

Data for queer lives: How LGBTQ gender and sexuality identities challenge norms of demographics

Big Data & Society
January–June: 1–12
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951720933286
journals.sagepub.com/home/bds

Bonnie Ruberg¹ and Spencer Ruelos²

Abstract

In this article, we argue that dominant norms of demographic data are insufficient for accounting for the complexities that characterize many lesbian, gay, bisexual, transgender, and queer (LGBTQ, or broadly “queer”) lives. Here, we draw from the responses of 178 people who identified as non-heterosexual or non-cisgender to demographic questions we developed regarding gender and sexual orientation. Demographic data commonly imagines identity as fixed, singular, and discrete. However, our findings suggest that, for LGBTQ people, gender and sexual identities are often multiple and in flux. An overwhelming majority of our respondents reported shifting in their understandings of their sexual identities over time. In addition, for many of our respondents, gender identity was made up of overlapping factors, including the relationship between gender and transgender identities. These findings challenge researchers to reconsider how identity is understood as and through data. Drawing from critical data studies, feminist and queer digital media studies, and social justice initiatives like Data for Black Lives, we call for a reimagining of identity-based data as “queer data” or “data for queer lives.” We offer also recommendations for researchers to develop more inclusive survey questions. At the same time, we address the ways that queer perspectives destabilize the underlying logics of data by resisting classification and “capture.” For marginalized people, the stakes of this work extend beyond academia, especially in the era of algorithms and big data when the issue of who is or is not “counted” profoundly affects visibility, access, and power in the digital realm.

Keywords

Data, digital media, identity, LGBTQ, gender, sexuality

Introduction: Queer data, queering data

The issue of data and the ways in which it does or does not ethically represent marginalized people is of pressing importance at the contemporary moment, especially in the context of digital media and the documented rise of big data (Gieseeking, 2018). As many scholars of media and communication have argued, users’ personal data has become the “currency” of the contemporary online environments (Sadowski, 2019), “driven by the ascendancy of algorithmic structures that trace, track, combine, compare, and predict our moves in the digital realm” (Huntemann, 2013: 43). Cultural values shape and are shaped by the ways that data is imagined and put to use; embodied histories of power and oppression underlie assumptions about what data is, “who counts, in what ways, and why” (Wernimont, 2018: xiii).

Today, such data has the ability to perpetuate discrimination, for example when search engine algorithms reinforce racial biases (Noble, 2018) or predictive policing applications recreate patterns of profiling and harm (Benbouzid, 2019). At the same time, data science can be used to “create concrete and measurable change” in the lives of those very people whom data itself has often misrepresented or even oppressed, as the organizers of the research group Data for Black Lives compellingly

¹Department of Film and Media Studies, University of California, Irvine
²Department of Informatics, University of California, Irvine

Corresponding author:

Bonnie Ruberg, University of California, Irvine, CA, USA.
Email: bruberg@uci.edu



state (<http://d4bl.org/about.html>). Whether we see data as a tool of control or a potential instrument of empowerment, its influence is undeniable, raising the need for new critical frameworks that move beyond positivist approaches and toward questions of epistemology (Resnyansky, 2019). For these reasons, it is crucial to critique and reimagine the role that data plays in both constructing and valuing—or devaluing—the identities and experiences of people who have been pushed to the margins, both of contemporary Western society and the digital realm. Demographic data, through which individuals and populations are categorized and counted, plays a particularly important role in setting the terms for identity. It therefore represents a key site of intervention for discussions of data and social justice.

In this article, we argue that dominant standards of understanding and collecting demographic data—such as in research surveys or government censuses—are insufficient for accounting for the complexities of lesbian, gay, bisexual, transgender, and queer (LGBTQ, or broadly “queer”) lives. Building from work by feminist scholars like Carole M. McCain (2017: 32), we understand the conceptualization of demographic data as “a historically situated, politically inflected, interpretive act” that translates people, their identities, and often their bodies into pre-determined statistical categories. As we show here, traditional notions of demographic data do not allow for the fluidity and multiplicity of gender and sexual identities that characterize the lived experiences of many LGBTQ people. Our research suggests that a considerable percentage of people who currently identify as non-heterosexual or non-cisgender have shifted in their understandings of their sexual identities over time. In addition, our research demonstrates that, for many LGBTQ people, the complicated relationship between gender identity and transgender identity destabilizes the dominant assumption that gender can be effectively described along a single axis of data. Drawing from these findings, we argue that an individual’s sexual and gender identities, especially for LGBTQ people, cannot be understood as a set of static, fixed data points. Rather, these identities are sites of complex temporality and intersectional multiplicity. Indeed, this friction between LGBTQ lives and data as it is commonly conceptualized challenges us to reconsider the logics of data itself, inspiring us to “queer” accepted notions of how identity is categorized, quantified, captured, and rendered into meaning. Written in the lead-up to the 2020 U.S. census, amidst debates about the census’ relative dearth of questions regarding LGBTQ identities (Lang, 2019) and calls to “queer the census” (<https://www.thetaskforce.org/thanks-for-keeping-the-census-queer/>), this work has timely implications for

broader discussions of data and the cultural politics of demographics.

For this article, we draw our insights from the design and results of a survey that we conducted in February 2019. This survey asked adult respondents currently in their mid-20s to mid-30s about their participation in sexual activities and engagement with sexual materials on the internet during their pre-teen and teen years. As part of this survey, we included demographic questions regarding participants’ sexual and gender identities; these questions (rather than respondents’ online sexual activities) are the focus of our work in this article. Because of our own backgrounds in gender studies and queer studies, we intentionally designed our survey’s demographic questions to allow for complexity in how respondents self-reported their sexual and gender identities. Of our respondents, 178 described themselves using identity terms that denote them as non-heterosexual or non-cisgender (i.e. “queer”). Among those respondents, a striking 83% reported that their sexual identities had changed between their teen years and the present. Another 13% chose more than one term to describe their gender identities and/or described themselves through a combination of markers related both gender and transgender identity. These findings demonstrate that, contrary to common perception that “diverse” individuals have specific, fixed sexual and gender identities that remain true over the course of their lifetimes, the sexual and gender identities of many LGBTQ people are in fact neither static nor singular. Such identities are likely to shift over time or to contain multiple elements that are commonly imagined, outside of LGBTQ communities, to be mutually exclusive. Our research shows that LGBTQ people often understand their own sexual and gender identities as overlapping, incomplete, or in flux. This is due to factors that are at once personal and cultural, arising from shifting experiences of the self and processes of self-exploration, hetero- and cis-normative societal pressures, evolving concepts and terms that emerge from within LGBTQ communities, and queer understandings of identity as nuanced and individual.

