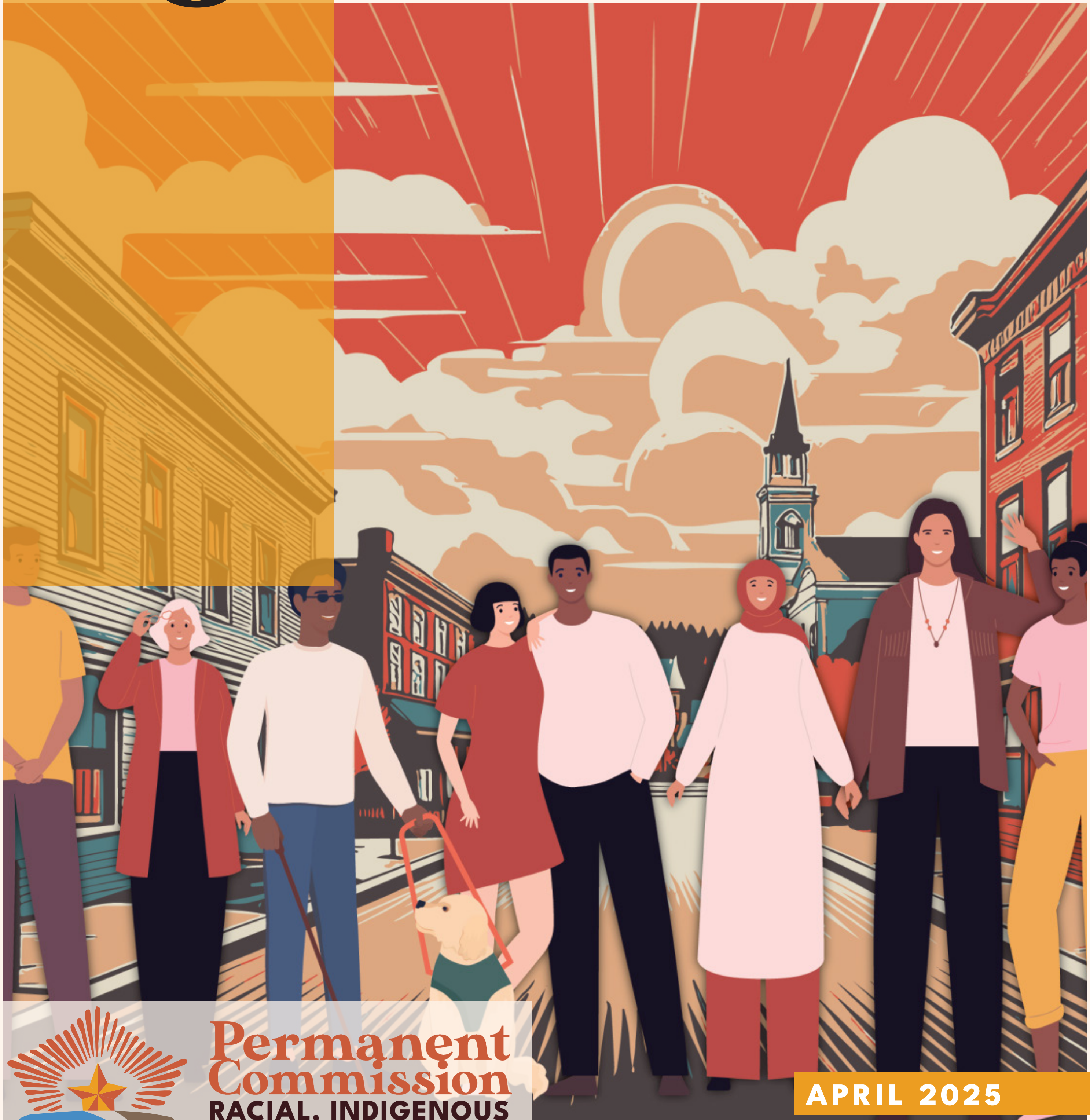
Data Justice



Permanent Commission
RACIAL, INDIGENOUS & TRIBAL POPULATIONS

APRIL 2025

Contents

Introduction

In 2021, the Maine State Legislature enacted and the Governor signed PL 2021, c. 717, “An Act to Promote Equity in Policy Making by Enhancing the State’s Ability to Collect, Analyze, and Apply Data”[*] and its companion legislation PL 2021, c. 21, “An Act to Require the Inclusion of Racial Impact Statements in the Legislative Process.” Through these combined efforts, the state of Maine aims to improve how demographic data is collected, analyzed, and governed so that it can be used in the policy process. These efforts also seek to improve the efficiency with which state agencies — who are often at the forefront of data collection — can share information.

