A Note on the Treatment of Gender and Race in My Statistics Notes – DRAFT

Richard Williams, University of Notre Dame, https://academicweb.nd.edu/~rwilliam/
Last revised August 26, 2024

My personal web pages at https://academicweb.nd.edu/~rwilliam/ contain dozens, perhaps hundreds of handouts on statistics. Early versions of some of these first saw life more than 30 years ago. Many of these are from classes I no longer offer and may never get revised again, while others will continue to be tweaked as I teach courses using them.

There have been many changes in statistical practice and preferred wording over those 30 years. Traditionally most studies have (and perhaps still do) treat gender as binary and fixed at birth. These practices are unfortunate, not only because they are inaccurate, but because they can perpetuate harm experienced by transgender and non-binary communities (Cameron & Stinson, 2019). I became acutely aware of such problems when one of my own students said that my terminology and examples sometimes made them feel "unwelcome" and "uncomfortable" in class.

It has also been common to treat race as an objective trait fixed at birth, and to oversimplify and misuse the concept of race. Preferred wording has also changed over time. While it used to be common to use the terms *black* or *blacks* (note the lower-case b's) today it is often seen as better to use phrases like *Black Individuals* or *Black People* or to at least capitalize *Black*. (See the references for historical and ongoing debates about racial terminology.)

My notes (and for that matter probably most of my publications) reflect common past practices. As I update handouts, I am making changes to reflect current, more inclusive terminology and methods. When I revise my notes, I am not going to engage in revisionist history and pretend studies were conducted differently than they were. Terms like *female* and *black* will continue to appear as they did in survey questions and in value and variable labels, which are often very short anyway to keep output easily readable. But, when interpreting results, I am going to try to use more contemporary phrasings, e.g. refer to *Black Individuals* rather than *blacks*. I'll also work on this verbally and try to avoid slipping into old habits.

I'd also like to incorporate examples where gender is not treated as binary and unchanging. If anyone knows of good empirical (and preferably replicable) examples where this is done I'd love to hear about them. I'll also periodically remind people that if studies I use in examples were being conducted today, they might handle race and/or gender differently.

Cautionary Note: Small Ns may greatly limit possible alternative gender identity analyses though. A recent study estimated that 1.2 million American adults identify as nonbinary. That is a lot of people, but in most national studies only a few nonbinary individuals would be included. As explained below, the General Social Survey made major changes in its questions on gender identity in 2018. But, out of 1409 respondents, only 9 cases were identified as transgender or another alternative gender designation. As a practical matter, I suspect most quantitative analyses will be limited to treating sex as binary, unless the sampling strategy has explicitly targeted collecting data from respondents with less common gender identities. Some promising-sounding data sets that have done so are listed in the appendix.

Appendix: Additional Resources

My thanks to those who have made suggestions on how my handouts can be improved. I appreciate everyone's patience as I work on them further. Here are some additional materials on relevant publications and data sets that may be useful, especially for those who would like to do research involving less common alternative gender identities.

I.　References on the measurement of gender identity

Cameron, JJ, Stinson, DA. Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. Soc Personal Psychol Compass. 2019; 13:e12506. https://doi.org/10.1111/spc3.12506. This article also offers good suggestions for what researchers should to do instead.

Carian, Emily K. 2019. More Inclusive Gender Questions Added to the General Social Survey. https://gender.stanford.edu/news-publications/gender-news/more-inclusive-gender-questions-added-general-social-survey. Retrieved January 4, 2021.

Mize, T. D. (2015) What Social Psychology Can Contribute to the Study of Sex, Gender, and Sexual Orientation. *Sociology Compass*, 9: 1066– 1081. https://doi.org/10.1111/soc4.12331

Smith, Tom W., and Jaesok Son. 2019. Transgender and Alternative Gender Measurement on the 2018 General Social Survey. *GSS Methodology Report No. 129*. http://www.gss.norc.org/Documents/reports/methodological-reports/MR129%20transgender%202018.pdf. Retrieved January 4, 2022.

