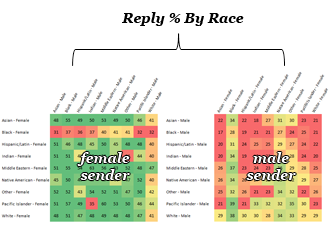
**OKCupid | Redesign Project**

***Introduction:***

OKCupid is an online dating website founded in 2007, which today aids over five million users in connecting with potential dating partners. The site allows date seekers to send messages to those they find interesting based on posted profile information (e.g., age, sex, race, interests, or the like). Recipients of these messages may view the sender’s description and decide for themselves whether or not they would like to respond.

This redesign focuses on an article published with the help of data straight from OKCupid’s servers, aiming to illustrate how race affects the percentage of responses a user receives[[1]](#footnote-1). The initial graphic is provided in a reduced scale below. The reply percentage for messages exchanged from a sending race to a receiving race is displayed for both male and female senders. Relative colors on the heat map illustrate the reply rate percentage. The goal of the article is simply to explore the depicted trends between races, genders, and reply rate chances and pick out some interesting trends.



***Original graphic***

Several aspects of the original graphic can be improved in the redesign. First, while the color encoding emphasizes the overall reply rate difference between genders, it does not allow the reader to readily establish the magnitude of this difference without looking at the particular values. Transitioning to a position along the scale encoding will remedy this deficiency. Next, it would be advantageous to group the data based on sender race instead of by gender. This would help the viewer determine the magnitude of the reply rate difference males and females of the same race experience. Moreover, if gender differences exist, it would make it clear that males and females of the same race have different races interested in them. In a larger sense, additionally it would be beneficial to clearly show which race of each gender receives the most replies on average. One could then quickly determine the most/least successful senders of each race and that alone would be a significant enough takeaway. These and other features will be addressed in the redesigned graphics presented below.

***Redesigned Graphic:***



The redesigned graphic utilizes a position along the scale encoding to show, for both men and women, how successful each race is at receiving replies to their online dating messages. The ggplot facet wrap layout allows us to compare, for each race, the difference in male and female reply rate side by side. The overall difference between the genders’ reply rates is shown numerically in the legend. On average, females receive replies 20% more often than males. This is a staggering statistic and the data shows that this difference is more or less retained for each race. Larger gaps exist between some races (e.g., Middle Easterners with around 27%) and smaller gaps for others (e.g., Blacks with around 14%), all of which are obtainable via the mean lines plotted for each gender and race.

Providing mean lines also aids in comparing the overall reply rates for each gender. The sender race panels are arranged based on the male reply rate percentage. Thus, in the top left, it is seen that white males receive the most replies on average (~31%) and, in the bottom right, black males receive the fewest on average (~23%). Females do not follow exactly the same trend, but their order can also be obtained via the red mean lines and the provided, color-coded mean value labels. Middle Eastern women receive the most replies (~54% ) and Blacks once again receive the fewest with a considerably low percentage (~36%).

An interesting metric can be created based on the data’s spread from their mean lines. While differences in the mean line values for each gender can be most likely explained by an overall tendency for men to reply more than women, the amount of spread from the mean lines may represent an amount of unfairness that a particular race faces. For example, White women receive roughly the same response rate no matter which race they enquire (except for White males). This may show that males of all races prefer white women in roughly the same amount, and no race is particularly enamored or disinterested in replying. However, when considering Indian women it is apparent that they struggle to receive messages from ‘Other’ and White males, and receive copious messages from Indian males. As it happens, Indian males seem to receive few messages Indian females, indicating that favoritism may not always be reciprocated. Therefore, the distributions around the mean lines may represent an unfairness or favoritism each race/gender experiences.

Visually, the replier axis is sorted based on the male mean reply rate to produce the shown linear like trend. This trend was the most prominent out of the two genders and simplifies the visual burden. Black outlines surround the generously sized points to make the data stand out and appear more inviting; careful attention is taken to avoid points touching axis lines. Males and Females are represented with standard ggplot blue and red colors respectfully. This reduces memory burden as blue is commonly associated with males and red/pink with females. Title, axis, and legend fonts are bolded and appropriately sized to be noticeable but not detract from the focus of the plot. A final consideration was taken to include only as many percentage markers as was necessary to reduce clutter and ensure that all plots could hold all of the data points on the same axis length as this facilitates magnitude comparisons.

***Shiny Application[[2]](#footnote-2):***

There are several other interesting features in this dataset and a single plot is not sufficient to capture every aspect. A Shiny application was created to hold all of the relevant graphics.