This report is a supplement to Maine’s ongoing data governance effort outlined in PL 2021, c. 717, laying out the importance of robust and targeted demographic data collection. To the Permanent Commission, this means not just **more** data, but new types of data and analysis that uplift the unique needs and assets of Maine’s communities so that policy is based on the knowledge and experiences of real people. While we know that as state actors we have work to do to ensure that our data represents all Maine communities, we also recognize that quantitative data alone — even at fine levels of granularity — can obscure the unique needs of smaller communities or those that are not fully or accurately reflected in statistical averages.

In addition, while we strongly support efforts to improve data, this report acknowledges concerns around data privacy, ownership, and ethical use, especially for people of color and those from Indigenous communities who are often the target of data collection but rarely consulted about what data will be collected or how it will be used. As we advance an authentic desire to expand and improve Maine’s current demographic data to better serve all Maine people, history tells us we must also acknowledge these valid concerns, and explore models of data governance that address them. In this report, we lay out some perspectives for consideration and focus in particular on three important points:

1. **Racial data — even as raw numbers in a database — are not “neutral,” but instead draw upon particular ways of dividing human communities that have real implications for individual people and for our society as a whole.**

2. **The collection of more data does not inherently resolve issues of racial inequality, unless that data “is used, for good, and only for good.”1**

3. **As data becomes more central to state functions and decision-making processes (e.g., through the use of AI technology), risks associated with data collection are amplified in communities of color.**

The intention of this report is not to make specific policy recommendations for how to advance data governance in Maine. Instead, it is our hope to spur discussion among state actors and our private and non-profit partners about the need to think critically about what data is and how we use it. At the Permanent Commission, we have only begun to explore these ideas, and recognize that we do so in the midst of a rapidly changing technological and social context. We enthusiastically invite further conversation about these issues and their possible solutions at the broadest scales, and in particular, as they pertain to Maine and Maine’s ongoing data governance work.

The Case for a Conversation About Data Justice

In early 2020, the COVID-19 pandemic swept through every corner of the United States, filling hospitals, closing schools, and impacting the lives of millions of people. As it spread, it became abundantly clear that while few Mainers were safe from the disease, some were significantly less safe than others. By July of 2020, data showed that 1 out of every 725 Mainers had tested positive for COVID-19. For Black Mainers — that number was 1 in 27.[2]

This data is a powerful testament to the unequal burden that communities of color shouldered during the pandemic, but does not exist in isolation. Data from 2023 shows that despite making up less than 2% of Maine’s population, Black people accounted for nearly 50% of those who are unhoused.[3] In the five years prior to the COVID-19 pandemic, our public school records show that Black, Indigenous, and other students of color were suspended at nearly twice the rate that they were enrolled.[4] We have the nation’s largest racial disparity in home ownership[5] and are in the top ten out of 50 states for racial disparities in incarceration.[6] When we look at the patterns in the data, it quickly becomes clear that Maine is a state deeply impacted by racial injustice.

Despite this fact, Maine has a mostly haphazard approach to demographic data collection, which means that the disproportionate racial impact of

---

[*] *Amended by PL 2023, c.581.*

social problems is often obscured in the data we have. By not attaching program evaluations to racial data — for example in Maine's "Death with Dignity" initiative[*] or the "Maine Won't Wait" climate mitigation programs[†] — we cannot assess whether these programs are reaching vulnerable populations or meaningfully improving peoples' quality of life. **To really address disparities, we need better data — better stories, statistics, and syntheses — that help us to understand, track, and make sense of complexity.** With that in hand, we can take the appropriate steps, in the right places, to help develop and fund programs that lift all people from poverty, ensure access to sustainable housing and employment, and an equal chance to thrive and build a future here.