Based on these findings, we argue for a shift in how identity is understood as and through data. This shift is simultaneously conceptual and pragmatic. In part, we are offering prompts to encourage fellow researchers, when creating surveys or other mechanisms for collecting demographic data, to reflect on how they can design questions that respond to a call for “data for queer lives” (to mirror the language of *Data for Black Lives*): that is, questions that more accurately and respectfully engage with the complexities of LGBTQ identities and experiences so that the data they produce can better serve LGBTQ people. This includes

constructing questions that account for how respondents' sexual identities may have shifted over time and that invite respondents to describe their gender identities using multiple markers—even (or perhaps especially) when these complexities make demographic data ill-suited to current standards of quantitative data analysis. At the same time, our aim is broader. It is our contention that LGBTQ lives, when considered alongside questions of data, require us to radically reimagine how we conceive of data itself, especially as it relates to and attempts to represent marginalized identities. As a political project, our research serves as an argument for what we describe as “queer data.” It prompts those who research the cultural implications of data (or simply those who use demographic data in research that they envision to be socially just) to attend to how the experiences of LGBTQ people are or are not reflected in the current data standards. In addition, a meaningful consideration of LGBTQ lives opens opportunities for queering data itself—that is, for shaking loose its heteronormative assumptions and destabilizing the very belief that demographic data can sufficiently reflect the realities of identity. The stakes of this work also extend beyond academic conversations into the lived experiences and histories of LGBTQ people. This is especially true in an era of algorithms and big data, when data increasingly shapes power, access, governance, and culture in the realm of the digital and beyond, raising the stakes of these questions about whose identities are or are not “counted.”

Critical (queer) approaches to demographics and data

Many scholars from critical data studies and related areas have addressed the cultural biases of data and the discriminatory ideologies that underlie data's historical formulations (Wernimont, 2018). This article builds from and contributes to that work by offering a concrete demonstration of how existing models of demographic data—such as those typically deployed in research surveys and government censuses, which often imagine information about an individual's identity to be fixed, discrete, self-evident, and “true”—are not fit to LGBTQ lives. As described by Rob Kitchin and Tracey P Lauriault (2014: 5), critical data studies “applies critical social theory to data to explore the ways in which they are never simply neutral, objective... representations of the world, but are situated, contingent, relational, contextual, and do active work in the world.” Although many still believe that data and algorithms are objective (Glantz and Martinez, 2019), both data itself and the ways in which meaning

is made from that data, such as through data visualizations or algorithmic calculations, are fundamentally constructed and subjective—always “cooked” and never “raw” (Bowker, 2005; Gitelman and Jackson, 2013). Much of this critical data studies work has been done against the backdrop of big data, which both enables and constrains certain elements of culture (Dalton and Thatcher, 2014). Big data, which is often structured around problematic assumptions and biases, has been the object of much-needed social critique (Boyd and Crawford, 2012). Some of these critiques relate explicitly to queer lives, such as in Jen Jack Giesecking's call for a “queer feminist approach to the scale of big data,” in which Giesecking argues that the big-ness of big data undermines the importance of groups that have been made “small” by histories of erasure and violence. Giesecking writes (2018: 150), “Society's obsession with big data further oppresses the marginalized by creating a false norm to which they are never able to measure up.” Other critiques, like Shaka McGlotten's (2016) “Black Data,” respond to big data through the intersectional perspectives of queer people of color. Indeed, it is especially important to confront data as it affects the lives and bodies of those whom data has, across its cultural history, sought to regulate, surveil, devalue, and even dehumanize including LGBTQ people but also women and people of color (Johnson, 2018; Posner, 2015).

Particularly notable for our work are two threads within existing critical data studies research: writing on demographics and writing on queerness. Demographic data commonly understands individuals as fitting into stable, unchanging identity categories, using quantification to “‘freeze’ the subject, just like a substance within the chemical periodic table, where one is born a certain element” (Terranova, 2000: 41). Notions of demographic data are fundamentally tied to census data, through which governments record, measure, and (importantly) mold collective identities (Browne, 2010: 234). Questions about gender and sexuality are increasingly common elements of demographic data, and demographic data has, in turn, often been used to dictate the terms of socially acceptable gender and sexual identities. For example, McCain writes about women's reproductive health and how commentators in the United States have used demographics to promote racist moral panics around birth rates in the Global South (McCain, 2017). Queer approaches to data intersect with demographics, which are inherently tied to identity. As Kat Browne and Catherine J Nash (2010: 1) ask in their introduction to *Queer Methods and Methodologies*, “If, as queer thinking argues, subjects and subjectivities are fluid, unstable and perpetually becoming, how can we gather ‘data’ from those tenuous and fleeting subjects

using the standard methods of data collection?” Some scholars have called for alternative approaches to understanding quantification and its importance for LGBTQ histories and communities, such as in Michelle Schwartz and Constance Crompton’s (2018) writing on self-published lists of lesbian literature and authors. Others have pushed back against quantification and categorization themselves, underscoring how queer theory destabilizes hegemonic notions of identity classification (Drabinski, 2013: 96) and reminding us, in the vein of queer of color scholarship, that identities are always intersectional and multiple. Challenges of LGBTQ classification also face archivists and librarians (Adler, 2017), prompting calls for collaboration between LGBTQ communities and those who establish the labels through which such communities are categorized (Baucom, 2018).

