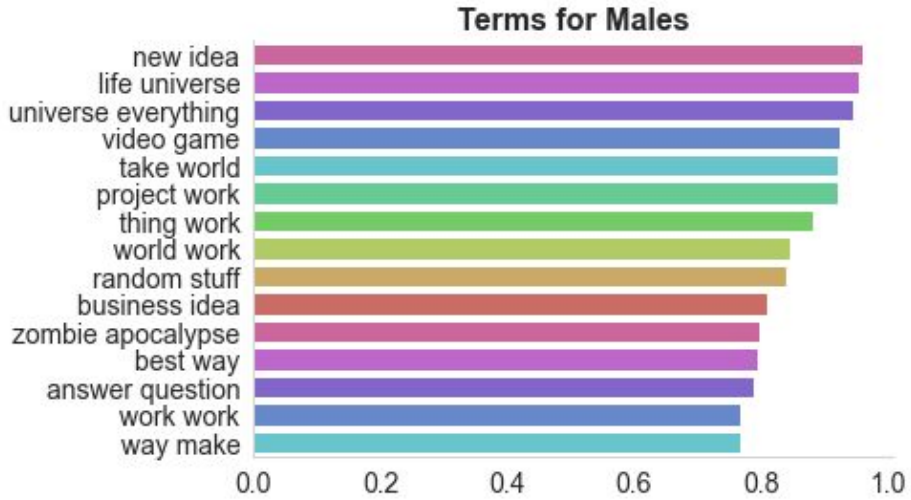


Parallel parking and hula hoop?

Appendix: Sex

Terms strongly associated with one sex in **‘I Spend a Lot of Time Thinking About’** section:





Males seem to be more grandiose- see ‘jack (of all) trades’ and ‘fix anything’ in prior slide and ‘life universe’ here.