At the same time, communities of color who are often under-represented in databases that support the development of community programs may be over-represented in databases that monitor citizen behavior, threatening their autonomy and the sense of security that is the right of all people. Ongoing debate around the efficacy of Maine's Information and Analysis Center — an entity designed to collect, analyze, and share data with the federal government — highlights the applicability of these concerns in Maine.[‡] **These issues require us to reckon with the confusing reality that both the collection of data by the state, as well as the absence of it, has resulted in the marginalization of minority communities.** What then do we do to make evidence-based decisions that can improve the lives of Maine's racial and tribal populations?

For us at the Permanent Commission, our starting point has been thinking differently about the data we collect and how we use it. We have been drawn in particular toward the emerging concept of data justice – or "fairness in the way people are made visible, represented, and treated" through the process of data collection and analysis.[7] Data justice draws on foundational ideals of justice as fair distribution of benefits and harms in society,[8] but also concerns itself with the practices and mechanisms that allow us to get there: equal access to information, meaningful participation in decision-making about data, and the intentional transfer of authority over data from the state to communities.[9] In this report, we outline what we have learned that has shaped our approach and introduce key questions to think through together as we advance data governance in Maine.

---

[*] *Maine's "Death with Dignity" 2023 annual report for example only includes information about age, limiting our ability to assess who is aware of these programs and seeking out their use. See https://www.mainedeathwithdignity.org/wp-content/uploads/2024/04/2023_MaineStateAnnualReport.pdf.*

[†] *Maine's climate response aligns with the federal Justice40 initiative, requiring that 40% of federal funding for climate and clean energy goes to marginalized communities. Evaluating whether that is the case in Maine will require more comprehensive data collection at the state level. See https://www.maine.gov/future/sites/maine.gov.future/files/inline-files/Maine%20Climate%20Council_Equity%20Subcommittee%20Final%20Report_March%202023.pdf.*

[‡] *For more on stated concerns related to MIAC, see 2021 testimony by Senator Rick Bennet and cosponsor of the bill to end MIAC programs: https://www.mainelegislature.org/legis/bills/getTestimonyDoc.asp?id=152154.*

Part I

What Do We Mean By Data?

Data can refer to many things, but for our purposes, we define it as pieces of measurable or observable information that can be aggregated and disaggregated to understand patterns in our environment, human behavior, or social systems.* Data is often quantifiable (e.g., temperature, population density, wealth, etc.) or can be made quantifiable through surveys and statistical analysis (e.g., happiness index, social vulnerability, socio-economic status, etc.). In cases where quantification is not possible or desirable, data can also be qualitative, describing the qualities of a thing, experience, or way of being. This can include things like stories, pictures, or models that capture nuanced information about human communities. Data can be collected from people, places, or events directly (primary data) or from records kept about people, places, or events after the fact (secondary data).

While data tells us a lot about the world, we also know that what we learn from it is filtered by the lens we see it through. The questions that drive data collection, the metrics and instruments we invent to answer those questions, and the ways that we use that information are all framed by particular historical and cultural frameworks. In other words, we collect data as a way of understanding the world that we live in. We then use that data to build new policies, programs, and technologies that change the world, which in turn, changes the questions we ask and the ways that we answer them in the future. In the words of historian Bruno Latour, data is not neutral because science and the society it informs “cannot be separated, they depend on the same foundation.”[10]

IS DATA NEUTRAL?

We often hear the expression that “data is neutral.” In reality, however, all data — even numbers on a page — are encoded with social, historical, and cultural information that give shape to what that data means. Let’s say for example that you look at a thermometer and read today’s temperature as 61°F. If you are sitting outside on a cloudy day in Florida or Southern California, this may be winter coat weather. But to someone in northern Maine on a sunny day, this data could be interpreted as comfortably warm. While two people may read the data the same, **the meaning of the data depends on the reader’s subjectivity.** This is not the only point, however, at which meaning enters the data. The reading on the thermometer itself is dependent upon context: its location in the home, its proximity to sunlight or other heat sources, or indeed whether the instrument itself is accurate or is consistently too warm or cool. And if you were to tell someone else that it is 61° today, their ability to infer your intended meaning from your statement **may require a shared understanding of the specific historical, geographic, and institutional context that you are speaking from.** Your reading on the thermometer is in degrees Fahrenheit, a temperature index constructed by a 16th century European physicist and once used around the world. Today, however, this metric is exclusive to the United States as a result of decisions made by the nation’s founders and later, by politicians who chose not to adopt degrees Celsius as global instrumentation standards changed. **When we use data, it is not just biased in how we choose apply it, it is biased at every point of imagination, measurement, collection, and use.**