To an extent, the work that we present here joins in conversations about not only the problems with how data is conceptualized and collected, but how to make that data better. Especially relevant is existing writing on better ethical practices for the design of survey questions about demographic data that relate to gender and sexuality identity. However, as we discuss at length below in our “methods” section, many of these current guidelines still fall short of accounting for the complexities of LGBTQ lives. As Scheurman et al. write in their online resource “HCI Guidelines for Gender Equity and Inclusivity” (<https://www.morgan-klaus.com/sigchi-gender-guidelines>), “how to respectfully report gender on surveys has been a tricky question” for many researchers who work at the intersection of technology and human concerns. Current approaches to improving gender and sexual inclusivity on surveys often emphasize increasing the number of possible options that respondents can choose from to describe their identities. This also manifests in the personal data collected by large social media corporations like Facebook, which now offers roughly 50 different gender options for users to choose from (Walker 2019)—though, as Rena Bivens (2017) has pointed out, the 2014 Facebook update that brought these additional options did not change the site’s underlying algorithmic gender binary. While we agree that offering users a wider range of gender options is a valuable step toward inclusivity, our work demonstrates that allowing users to choose only *one* out of a list of identity options remains insufficient to account for the multiple, overlapping experiences of self that often characterize queer identity. At the same time that we are pushing for better approaches to representing LGBTQ lives through demographic data, we remain aware and indeed wary of the negative implications of making those lives countable and counted. As Joanna Drucker (2011) writes, rather than thinking of data as

“data” (that which is given, like a given truth), we must think of it as “capta” (that which is created and also *captured* or taken). When it comes to identity and desire, the relationship between data and capta can be a particularly complicated one, as Patrick Keilty writes in his analysis of how tags on the pornography website Xtube both emerge from and codify the folksonomies of sexual subcultures (2012). Susan Stryker and Paisley Currah and their contributors tie this concern to queer and transgender lives in their special issue of *Transgender Quarterly*, “Making Transgender Count” (2015). Even as we strive for social justice through and within data, we must acknowledge the worrisome tension in calling for marginalized lives to be better “captured,” translated into data, and put to use by corporations and regulatory bodies.

Because readers may not be familiar with the concepts related to gender and sexual identity that have emerged from LGBTQ communities and queer studies, we pause here for a note on terminology. In referring to *LGBTQ people*, such as the LGBTQ respondents to our survey, we are referred to all people who identify as non-straight and or non-cisgender. When we talk about *sexuality*, we understand it to mean “the way in which we experience and express ourselves as sexual beings” (Kannabiran et al., 2011: 696). Put another way, sexuality represents the summative components of sexual interests, identity, orientation, performances, expressions, and desires. When we talk about *sexual identity*, we are referring to the identities related to one’s sexuality. This might mean one’s identity as gay, lesbian, queer, bisexual, asexual, heterosexual, etc. *Sexual orientation* is closely related to sexual identity. It describes an individual’s sexual and romantic interests in others. The word *queer* has two primary, interrelated meanings. In the simplest terms, it is both an umbrella term for those in the LGBTQIA+ spectrum and a descriptor of identities, experiences, and political positions related to gender and sexuality that resist dominant societal norms (Tongson, 2017: 157)—although calls have come from within queer studies itself to question the relationship between queerness and antinormativity (Wiegman and Wilson, 2015). We understand *gender* as distinct from biological sex. Gender represents a set of identities (not simply a “male” and “female” binary) that are at once culturally constructed and deeply personal. *Transgender* refers to individuals whose gender does not match the gender they were assigned at birth. Conversely, *cisgender* refers to individuals whose gender does match the gender they were assigned at birth. *Non-binary* refers to individuals who do not identify as either men or women but whose gender identity falls outside the normative binary.

Survey and question design: Reflecting and respecting queer lives

The objects of analysis from which we draw our findings are the answers to a set of demographic questions regarding sexual and gender identities that we designed and deployed as part of a larger cross-sectional online survey. Although the later questions in the survey and the responses that we received to them are not the subject of this article, we provide a brief description of the overall survey topic here for context. This survey addressed the roles that mid-1990s and early 2000s internet technologies played in the development of sexual identity for users in the U.S. In constructing this survey, we were particularly interested in how respondents who are currently between the ages of approximately 25 and 35, and who had access to the internet as teens, now understand the relationship between their internet practices as youth and their sexual identities as adults. The survey opened with questions about demographics (including the respondent's age, racial and ethnic identities, gender identity, sexual orientation, geographic residence, and socioeconomic status). As we originally conceived of them, the demographic questions that represent our focus in this article were intended to assist with screening participants for eligibility and situating our findings within considerations of respondents' identities and personal histories. However, as we discuss in the sections below, we found the responses we received to these demographic questions themselves to be both surprising and meaningful, with implications that extended beyond the context of the original survey topic.

This survey was deployed online over the course of two months (February 2019 to April 2019). Our target population was adults born between 1980 and 1996 who grew up primarily in the U.S. with access to internet-enabled technologies in their pre-teens and teens; the demographic data we discuss here is drawn from individuals who fall within these parameters. The online survey was exploratory in nature and the size of our target population was undeterminable. Therefore, we necessarily relied on convenience sampling. Our data does not constitute a representative sample and is not a generalizable reflection of the overall population. To reach potential participants, we openly circulated the survey in thematically appropriate groups on several social media sites and other online spaces, such as Twitter, Facebook, Reddit, and email listservs. Because we were particularly interested in learning about queer individuals' experiences with technology, we also posted our survey in a number of digital spaces dedicated to LGBTQ issues, such as LGBTQ-focused Facebook groups and sexuality-related subreddits. An initial review of responses to the survey in March 2019

revealed that the overwhelming majority of respondents were white. In hopes of increasing the representation of people of color in our survey, we additionally publicized the survey on social media groups and other digital spaces focused on people of color, queer people of color, and communities of color with an interest in technology and digital culture.

In total, after discarding responses from ineligible respondents, we collected a 227 responses, a majority ($n = 178$) of which came from LGBTQ respondents (i.e., those who did not identify as solely heterosexual and cisgender). Of these LGBTQ respondents, 78% ($n = 139$) were white and 22% ($n = 39$) were people of color; we acknowledge this ratio as a limitation and strive to better address the experiences of queer people of color in our future research. For the purposes of this article, the findings that we present here are drawn only from the survey responses from LGBTQ individuals. In addition, as described, our focus is on a specific subset of demographic questions contained within the survey—namely those that relate to the respondents' gender and sexual identities. The text of these questions, as found within the survey, read as follows:

1. What is your gender? (Select all that apply.)

Response options: man, non-binary, woman, other (please specify).

