1. How did the most recent US Census gather data on race?

   The most recent US Census gathered data on race through a question posted to all US households as part of the census questionnaire. The format asked respondents to "Select one or more boxes and print origins" and included the categories:

   - **White**
   - **Black or African American**
   - **American Indian or Alaska Native**
   - **Asian**
   - **Native Hawaiian or Other Pacific Islander**
   - **Some Other Race**

2. Why do we gather these data? What role do these kinds of data play in politics and society? Why does data quality matter?

   The reason the Census gathers race data is because it aids the government in shaping public policy, ensuring representation, and understanding demographic trends. For example, race data aids allocating funds to school districts for bilingual services under the Bilingual Education Act. Data quality matters as that allows effective policy implementation, accuracy of representation, and responsiveness from the government.

3. Please provide a constructive criticism of how the Census was conducted: What was done well? What do you think was missing? How should future large scale surveys be adjusted to best reflect the diversity of the population? Could some of the Census' good practices be adopted more widely to gather richer and more useful data?

   The 2020 Census allowed for digital responses, making it more accessible. The race and ethnicity questions were detailed, and there were extensive outreach and awareness campaigns to complete the census. I think there could be a little more flexibility in race and ethnicity categories to be even more specific. There is still underrepresentation and miscounts, causing racial and ethnic minorities to possible unequal representation and resource allocation. Future large scale surveys could be adjusted by creating more innovative data collection methods (being digital was huge) and I think there can be more expansive community engagement to encourage more people to complete the Census, as I think people don't see it's importance. These practices can enhance participation rates, improve data quality, and ensure that data collection efforts more accurately reflect and serve the needs of a diverse population.

4. How did the Census gather data on sex and gender? Please provide a similar constructive criticism of their practices.

   The Census gathered data on sex and gender by asking participants to indicate their sex as either male or female, with no questions included to capture gender identity. This is a clear lack of inclusivity, which in turn leads

to less participants. The Census seems to be alienating those who may identify as non-binary, transgender, and so on. This lack of data leads to inadequate representation for our gender-diverse population, which can lead to lack of support in health care, education, employment, and legal protections.

5. When it comes to cleaning data, what concerns do you have about protected characteristics like sex, gender, sexual identity, or race? What challenges can you imagine arising when there are missing values? What good or bad practices might people adopt, and why?

   Concerns that I have about protected characteristics in reference to cleaning data is that traditional gender values such as "M" or "F" don't cover the whole picture anymore and disregards complex human identities, same thing with racial identity. So, data cleaning may not actually correctly represent a population. Another concern is that misuse with data cleaning can very easily lead to bias and discrimination, so those who data clean must be aware of these implications before they begin. Missing values obviously causes an issue with inaccurate representation, but imputing values would be challenging as the standard methods may not account for diversity and complexity of gender identities and can introduce bias.

   I think one good practice that people may adopt is documentation that is clear, coherent, and transparent. Documenting the process of how the data was cleaned would show rationale behind methods and reasoning and thus help limit bias and misunderstanding. Another good practice would potentially be multiple imputation techniques, which may involve models that better reflect the distribution and diversity of these protected characteristics in the population. A bad practice people may adopt is cleaning data, attempting to leave the most accounted for result, when those minority results matter as well.

6. Suppose someone invented an algorithm to impute values for protected characteristics like race, gender, sex, or sexuality. What kinds of concerns would you have?

If someone had invented an algorithm to impute values for protected characteristics, this would be a concern of consent, as it may infringe on individuals' right to self-identify by overriding their personal experiences/identities with an algorithmic guess. This would also obviously be a bias/fairness concern, as it could perpetuate existing biases or introduce new ones. Gender identity is very complex with factors that are unique to every individual, so to have an algorithm make a guess upon that is unrealistic. If this imputed data was used in any type of decision-making, like healthcare or employment, then it would most definitely be a risk of access to services and opportunities.