The first plotted feature which is missing is a plot showing how often each race/gender replies to messages. While the redesigned plot above helps clarify which race receives the most responses and contains all of the data of the original graph, the transpose of the dataset is just as interesting. A copy of this plot is provided in the Appendix below. One of the key takeaways is realizing that the deviation from the mean lines here indicates a measure of pickiness, wherein some races favor replying to particular races separate from the overall gender tendencies. It is easy to determine by eye that men are pickier repliers than women even though they reply more on average.

The next plot is two bar graphs showing how many replies each gender receives and how often each gender replies for each race. Essentially, the main point is that White men receive the most replies and respond the least out of all men and black females receive the least replies and respond the most out of all women. This fact is represented in the already presented plots but is more prominent in this graphic as shown in the Appendix.

Finally, the last graphic is an interactive graphic tailored to an individual who is interested in their reply preferences. This application allow a user to input a gender and a race and view (1) who typically replies back to them and (2) to whom do they typically reply. In this way, a user can examine their dating chances by hand and investigate other user’s assumed preferences.

***Future Improvements/Points of Consideration:***

Ideally, groups of four to five are easier to track across a page however due to the nature of the dataset the overall shape of each race’s distribution was deemed more important. To improve the graphic, dividing each panel horizontally into 4x1x4 sections (with Middle Eastern as the center section) would aid in readability. Additionally, it would be easier to identify which race received the most/least replies if the panel was grouped into a single sorted column. Space constraints and x-axis resolution hampered this implementation. Also, I believe that this configuration aids in distinguishing trends between races which is similarly interested to the overall values.

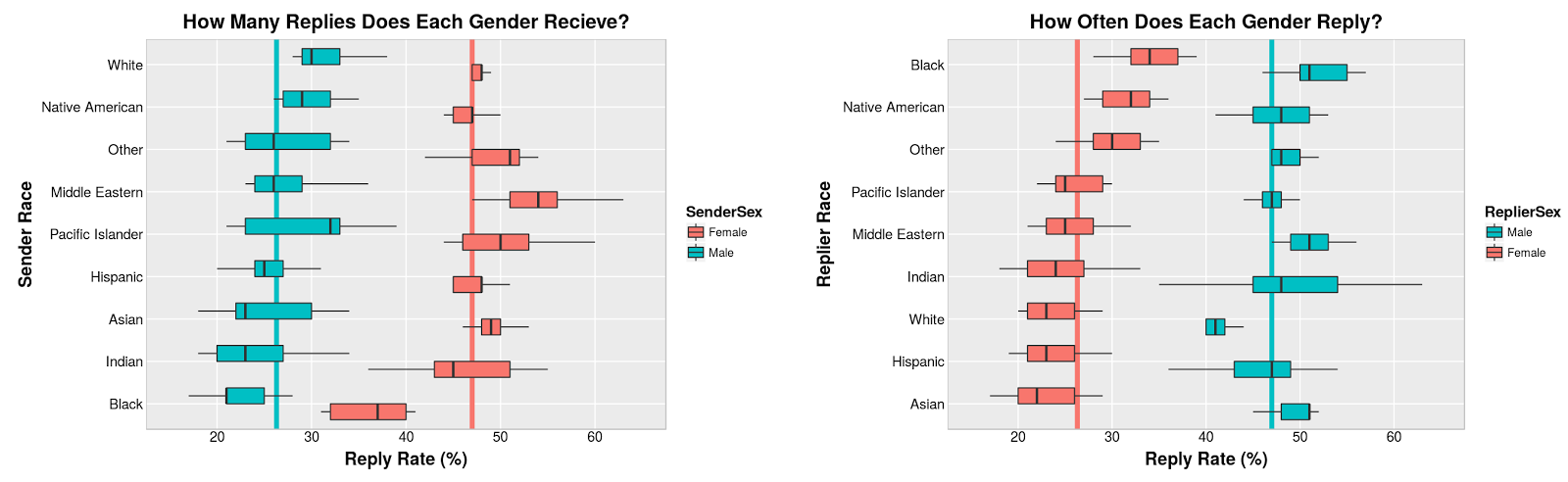
An unexamined metric exists in the most efficient race/gender. This is the race where the reply rate gap between the users they reply to and the user’s they are interested in is the smallest. An additional plot should be constructed based on this calculated metric to determine which races/genders are most effective at receiving the replies from races they desire.

**APPENDIX:**

**Redesigned Graphic – Repliers Plot**



**Overall Gender Differences**





1. <https://theblog.okcupid.com/how-your-race-affects-the-messages-you-get-39c68771b99e> [↑](#footnote-ref-1)
2. <https://ermlick.shinyapps.io/OKCUPID/> [↑](#footnote-ref-2)