2. Do you identify as transgender?

Response options: yes, no, decline to state.

3. What is your sexual orientation? (Select all that apply.)

Response options: asexual, bisexual|pansexual, gay, heterosexual|straight, lesbian, queer, other (please specify)

4. Prior to the age of 18, how did you identify your sexual orientation. (Select all that apply.)

Response options: asexual, bisexual|pansexual, gay, heterosexual|straight, lesbian, queer, other (please specify).

In addition to the responses that we received to these questions, the design of these questions themselves is itself notable, since they differ from more commonly encountered versions of demographic questions related to gender and sexual identity. Designing these questions in ways that allowed for the multiplicity and complexity of queer identity was an important part of the methodology—and, relatedly, the politics—of our research. In designing these questions, we drew from our backgrounds as queer (and queer of color) studies scholars, as well as our own experiences as LGBTQ people, with the goal of allowing respondents to describe their identities in ways that more respectfully and accurately reflected their lived experiences of sexuality and gender.

To design our questions, we both built from and also conscientiously deviated from existing standards for collecting demographic data. While we value the work of those who are striving to create more inclusive standards, we see many of widely used examples as insufficient or even problematic. For example, current standards for collecting demographic data on gender in health and clinical research suggest creating two questions about gender, one about gender identity and one about sex assigned at birth (The GenIUSS Group, 2014). This two-step approach uses discrete answer possibilities, often relying on sexed terms (male and female) for both questions. While this approach has been validated by the research community, we consider it to be inappropriate, because it requires trans and non-binary folks to report their sex assigned at birth. Additionally, because this standardized two-step demographic question only offers “female,” “male,” “trans female,” and “trans male” as possible answers, it reifies cisnormative structures that presumes gender to be binary and positions transgender people as “not really” male or female. Other standards for collecting demographic data on sexuality (Human Rights Campaign, 2016; Sexuality Minority Assessment Resource Team, 2009), although valuable, also fall short because they present respondents with discrete answer categories. While these guides encourage a diverse range of possible answers related to gender and sexual identity, they rarely allow respondents to choose multiple responses, which might more accurately allow for a range of sexual identity expressions. By addressing sexuality as a single question, they also inadvertently reproduce the assumption that an individual’s sexual orientation is static, single-axis, and unchanging.

The questions that we designed for our survey differ in multiple ways from standard demographic questions regarding sexuality and gender. These differences and their implications bear articulating, since the questions themselves may initially appear straightforward. Firstly, we broke gender identity out into two questions: one about gender identity in general and one about transgender identity; participants were also offered the option to “decline to state” whether they identified as transgender. This acknowledges that a person’s gender includes multiple, intersecting elements that cannot be captured as one data point. It is also more respectful of transgender people because it does not suggest that an individual’s gender identity is not fundamentally altered, characterized, or mitigated by their trans-ness. Secondly, our questions also separated sexual orientation into multiple elements by asking respondents to address how they understood their sexual orientations before the age of 18 *and* how they understand their sexual orientations today. This reflects a belief that sexual orientation, and the way that one

understands that orientation, may change over the course of an individual’s lifetime. Thirdly, in the questions regarding sexual orientation, respondents were given a larger and more inclusive list of options to choose from than is standard for collecting demographic data, including commonly overlooked or marginalized queer sexualities, such as asexuality and bisexuality/pansexuality. This demonstrates greater inclusivity in demographic options, with terminology drawn from queer communities themselves. Lastly, three out of four demographic questions related to gender and sexual identity allowed respondents to select multiple options, recognizing the complexity of LGBTQ identities.

The design of our demographic questions related to gender and sexuality represents an extension of recent calls by other researchers, such as Spiel et al. (2019), for questions of gender on research surveys to be optional, allow for multiple checkboxes, and represent diverse range of identities that do not reify the gender binary. Our work embraces similar, gender-affirming values. At the same time, we push these calls for more inclusive survey design in new directions by turning much-needed attention to sexual orientation as well as gender and an awareness of how gender and sexual identities can shift over time. Our design of these questions allowed LGBTQ respondents to express some of the complexities of their sexual and gender identities, and made visible the ways in which they identities challenge traditional notions of demographic data.

Complicating expectations about gender and sexual identities

The responses that we received to the demographic questions related to gender and sexuality identity in our survey suggest key findings in two interrelated areas—sexual identity and gender identity. Taken together, these findings offer valuable insights into queer lives. They also push us to reconsider and reimagine what it might mean to understand gender and sexual identity as and through demographic data. Our core findings are:

1. For an overwhelming majority of LGBTQ people who responded to our survey, sexual identity was not static or singular. Rather, it was something that shifted over time. Sexual identity had shifted over the respondents’ lifetimes and/or was best characterized by respondents themselves using multiple identity markers.
2. A notable percentage of LGBTQ respondents described their gender identities using multiple markers, even when given the option to write in

their own term for their gender identity. This suggests that gender identity, for many LGBTQ people, is characterized not simply by one term selected from an inclusive list but by multiple, overlapping elements of identity.

These findings and the evidence that supports them are presented in depth in the subsections that follow.

Sexual identity: Shifts over time and multiple markers

The first demographic question regarding sexual identity that appeared in our survey asked respondents to identify their sexual orientation using a select-all-that-apply format. From within our LGBTQ sample, the breakdown of descriptors that respondents chose to describe their sexual orientations was as follows: asexual (9.55% of respondents), bisexual/pansexual (52.25% of respondents), gay (22.47% of respondents), heterosexual/straight (3.93% of respondents), lesbian (11.24% of respondents), queer (44.38% of respondents), and “other” (5.06% of respondents) (Figure 1). Two observations are immediately notable within this dataset. First, it totals more than 100% (148.88%), indicating that a sizeable number of respondents described their sexual orientations using more than one term. Additionally, there is a strong representation of respondents who identify as “bisexual/pansexual” and/or “queer.” In the case of the term “queer,” this fits with our understanding of contemporary LGBTQ identities in the United States, since “queer” can be an umbrella term with which many non-heterosexual, non-cisgender individuals identify. However, given that bisexuality is widely considered a misrepresented and stigmatized sexual orientation even within LGBTQ communities (Garelick et al., 2017), it is striking that so many of our respondents identified as “bisexual/pansexual.” This result alone demonstrates the need for a greater diversity of options for describing sexual

identity in demographic data, which might otherwise preclude bisexual or pansexual individuals (grouped for the purpose of our survey because they are often used together within queer communities) and others from accurately describing their sexualities.

