Respectful Collection of Demographic Data

Demographic data may be critical to your mission as a community center, legally required diversity disclosure of a corporation, or an idle curiosity of a blogger to understand their followers.

Whatever your reason, this article establishes some guidelines for respectful design of your form and language for collecting demographic data.

**An overview:**

1. Ask affected communities for their input.
2. Identify whether you truly need all of the information you ask for.
3. Explain your purpose and your privacy policy.
4. Offer multi-select checkboxes, not single-select radio buttons.
5. Allow users to self-describe.
6. Do not require a response.
7. Consider your defaults.
8. Consider the presentation and influence of your survey.
9. Learn how to write questions about gender and sexuality.

Everything that I’ve learned about creating effective and respectful demographic forms is something I learned from research or discussion. As a queer woman myself, I can tell you that our communities don’t always agree on what is the best approach, but offering...
multiple-choice responses and widely-accepted language along with an option to multi-select and self-describe help reduce these concerns.

**Ask affected communities for their input**

And reward them for their contributions! Women — and black women in particular — are frequently asked to volunteer their time and energy into helping others understand issues that affect them.

For Product Managers, incorporate feedback requests into your user research, heed their responses, and iterate upon that feedback to continuously improve your product.

Some of the best feedback you can receive is from A/B testing, or from directly asking users which design they prefer. However, where marginalized communities are affected, beware that an affected minority group may respond distinctly from the majority, and that opinions among that minority group may be split. Be prepared to reject all designs in favor of a new design that can accommodate all known concerns.

**Identify whether you need the information you ask for**

A friend recently told me about an uncomfortable experience with her doctor, in which her doctor asked about her partner’s gender. The doctor wanted to advise her on risks. Another friend’s doctor used a different approach: she explained the risks of certain sexual activities, and she offered a couple different options for treatments and tests they might consider.

An employer might ask, “Are you available to work on Saturday?” instead of asking about an employee’s religious practices.

Many services ask about the user’s gender when they mean to ask what pronouns they should use. Some services use gender neutral pronouns (they/their) for all users by default, thereby skipping the matter entirely, and most users don’t notice — and have you noticed that LinkedIn doesn’t use third-person pronouns for users at all?
Explain your purpose and your privacy policy

Users will be more comfortable providing answers if you explain how you will use the data, how it benefits them, and how you will protect the data they provide.

Image offers two form examples. The design on the left asks for Pronouns and has a pop-up explanation, “Our product refers to users using the pronouns that you select below. E.g., “This user updated their profile”. The form offers three pronoun choices, and a textbox for the user to write their own pronouns. This design informs your users and lets them make flexible choices that meet their needs. The design on the right offers only two options for Gender (male, female), and no explanation. The design on the right is not recommended.

The example above communicates to the user why their choice is relevant to them, and how it is displayed to other users. Ideally, the form above would also demonstrate use of the self-description provided with an example.

State clearly in your form who will have access to the data that you collect, and be particularly clear on what data other users can view. Be specific with what you do with the data that users provide, and refer them to your privacy policy for details. Personally Identifiable Information (PII) may be subject to privacy laws. If you are unsure which laws may be applicable and how to protect yourself and your users, please consult with an attorney.
Also, as long as we’re talking about privacy — if your form is on the web, you have an HTTPS certificate for SSL/TLS encryption of all transmitted data, right? You know it’s free and easy, right?

Do not store user data associated with the user, unless you need it.

**Offer multi-select checkboxes, not single-select radio buttons**

Particularly when it comes to details about a user’s identify, it is critical to offer multi-select checkboxes in any situation when a user could potentially identify as multiple options that you offer.

**Allow users to self-describe**

Users feel dehumanized when they are forced to make choices between discrete categories that don’t fit them. Data-collection purposes like advertising are often driven by stakeholders that require categorical data. However, if the options you present are representative of your audience, most users will select not to self-describe. Periodically, sample user self-descriptions to determine if you should add new options to your form.

Image offers two form examples. The design on the left asks the user to identify their ethnicity (select all that apply), and offers a multi-choice list with checkboxes, including an option “Prefer not to answer” and an open
textbox for self-description. This design allows your users to select multiple options, choose not to answer, or write their own description. The design on the right offers a list with single-choice radio buttons, including options for “Mixed Race” and “Other”. Users feel more dehumanized when they are forced to make a single choice from a list like this. The design on the right is not recommended.