II.　References on Racial Terminology & the Very Concept of Race

Racial terminology is hotly debated and has been for several decades. Terms that might today be viewed as antiquated or even racist today were often not viewed that way in the past. For that matter, the very concept of race has been challenged. For some good discussions, see:

Darren Curnoe claims that classifying people by race is the biggest mistake in the history of science. https://theconversation.com/the-biggest-mistake-in-the-history-of-science-70575

In three is not enough, scientists debate whether or not race is a useful concept. https://www.newsweek.com/three-not-enough-184974

Nicholas D. Kristof asks Is Race Real? https://www.nytimes.com/2003/07/11/opinion/is-race-real.html

Eligon, John. 2020. A Debate Over Identity and Race Asks, Are African-Americans 'Black' or 'black'? https://www.nytimes.com/2020/06/26/us/black-african-american-style-debate.html

Singh, Amerdeep. Last updated 8/2/2024. A Note on Historical Terminology. https://scalar.lehigh.edu/african-american-poetry-a-digital-anthology/a-note-on-historical-language-is-negro-a-racial-slur Part of the African Poetry: A Digital Anthology Project. Retrieved 8/22/2024.

Appiah, Kwame Anthony. 2020. The Case for Capitalizing the B in Black. https://www.theatlantic.com/ideas/archive/2020/06/time-to-capitalize-blackand-white/613159/

Wong, Brittany. 2020. Here's Why It's A Big Deal To Capitalize The Word 'Black'. https://www.huffpost.com/entry/why-capitalize-word-black_l_5f342ca1c5b6960c066faea5

III. Possible Data Sets that include gender-identity questions

Many data sets have good measures of sexual orientation, but I suspect good data related to sex and gender identity (e.g. transgender, intersex, non-binary) may be harder to find. Questions about gender identity are often not asked, and when they are the Ns for people who list themselves as transgender/intersex/non-binary may be too small to provide useful information.

The General Social Survey. https://gss.norc.org/. The General Social Survey (GSS) is a nationally representative survey of adults in the United States conducted since 1972. Based on the recommendations of several social scientists, the GSS modified its questions on gender identity in 2018. For the first time in GSS history, a two-step gender question asked respondents to separately identify the sex they were assigned at birth and their current gender (Carian 2019). The first question reads, "What sex were you assigned at birth? (For example, on your birth certificate)" and allows respondents to select "Female," "Male," "Intersex," or "No answer." The second question asks, "What is your current gender?" Respondents were able to select "Woman," "Man," "Transgender," "A gender not listed here," and "No answer." *But, out of 1409 respondents, only 9 cases were identified as transgender or another alternative gender designation.* However, a person at the GSS told me that *the Ns may get higher, more questions are coming, and that gender scholars are excited for the new content*. It will be exciting to see what the GSS produces in the future.

Social Justice Sexuality Project: 2010 National Survey. https://www.icpsr.umich.edu/web/RCMD/studies/34363. The Social Justice Sexuality Project (SJS) is one of the largest national surveys of Black, Latina/o, Asian and Pacific Islander, and multiracial lesbian, gay, bisexual, and transgender (LGBT) people. The purpose of the SJS Project is to document and celebrate the experiences of lesbian, gay, bisexual and transgender (LGBT) people of color. It has many questions I haven't seen anywhere else. For Gender Identity, 2279 listed themselves as male, 2162 were female, 111 were trans M-F, 62 were trans F-M, 132 were other, 28 claimed multiple identities, and 179 were missing. *That is over 300 cases with alternative gender identities, which is far higher than I have seen in most data sets*.

LGBTQData.com. http://www.lgbtdata.com/. LGBTData.com serves as a no-cost, open-access clearinghouse for the collection of sexual orientation & gender identity data and measures. The website encourages the collection of sexual orientation data and the analysis of data sources that

have already collected such data. While a few of these data sources were designed to collect data primarily from LGBs, most of them are large surveys that decided to include a sexual orientation variable. I haven't checked these sources out.

LGBTQ datasets available on data.world. https://data.world/datasets/lgbtq. The site currently lists three LGBTQ datasets that it has available. I haven't checked these sources out.

Searchable LGBTQ Resource Database. https://lgbtq-ta-center.org/resource/. The California Reducing Disparities Project (CRDP) aims to implement and evaluate community-based programs that support mental health among California's Lesbian, Gay, Bisexual, Transgender, and Queer/Questioning populations. *I'm not sure if you can actually find data on the site, but it does provide links to 100s of resources related to LGBTQ-related topics*.