Following the question about respondents’ current sexual orientation, we asked respondents to select terms that described how they identified their sexual orientation before they were 18 years old. The breakdown of descriptors that our LGBTQ respondents chose is as follows: asexual (2.81% of respondents), bisexual/pansexual (34.83% of respondents), gay (15.73% of respondents), heterosexual/straight (54.49% of respondents), lesbian (7.87% of respondents), queer (4.49% of respondents), and “other” (3.93% of respondents) (Figure 2). There also a number of observations that merit note in this dataset, especially when contrasted with responses to the prior question regarding current sexual orientation. Taken together, the percentages to this second question total 124.15%. This is less than the total percentage for the previous question (148.88%), suggesting that respondents have not only shifted in their understandings of their sexual orientations over time but also that the number of descriptors through which they characterize their sexual orientations has increased. Perhaps most striking is the high number of respondents who, prior to the age of 18, identified as “straight” and the comparatively low number who identified as “queer.” It is likely that the discrepancy in identification with the term “queer” can be partially explained by cultural changes in the term’s use and adoption. However, the large number of LGBTQ respondents who would have described themselves as “straight” before the age of 18 demonstrates that many LGBTQ people change in the way that they understand their sexual identities, for example by moving from an understanding of themselves as heterosexual to an understanding of themselves as queer. (Note: This is distinct from how they *present* their sexual identities; these numbers do not reflect whether LGBTQ respondents were “out”

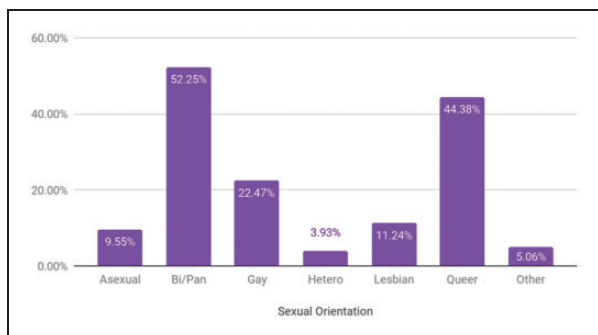


Figure 1. Sexual orientation for LGBTQ respondents (n = 178).

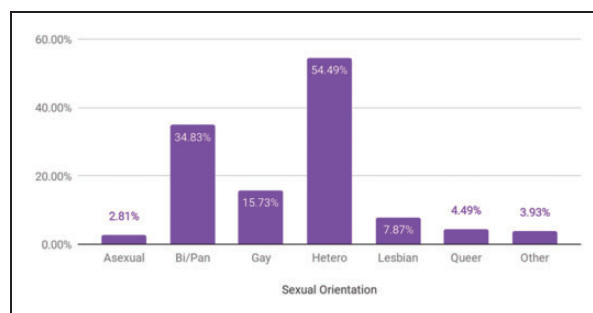


Figure 2. Sexual orientation prior to age 18 for LGBTQ respondents (n = 178).

before the age of 18, but rather whether they understood themselves as heterosexual or non-heterosexual at that time.) Considering these two bar graphs side-by-side, their differences are striking, given that they represent the identities of the same group of LGBTQ individuals.

In working with this data regarding sexual orientation, we were particularly interested in respondents who chose multiple terms to describe their sexual orientations or who described their sexual orientations at different times using different terms. Our initial review of these responses raised the following questions:

- When asked to select terms to describe their sexual orientations, how many respondents chose multiple terms? How does this number differ across the two demographic questions related to sexual orientation?
- How many respondents choose different answers to these two demographic questions—that is, how many selected different terms to describe their sexual orientation as they currently understand it versus their sexual orientation as they understood it prior to the age of 18?

Answering these questions allowed us to nuance our findings regarding how the sexual orientations of LGBTQ respondents shifted over time and how LGBTQ individuals understood their sexual orientations.

In total, we found that 83.15% of respondents ($n=150$) gave a different answer when asked about their current sexual orientation than they did when asked about their sexual orientation before the age of 18. This is a meaningful finding: an overwhelming majority of the LGBTQ respondents to our survey reported, through these questions, that their understanding of their sexual orientations has shifted over the course of their lifetime. Although we identified patterns within these findings, the complexities of this dataset also speak to the many, varied forms that LGBTQ identities can take. Of these 150 LGBTQ respondents who described their identities at these two time periods using different terms, a notable number (49%, $n=74$) shifted from identifying as solely heterosexual to identifying as non-heterosexual. Within these 150 respondents, 10.66% ($n=16$) shifted from identifying as heterosexual to bisexual, 6.66% ($n=10$) shifted from identifying as heterosexual to gay, and another 6.66% ($n=10$) shifted from identifying as heterosexual to bisexual and queer. In addition, several respondents shifted their understandings of their sexual orientations by including the term “queer” in their selection of markers that describe their current sexual orientations, while otherwise choosing descriptions that were consistent across both responses

(9.33%, $n=14$). It is also meaningful that many responses could not be sorted into groupings. Forty-five of 150 respondents (30%) had unique answer combinations, meaning that no other respondent answered both questions in the same way. Therefore, the number of unique combinations of shifting identity categories is nearly twice as frequent as the largest group that shared shifting identity categories. This speaks to both what is similar and what is different across LGBTQ experiences. While the majority of LGBTQ respondents had in common that their understandings of their sexual orientations shifted over time, comparatively few shared the same combination of past and present sexual orientations.