Making Meaning from Racial Data

All data is influenced by the historical and cultural frameworks that we bring into our work. However, for data about human characteristics, it is particularly important to understand how our measurements bring meaning to social categories like “race.”† Race, as a concept, is broadly defined as a construct used to describe categories of people based on the perception of shared physical traits, where no biological basis exists.[11, 12, 13] That does not mean that racialized experiences are not real, but rather that **the realities and impacts of race are regularly produced and reproduced through our efforts to define and categorize people along these lines.** In other words, what makes race “real” is that we measure it, and then use those measurements to inform policy and practice in the real world, with real outcomes for people alive today.

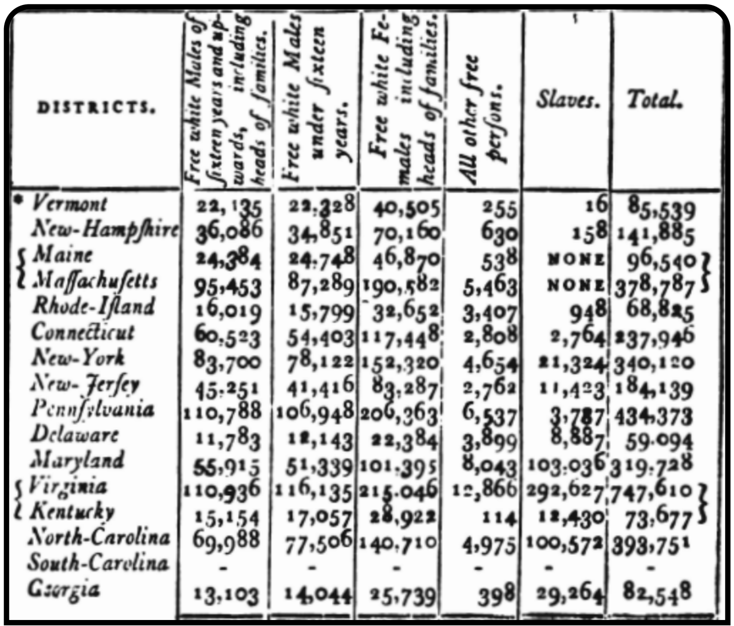
The formative power of racial data is visible in state efforts to document who lives where.[14] Within the US, the first large-scale racial data collection effort occurred in 1790 when the US Congress, empowered through Article I, Section 2 of the Constitution, enacted the first national census. This census was designed to count the number of individuals residing in each state to collect taxes, allocate federal resources, and apportion Congressional representatives.[15] To do this, the census needed to create categories into which it could group people for counting. A full count was given to “free persons, minus non-tax paying Indians,” while a three-fifths count was given to “all other persons”— namely enslaved Black people.[16] Not only did this choice carry moral implications for the treatment of people in captivity, it also restructured political power and financial capital in the United States through a system that discounted Black population centers, the impacts of which can still be mapped today in north/south economic disparities.[17]

---

* *Drawn in part from University of York (2024). Data: A Practical Guide. https://subjectguides.york.ac.uk/data.*

† *And gender, sexuality, ethnicity, and a number of other socially constructed demographic concepts.*