Update: Using “Caucasian” instead of “white” can make some people uncomfortable. Making people uncomfortable isn’t necessarily bad if you have an educational purpose, but you want people to understand and complete your form.

Demographic forms often distinguish “race” from “ethnicity”, particularly in the case of identifying whether a person is hispanic or non-hispanic. The FDA guidelines for collection of race and ethnicity recommend the “two-question format” for this reason. Some forms ask for “race and ethnic background”, or avoid using those words entirely, as a possible 2020 U.S. Census question below suggests:

Possible 2020 census question: “Which categories describe Person 1? Select all boxes that apply.” Options include “White”, “Hispanic, Latino, or Spanish origin”, “Black or African American”, “Asian”, “American Indian or Alaska Native”, “Middle Eastern or North African”, “Native Hawaiian or Other Pacific Islander”, and “Some other race, ethnicity, or origin”.

The language used in the example above is mostly good, though self-description is always preferable. While the census data may end up collecting all self-description as “other”, some users feel validated when they have the opportunity to write their own entry when other entries do not accurately describe them.

The Spanish language has binary gendered forms for “Latino” and “Latina”. In the last few years, a movement has started to replace Latino/Latina with Latinx to be inclusive with a single word of Latin American people — not just of women and men, but of nonbinary Latinx people. While Latin@ also has some popularity, its form merges Latino and Latina and is not inclusive of all nonbinary Latin American people.

Do not require a response

If you absolutely must ask a question (e.g., for advertising purposes), if it’s feasible, offer an option for the user to choose not to respond.

That being said, if you can elect not to require a response, did you need to ask for this in the first place? Best approach: don’t ask. Second-best approach: allow user to opt-out of responding to a particular question (which allows them to answer other questions).

Consider your defaults

Current American media and technology often offers white men as the default option and the default representation. If you must select a default, or you must select an order in which to present elements of a set, challenge your own assumptions about your users, and challenge your users’ assumptions by consciously choosing an order that defies societal power structures.

Challenge our assumptions that our users are American, white, male, cisgender, able-bodied, and hearing. When users see a default value selected for them (e.g., Country: United States of America), this may be convenient if the default matches the majority of your users. If you challenge this assumption and set a distinct default or ask the user to select from a list that does not prioritize the majority option, the user is driven to think about the variety of other users that might be using your product.

In a game, a player might recognize that nearly half of all players are female if the default player model is female. It might be jarring for some users, but that experience could drive a realization about the diversity in the user community. Users who do not
often find that the default values fit them will frequently be pleased to be represented even in a form, which generates goodwill for your company.

While representation is gratifying, don’t make your demographics collection a smorgasbord of options, which dehumanizes and can offend users. Terminology in marginalized communities evolves, and often outpaces language used by those with the social power to publish. The terms you see used in the NYTimes or Oxford University Press are sometimes years or decades behind currently accepted terminology. (*A November 2016 NYTimes headline reads “No Rest at Rest Home: Fighting Bias Against Gays and Lesbians”*. The article proceeds to repeatedly refer to “gays and lesbians”. The National Association of Black Journalists writes, “In news copy, aim to use black as an adjective, not a noun. Also, when describing a group, use black people instead of just blacks. In headlines, blacks, however, is acceptable.” A similar attitude is prevalent in the gay and lesbian community: *use gay as an adjective, not as a noun.*)

Consider the presentation and influence of your survey
Default values can lead to survey bias. For example, an impatient user may accept default values to complete the survey faster. Impatient users may also select the first option available. To reduce these biases in your data, consider randomizing default values (if you must have a default value) and randomizing the order of the options presented.

Psychology studies often find that participants respond differently after being asked for demographic data. If your demographics collection is associated with a task that you want users to perform (e.g., user research or job applications), consider asking for demographics after they complete other tasks.

Approaching Gender and Sexuality
When I redesigned the demographics form for the Sacramento LGBT Community Center, the Human Rights Campaign and GLAAD had not yet released any guidance on surveys or writing about trans people. I reached out to the trans community to get feedback on the forms I was designing, and I learned that many surveys trying to be inclusive were unintentionally offensive.