Additionally, we found that 41.01% of the 178 LGBTQ respondents ($n=73$) selected multiple identity markers to describe their current sexual orientations. By contrast, 18.54% of LGBTQ respondents ($n=33$) selected multiple identity markers to describe their understandings of their sexual orientations prior to the age of 18. In responses to the question regarding individuals’ current sexual orientations, the most common combination of markers was the pairing of the terms “queer” and/or “asexual” with other LGBTQ identities, including “gay,” “lesbian,” and “bisexual.” In response to the question regarding individuals’ sexual orientations prior to the age of 18, respondents more commonly paired “heterosexual” with other identity terms. This demonstrates that, for a sizeable portion of LGBTQ individuals who responded to our survey, sexual orientation could not be adequately described using just one identity marker, especially as they presently understand their identities. Instead, these respondents understood their sexual identities to include multiple dimensions—such as considerations not only of whom they are attracted to, as signified by terms like “gay” or “lesbian,” but also their levels of sexual interest and their affinities to broader communities, as signified by terms like “asexual” and “queer,” respectively. The fact that so many LGBTQ respondents chose to select multiple terms to describe their sexual orientations also confirmed our own beliefs and expectations, as queer people ourselves, that offering respondents the possibility of selecting multiple options would allow them to more fully and accurately express the complexities of their sexual identities.

Gender identity: Overlaps and intersections

To understand our respondents’ gender identities, we included two demographic questions related to gender. The first asked respondents to identify their gender. In response to this question, 46.06% of the 178 LGBTQ

respondents (n = 82) identified as women, 34.83% (n = 62) identified as men, 21.91% (n = 39) identified as non-binary, and 11.24% (n = 20) chose “other.” As in the case of our survey questions regarding sexual orientation, we allowed respondents to choose multiple markers to describe their genders. In total, 22 respondents (12.36%) selected more than one option to describe their identities (Figure 3). Among these, eight respondents (36.36%) identified as “woman” and “non-binary” and three (13.64%) identified as “man” and “non-binary.” The remaining 11 respondents (50%) selected one or more of these predetermined categories and used the “other” option to further clarify their gender identities. Respondents who selected “other” were prompted to specify in a text field. Examples of identity descriptors that respondents entered include: agender, intersex, gender non-conforming, anti-gender, transfeminine, and trans questioning. Although the percentage of respondents who chose multiple gender markers is lower than the percentage who choose multiple sexual orientation markers, this number is still notable, given that we provided a more limited set of gender options from which respondents could choose (without inputting their own additions). The choices of the 22 LGBTQ respondents who selected multiple gender identity descriptors suggest that gender is still complex. As these responses indicate, in some instances, gender identity cannot be accurately represented by any one box—even one that allows respondents to state their gender in their own words.

The second demographic question in our survey related to gender asked participants whether they identified as transgender. This was the only demographic question in our survey related to gender and sexuality for which we provided only discrete answer possibilities: yes, no, or decline to state. Of our LGBTQ respondents, approximately 19.66% (n = 35) identified

as transgender. While 75.28% (n = 134) stated that they did not identify as transgender, 5.06% (n = 9) declined to state. Comparing the responses to this question about transgender identity and the prior question about gender identity more broadly reveals notable insights. For example, of the 39 respondents who selected “non-binary” to describe their gender identities in the first question, 19 (48.72%) indicated in the second question that they *did* identify as transgender, whereas 16 (41.03%) indicated that they did *not* identify as transgender, and 4 (11.43%) declined to state. That is, among these 39 gender non-binary respondents, there was a nearly equal split between those who did or did not correlate their identity as non-binary to an identity as transgender. This indicates that, even among those who share specific LGBTQ identities, there may be no one, universally-accepted belief about how to characterize those identities. Historically, gender and sexual identities have shifted considerably over time. However, these findings demonstrate that, even within a shared historical and cultural moment, these identities themselves may not have an “objective” meaning; rather, their meanings are often contested and/or personal. Even surveys that include multiple questions about gender are unlikely to capture an unchanging “truth” about identity.

Reimagining data queerly

In this article, we have argued that dominant notions of demographic data, as those elements of data that seeks to accurately categorize and “capture” identity, do not sufficiently account for the complexities of LGBTQ lives. We have demonstrated this through an analysis of the responses we received from 178 LGBTQ individuals to an online survey that included four demographic questions regarding gender and sexual identity.

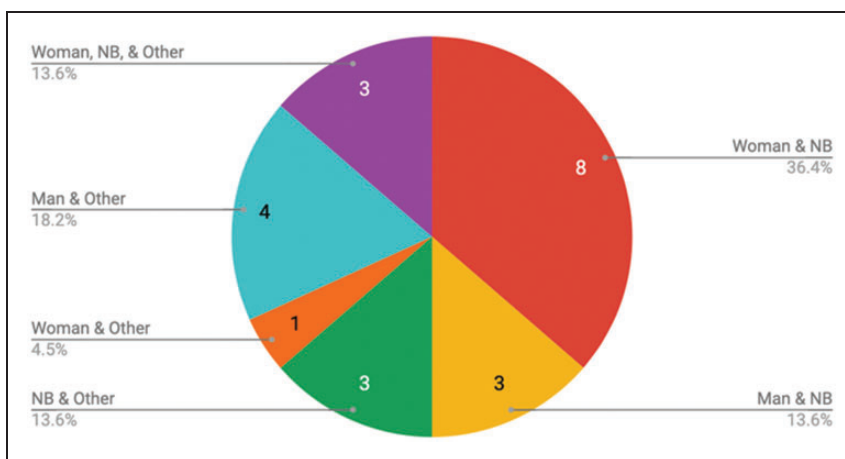


Figure 3. LGBTQ respondents who selected multiple gender categories (n = 22).

These responses indicated that, far from being singular or fixed, the gender and sexual identities of these LGBTQ individuals had often shifted over time or could best be described using multiple markers. This serves as a valuable, concrete demonstration of the ways in which dominant norms of conceptualizing data, which typically imagine an individual's identity to be static and discrete, are ill-suited to the complex, shifting lives of LGBTQ people, which by nature resist hegemonic identity categorization. As discussed above, many scholars from critical data studies, as well as related fields such as the digital humanities and feminist media studies, have voiced critiques about data, the quantification of identity, and the impact of big data and data-driven algorithms, especially as such issues relate to marginalized people. Our work here echoes these critiques while also expanding upon them through insights drawn from alternative approaches to the design of demographic data collection.