2015 U.S. Transgender Survey (USTS) (ICPSR 37229). https://www.icpsr.umich.edu/web/ICPSR/studies/37229. The 2015 U.S. Transgender Survey (USTS) is the largest survey ever conducted (27,715 respondents) to examine the experiences of transgender people in the United States. The survey included questions related to the following topics (in alphabetical order): accessing restrooms; airport security; civic participation; counseling; education; employment; family and peer support; health and health insurance; HIV; housing and homelessness; identity documents; immigration; income; intimate partner violence; military service; police and incarceration; policy priorities; public accommodations; faith; sex work; sexual assault; substance use; suicidal thoughts and behaviors; unequal treatment, harassment, and physical attack; and voting. *There are no publicly available data files for this study, but researchers can apply for access*. You can download documentation for the study.

National Transgender Discrimination Survey, [United States], 2008-2009 (ICPSR 37888). https://www.icpsr.umich.edu/web/ICPSR/studies/37888. The National Gay and Lesbian Task Force and the National Center for Transgender Equality are grateful to each of the 6,450 transgender and gender non-conforming study participants who took the time and energy to answer questions about the depth and breadth of injustice in their lives. These data provide information on transgender discrimination in every major area of life, including housing, employment, health and health care, education, public accommodation, family life, criminal justice and government identity documents, and demographic information such as citizenship, race, ethnicity, employment, and income. In virtually every setting, the data underscores the urgent need for policymakers and community leaders to change their business-as-usual approach and confront the devastating consequences of anti-transgender bias. *Data can be downloaded in Stata format. It appears to have many interesting variables you usually don't see in other studies.*

Resources for Research on Gender Identity, Sexual Behavior, and Sexual Orientation. https://www.icpsr.umich.edu/web/pages/RCMD/lgbtq-resources.html. (These resources are provided by the Resource Center for Minority Data, described below.) Here you'll find data resources for those interested in studying gender identity, sexual behavior, and sexual orientation, or in finding datasets in which those populations are well-represented. It includes data resources available at ICPSR and elsewhere. For instance, you'll find data from studies that focus on these populations and data from broader studies that contain questions that enable

researchers to identify participants in these populations. Currently about 40 datasets are listed in a downloadable spreadsheet.

You might also check out **Celebrating Pride Month with Resources From ICPSR**. https://www.icpsr.umich.edu/web/pages/RCMD/news.html?node=3290. Besides highlighting various data sets, it has interesting video and audio presentations. For example, "Quantifying Gender Identities and Behaviors" (https://www.youtube.com/watch?v=1j1U6b436QA) reviews challenges in creating gender inclusive categories in surveys using real-world examples (e.g. what if your sample has 1 trans person and everyone else is cis male/female, for example?) It also discusses how researchers, particularly grad students and undergrads, are using social media to target distribution of surveys to transgender populations.

IV. Data Sets that focus on racial and ethnic minorities

Many, perhaps most, data sets have questions on race and ethnicity. But, some data sets focus on racial and ethnic minorities much more than do others.

Resource Center for Minority Data (RCMD; Also referred to on their web pages as Really Cool Minority Data). https://www.icpsr.umich.edu/web/pages/RCMD/about.html, https://www.icpsr.umich.edu/web/pages/RCMD/mission.html
RCMD is a recent initiative of ICPSR (Inter-university Consortium of Political and Social Research). The mission of the Resource Center for Minority Data (RCMD) is to provide educators, researchers, and students with data resources to analyze issues affecting racial and ethnic minority populations in the United States. RCMD provides access and analytic tools for use with the vast array of minority data available from ICPSR. RCMD enhancements assist researchers, educators, and students in their quest to effectively utilize data on underrepresented populations. RCMD provides links to over 400 downloadable data sets, sortable by topic area (e.g. education, discrimination, immigration, health & well-being), at least some of which are already in Stata format.

Coursework offered by RCMD. To raise awareness of data resources about racial and ethnic minority populations, the ICPSR Summer Program offers (or at least has offered) a four-week course titled "Methodological Issues in Quantitative Research on Race and Ethnicity." This course examines research design, measurement, and analysis issues involved in research on minority populations in the United States. Topics include
- Reliable and valid measurement of racial and ethnic variables in major survey studies
- Innovative sampling strategies for diverse populations
- Race-related measurement errors in data collection
- Analysis traditions and trends within various social/behavioral science field

For more detail see https://www.icpsr.umich.edu/files/RCMD/brochure.pdf

Note: My thanks to Mara Kraemer for drawing my attention to the RCMD. I haven't used it yet myself but it looks like it has a lot of great resources that are freely available.