1. I looked only at the American Community Survey (ACS). On sex, the ACS only had the options “male” and “female” and asked them to answer based on the current time. For gender and sexuality, there was no data collected. For race, the census used an established list: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White where multiple options could be selected.. Ethnicity was categorized as “Hispanic or Latino” and “Not Hispanic or Latino.”
2. We gather these data because these demographics allow for our government to determine how to allocate resources properly and what policies should be proposed. These data play a massive role in politics and society. First, schools like UVA attempt to replicate these data in their student body (even if they fall short of doing so). In politics, we see candidates targeting each demographic as a group, so they often want to know how many people fall into these categories. The quality of this data matters because inaccuracies can lead to policies that hurt groups that need help and vice versa.
3. I believe that the census collected races as well as they could; by allowing multiple races to be selected, the data includes multiracial people. Maybe they could add percentages since someone who is 50% white, 50% black is very different from someone 90% white, 10% black in terms of diversity. By not having gender and sexuality, they’re missing out on a lot of important data that could better show the diversity of the population. For sex, I believe having a “does not apply” or “prefer not to answer” would improve it since not everyone feels comfortable fitting into one category of a binary system. Large surveys should try to allow for more personalized responses, but it is also understandable to simplify where possible to streamline the data collection and analysis (for example, if someone puts down 5% asian, do you include that individual in the asian population? If not, what is the cutoff).
4. Look at answers 1-3.
5. When it comes to cleaning and analyzing data, placing people into single categories doesn’t encapsulate the full picture. There would also be people that refrain from answering, and this missing data could skew the results. This doesn’t include why the person opted out, and placing values into the missing data would further damage the integrity of the results. Lastly, cleaning data could possibly “compress” complex data, such as race and sexuality, into one “main” category. For example, a race like Asian includes various groups like Chinese, Korean, Filipino, Japanese, Taiwanese, Vietnamese, etc. and cleaning the data into one group wouldn’t capture the nuisances between each group.
6. An algorithm that imputes values for characteristics like race, gender, sex, and sexuality could be concerning ethically and socially. Giving race or gender a value could lead to bias where one group feels superior to another because they have a “higher” value. For instance, my research involves male and female mice where males are given a value “1 and females are given a value “0.” If they were humans, the males might think they’re better since 1 is greater than 0, etc. Another concern would be algorithms determining

values without being transparent about how they are calculated. For example, if a straight cisgender caucasian male is given the value of “18” while a gay transgender asian woman is given the value “29,” we would have no idea which characteristics have which value. Lastly, these characteristics are on a spectrum and putting them into discrete values could be dangerous to both the individual and society.