Although the questions that we developed for our survey represent a more inclusive, “queerer” approach to understanding gender and sexuality through data, these questions—and, by extension, the research presented here—also have limitations. As with most online surveys with an undeterminable target population, the question of sample size warrants consideration. While 178 responses represent a sufficient sample for an exploratory survey, a larger sample size would strengthen our findings. Additionally, our sample set over-represents respondents who identify solely as white (78.09%, $n = 139$). We also recognize that our own arguments regarding the importance of attending to how LGBTQ identities shift over time would be better supported by asking respondents about how they understood their sexual orientations at numerous points in their personal histories, rather than simply at two such points (the present and prior to the age of 18). By having only two questions that address temporal changes, we risk flattening the experiences of the LGBTQ folks and falling back on legal notions of the divide between an individual's experiences in childhood and adulthood. Ultimately, this is not a comprehensive study. Rather, it is a window onto a larger set of issues that demonstrates the friction between how demographic data is traditionally conceptualized and collected and the realities of queer lives. It is also an argument for how demographic data could be collected differently in the future to better account for the complexity, nuance, and fluidity of queer lives and experiences.

This work has implications that are both pragmatic and conceptual. It prompts new considerations in the design of more inclusive demographic questions. As discussed above, existing standards of survey design demand reconsideration because of factors such as their reliance on discrete answer possibilities, the

reification of cisnormative ideologies that imply that trans women and men are not real women and men, and the implication of identity as static and unchanging. Drawing from our work, we make the following recommendations to researchers developing demographic questions:

- Remove discreteness in answer possibilities. This is in line with the work of Spiel et al. (2019), who suggest that respondents should be allowed to check multiple boxes for gender identity. We recommend that, in addition to a variety of possible answers, questions should invite respondents to address multiple elements of gender and sexual identity.
- Apply a general approach to sexual and gender identity that understands that these identities may change over time. Recognize that categories of sexuality and gender are dynamic, temporal, and contextual. Allow respondents to account for the complexities of their identities and remember that all of the elements of their identities are valid; unless a respondent states otherwise, no one element of their identity, in the present or the past, is more “real” or “true.”
- To create questions about people who hold marginalized identity positions, such as LGBTQ people, collaborate with those people. People from within these groups have the best understanding of how to respectfully represent and describe their own experiences.

Together, these suggestions promote a more nuanced approach that embraces a queer understanding of gender and sexuality—one that is more inclusive, acknowledges complexity, and affirms the identities of respondents. These are important steps toward the creation of “data for queer lives.”

However, at the same time that we encourage researchers to design demographic questions in ways that are more socially just, we recognize that queer thinking itself challenges the very notion that categorizations of identity, however nuanced, can ever be complete or “correct.” As Emily Drabinski (2013) writes in “Queering the Catalog: Queer Theory and the Politics of Correction,” information classification systems such as those used Library of Congress can never be sufficient to represent the experiences of LGBTQ people. Whereas knowledge organization schemes “[take] these identities as stable and fixed,” frozen in time and universal, “queer theory sees these identities as shifting and contextual.” According to Drabinski (2013: 96), the problem with classification systems is this fixity itself—the very idea that queerness and its related concerns could sufficiently be “captured” as data—precisely because “queer perspectives [themselves] challenge the idea that classification...can

ever be corrected once and for all.” This call for a queer reconsideration of information, classification, and meaning-making is echoed in Kath Browne’s work on the development of a sexuality-related demographic question for 2011 United Kingdom’s census. Browne (2010: 235–236) writes:

To deconstruct methods and methodology that count and create state sanctioned subjectivities could be read as a ‘queer’ pursuit . . . A queer deconstruction of quantitative research tools could (and some would argue should) conclude in using queer tools to deconstruct normative categorization impulses.

Along with our own, this work demonstrates how a reconsideration of data from queer perspectives can itself have direct implications for the design of data collection and classification.

We conclude with a call for an approach to data that could itself be considered “queer.” What is queer data? Queer data is data that represents and serves queer lives. It is also data that, at its very foundation, is constructed around queer reconsiderations of identity, information, and meaning. Queer data stands as a challenge to the underlying, heteronormative and cisnormative logics that currently structure notions of demographics and data more broadly. At the same time, queer data puts data to use in order to complicate dominant cultural narratives about the structures of LGBTQ lives. For example, the research presented here productively brings into question the commonly-heard statement that LGBTQ people are simply “born this way.” Our findings suggest that LGBTQ people’s sexual and gender identities do, in fact, often shift over time. This does not make their identities any less “real” or indicate that they have chosen their gender or sexual identities. However, as researchers, it does force us to reconsider the relationship between these identities and what we call identity-based data as that which we commonly envision to be unchanging, objective, and “true.” Especially for LGBTQ people, the realities of gender and sexual identity do not fit within the tidy, immutable categories that are used to produce “good, clean” data. Indeed, queer data is messy, in part because queer lives are themselves often messy, as seen through the lens of normative society (Manalansan, 2014). Respectfully and meaningfully representing LGBTQ people through data, such as by using the methods we recommend here, may well complicate the research process. The data it generates will not be discrete and therefore may not fit neatly into standard tools for quantitative analysis. We recognize that this poses challenges for researchers. This is precisely the point. Queer data does not fit the norms of data analysis because those norms are not made to fit

the experiences of queer lives; queer data is not immediately legible to analytical tools because those tools have not been built to see queer people. Yet, while the work of queer scholarship must remain committed to serving the lives and communities of LGBTQ individuals, queer data also has implications beyond the datafication of queer people. As Ruppert et al. (2013) have argued, digital data itself is messy: shifting, heterogeneous, and non-coherent. Perhaps then we might understand data, stripped of its supposed objectivity, as already queer. Indeed, creating data for queer lives requires us to reimagine data itself.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Bonnie Ruberg  <https://orcid.org/0000-0003-2759-7233>