At the Sacramento LGBT Community Center, the demographic information we gathered about people who used our services was used to apply for grants that support gender,
romantic, and sexual minorities (GRSM). The trans community pointed out that they were frequently asked about their trans status when it was not relevant. They felt less comfortable participating in a survey or using products when they were asked about their trans status without an explanation of why that was necessary or how their privacy would be protected.

Image offers two form examples. The design on the left asks two questions, “Gender” and “Do you identify as transgender?” The choices for the first question include female, male, nonbinary, prefer not to say, and an open textbox. The choices for the second question include yes, no, and prefer not to say. Both questions have blue question marks, which would explain why those questions are relevant (the explanations are not given).

This design allows binary users to identify their gender without disclosing their trans status, allows self-description, and allows users to choose not to share. The design on the right asks for Gender and offers three choices: male, female, and transgender. This design forces binary transgender people to choose between identifying their gender (male or female) or their trans status. The design on the right is not recommended.

The Human Rights Campaign recommends asking for transgender status in a separate question from gender and from orientation. Because this is a particularly private and sensitive piece of information, clarify to your users why you are asking this. Before you ask this of users, justify to your team why you need this data. If you don’t need it, don’t ask for it.
Update: The example above should allow selection of multiple gender options — for example, to allow a user to state that they are both non-binary and female, or male and self-describe to clarify that they prefer to use the term transmasculine.

Image says, “I identify my gender as”, and offers five options with single-choice radio buttons: man, transgender man, woman, transgender woman, and non-binary. This is not recommended.

Surveys often offer “transgender” as a gender choice, or offer separate choices for “trans woman” and “trans man” from “woman” and “man”, as in the above example. Most transgender people are within the gender binary (male or female), and binary trans people feel othered when forms imply that, for example, “trans man” is a separate category from “man”. Trans men are men. Trans women are women.
Image asks, “What gender identity do you identify with? Please select all that apply”, and offers 21 options, including “Human Being”, “Decline response”, and “Other, please specify”. This is not recommended.

The form above allows the user to select multiple options, implying that categories are not mutually exclusive, and allows the user to decline response or self-describe — all of which are commendable survey features. However, some of the options here are not “genders”, and the smorgasbord presentation is dehumanizing to trans people. This is not a recommended design.

Image asks, “I identify my sexual orientation as”, and offers 6 options: Straight/Heterosexual, “Gay or Lesbian”, Bisexual, Transgender, Queer, and Asexual. This is not recommended.

The form above restricts the user to a single option, suggesting that these categories are mutually exclusive or that a user must select only one aspect of themselves if multiple options apply. Transgender is neither a gender nor a sexual orientation — be careful in trying to ask in a single question what should be separated into multiple. The form also lacks a self-description option, which is critical for gender and sexuality.
Multiple of the examples above use the word “identify” or “identity”. A nonbinary friend recently pointed out to me that straight, cisgender, binary people have orientations, genders, and pronouns, while often queer and trans people are othered by emphasizing terms like “preferred” pronouns or gender “identity”. Avoid using these unnecessary adjectives that could feel derisive to queer or trans people.

Conclusions

Focus on the demographic information that you need, demonstrate interest in the human characteristics of the users you are surveying, and ask users to express aspects of themselves that are relevant to your product or service (e.g., to determine how you can improve accessibility or marketing of your product, or improve the community).

Before designing your survey, challenge your assumptions and question what details you need to collect from your users! Recognize that asking for sensitive or private details about your users is a significant ask: justify your need, and reward your users.

Further resources:

- How to Write Better Demographic Survey Questions provides a different perspective with a wider variety of demographics questions than I covered
- GLAAD Media Reference Guide is a spectacular collection of recommendations for writing about queer people
- Collecting Transgender-Inclusive Gender Data from the Human Rights Campaign

Credit:

I appreciate the feedback and suggestions I’ve received from Amaya, Aria, Christina, Crystal, Gigi, Heather, Indy, Jon, LaToya, Liz, Melissa, Nicole, Rin, and Sophie!
About the author

Sarai Rosenberg is a mathematician, security engineer, and queer femme woman dismantling systemic barriers in tech, one fencepost problem at a time.