References

- Adler M (2017) *Cruising the Library: Perversities in the Organization of Knowledge*. New York: Fordham University Press.
- Baucom E (2018) An exploration into archival descriptions of LGBTQ materials. *The American Archivist* 81(1): 65–83.
- Benbouzid B (2019) To predict and to manage: Predictive policing in the United States. *Big Data & Society* 6(1): 1–13.
- Bivens R (2017) The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society* 19(6): 880–898.
- Boyd B and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge: MIT Press, pp.184.
- Browne K (2010) Queer quantification or queer(y)ing quantification: Creating lesbian, gay, bisexual, or heterosexual citizens through governmental social research. In: Browne K and Nash CJ (eds) *Queer Methods and Methodologies: Intersecting Queer Theories and Social Science Research*. New York: Routledge, pp.231–249.
- Browne K and Nash CJ (2010) Queer methods and methodologies: An introduction. In: Browne K and Nash CJ (eds) *Queer Methods and Methodologies: Intersecting Queer Theories and Social Science Research*. New York: Routledge, pp.1–23.
- Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? *Society and Space*.

- Available at: <https://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/> (accessed May 15, 2020).
- Drabinski E (2013) Queering the catalog: Queer theory and the politics of correction. *Library Quarterly: Information, Communication, Policy* 83(2): 94–111.
- Drucker J (2011) Humanities approaches to graphical display. *Digital Humanities Quarterly* 5: 1.
- Garelick AG, Filip-Crawford G, Varley A, et al. (2017) Beyond the binary: Exploring the role of ambiguity in biphobia and transphobia. *Journal of Bisexuality* 17(2): 172–189.
- Gieseck JJ (2018) Size matters to lesbians, too: Queer feminist interventions into the scale of big data. *The Professional Geographer* 70(1): 150–156.
- Gitelman L and Jackson V (2013) *'Raw Data' Is an Oxymoron*. Cambridge: MIT Press, pp. 2.
- Glantz A and Martinez E (2019) Can algorithms be racist? Trump's housing department says no. *Reveal*. Available at: <http://revealnews.org/article/can-algorithms-be-racist-trumps-housing-department-says-no/> (accessed May 15, 2020).
- Human Rights Campaign (2016) Collecting transgender inclusive data in workplace and other surveys. Available at: <https://www.hrc.org/resources/collecting-transgender-inclusive-gender-data-in-workplace-and-other-surveys>
- Huntemann N (2013) Women in video games: The case of hardware production and promotion. In: Huntemann N and Aslinger B (eds) *Gaming Globally: Production, Play, and Place*. New York: Palgrave Macmillan, pp.41–57.
- Johnson JM (2018) Markup bodies: Black (life) studies and slavery (death) studies at the digital crossroads. *Social Text* 36(4): 57–79.
- Kannabiran G, Bardzell J and Bardzell S (2011) How HCI talks about sexuality: Discursive strategies, blind spots, and opportunities for future research. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, US, May 2011 pp.695–704. New York: Association for Computing Machinery.
- Keilty P (2012) Sexual boundaries and subcultural discipline. *Knowledge Organization* 39(6): 417–431.
- Kitchin R and Lauriault TP (2014) Toward critical data studies: Charting and unpacking data assemblages and their work. In: Thatcher J, Shears A and Eckert J (eds) *Thinking Big Data in Geography: New Regimes, New Research*. Lincoln: University of Nebraska Press, pp.3–20.
- Lang N (2019) Inside the battle to get LGBTQ Americans counted in the census. *Daily Beast*. Available at: www.thedailybeast.com/inside-the-battle-to-get-lgbtq-americans-counted-in-the-census (accessed May 15, 2020).
- McCain CR (2017) Figuring the population explosion: Demography in the mid-twentieth century. *Feminist Media Histories* 3(2): 30–56.
- McGlotten S (2016) Black data. In: Johnson PE (ed.) *No Tea, No Shade: New Writings in Black Queer Studies*. Durham: Duke University Press, pp.262–286.
- Manalansan MF (2014) The 'stuff' of archives: Mess, migration, and queer lives. *Radical History Review* 120: 94–107.
- Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
- Posner M (2015) What's next: The radical, unrealized potential of digital humanities. Available at: <https://miriamposner.com/blog/whats-next-the-radical-unrealized-potential-of-digital-humanities/> (accessed May 15, 2020).
- Resnyansky L (2019) Conceptual frameworks for social and cultural big data analytics: Answering the epistemological challenge. *Big Data & Society* 6(1): 1–12.
- Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.
- Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6(1): 1–12
- Schwartz M and Constance C (2018) Remaking history: Lesbian feminist historical methods in the digital humanities. In: Losh L and Wernimont J (eds) *Bodies of Information: Intersectional Feminism and the Digital Humanities*. Minneapolis: University of Minnesota Press, pp.131–155.
- Sexual Minority Assessment Resource Team (2009) Best practices for asking questions about sexual orientation on surveys. Report, The Williams Institute of UCLA School of Law, Los Angeles, November.
- Spiel K, Haimson O and Lottridge D (2019) How to do better with gender on surveys: A guide for HCI researchers. *Interactions* 26(4): 62–65.
- Stryker S and Currah P (eds) (2015) Making transgender count [Special issue]. *Transgender Quarterly* 2(1): 1–12.
- Terranova T (2000) Free labor: Producing culture for the digital economy. *Social Text* 18(2): 33–58.
- Tongson K (2017) Queer. In: Ouelette L and Gray J (eds) *Keywords for Media Studies*. New York: New York University Press, pp.157–160.
- The GenIUSS Group (2014) Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys. Report, The Williams Institute of UCLA School of Law, Los Angeles, September.
- Walker L (2019) How to edit gender identity status on Facebook. *Lifewire*. Available at: www.lifewire.com/edit-gender-identity-status-on-facebook-2654421 (accessed May 15, 2020).
- Wernimont J (2018) *Numbered Lives: Life and Death in Quantum Media*. Cambridge: MIT Press.
- Wiegman R and Wilson EA (2015) Introduction: Antinormativity's queer conventions. *Differences* 26(1): 1–25.