As early as 1820, the census began further segmenting populations along racialized lines. By 1850, in the lead up to emancipation, the census changed its classification system to remove a designation for those currently enslaved and began focusing categorization on the basis of skin color instead. By 1860, it included new designations for those of Indigenous and Chinese descent, populations likely to be encountered during westward territorial expansion. Every decade that followed, often in direct response to political issues facing the nation, more and more racial and ethnic categories were added, removed, aggregated and disaggregated to form and reform what we today think of as "racial groups."[18,19,20] These changes sometimes lump together individuals from across disparate geographies and cultures as if they are a homogenous group* (e.g., "Asian and Pacific Islander"), while other times forging new racial identities that had not previously existed.[21] Through these processes of construction and reconstruction, we see the challenge of conceptualizing race (even without political motives) as politically neutral; the fluidity in our understanding of race over time; and the powerful role of social statistics in shaping national discussion around belonging. This had — and continues to have — a profound impact, not only on the lived experiences of racialized people, but on the broader "imagined racial identity of the nation."[22]



DISTRICTS.	Free white Males of ſixteen years and upwards, including heads of families.	Free white Males under ſixteen years.	Free white Females, including heads of families.	All other free perſons.	Slaves.	Total.
• Vermont	22,135	22,328	40,505	255	16	85,539
New-Hampſhire	36,086	34,851	70,160	630	158	141,885
Maine	24,384	24,748	46,870	538	NONE	96,540
Maſſachuſetts	95,453	87,289	190,582	5,463	NONE	378,787
Rhode-Iſland	16,019	15,799	32,652	3,407	948	68,825
Connecticut	60,523	54,403	117,448	2,808	2,764	237,946
New-York	83,700	78,122	152,320	4,654	21,324	340,120
New-Jerſey	45,251	41,416	83,287	2,762	11,423	184,139
Pennſylvania	110,788	106,948	206,363	6,537	3,787	434,373
Delaware	11,783	12,143	22,384	3,899	8,887	59,094
Maryland	55,915	51,339	101,395	8,043	103,036	319,728
Virginia	110,936	116,135	215,046	12,866	292,627	747,610
Kentucky	15,154	17,057	28,922	114	12,430	73,677
North-Carolina	69,988	77,506	140,710	4,975	100,572	393,751
South-Carolina	-	-	-	-	-	-
Georgia	13,103	14,044	25,739	398	29,264	82,548

Figure 1. The first census, created to apportion political power across states, included counts of "freemen" and "slaves."[29]

Today, we take for granted that race is a meaningful way to segment and group human communities, and that those categories have meaning. We also see — often without any instrumentation at all — that there are differences and inequalities between these groups that fall to communities in patterned and predictable ways. In this context then, data on race — such as what is collected through the census — is vital to understanding inequality and its causes and consequences. So, will more data help us solve systemic racism?

> **"According to its constitutional mandate, the census does more than facilitate a body count; it also tells us whose body counts, and for how much."**
>
> **-Naomi Mezey, 2002 p. 1705**

Do We Need More Data About Race?

Looking at the multitude and complexities of racial disparities in Maine today — from inequalities in access to housing, transportation to wealth and employment[23] — it is entirely rational to wonder whether more data could help us to advance more just and equitable social outcomes. In some cases, this assumption may be true, and we can make changes to acquire that data at finer degrees of granularity. In some cases, this assumption may be false, either because the issue is already well-known or because it is impossible to gather data in a way that can answer our questions. And in other cases, the continuous pursuit of data, particularly data focusing on minority communities, can deepen the risks and harms being experienced by those communities. Based on the information outlined above, we are interested in exploring how to get more and better data, while also critically examining when demands for "more data" may exacerbate existing inequalities.

Our research suggests that there are multiple reasons why data related to racial disparities may be insufficient today. Historically, the state of Maine has not consistently collected demographic data related to race through its programs, which runs the risk of allowing disparities to go unnoticed. In other cases, however, data is collected, but its statistical applications are limited. Our ability to identify inequalities quantitatively depends in part on having sufficient numbers of people across groups to draw a comparison. The 2020 census suggested that 90.8% of residents identified as "white alone," while less than 10% identified as all other races.† Further disaggregating these data into smaller racial categories or smaller geographies creates challenges for statistical analysis, and vulnerable to misinterpretation.‡

** For a full documentation of changes to the census, see https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html.*

† Decennial census numbers also miss year over year variability, which is often high for minority populations in Maine. Over reliance on this data to understand Maine's demographic breakdown may obscure significant growth (and growing needs) in particular communities.

‡ When sample sizes differ dramatically, we run a higher risk of committing a type II error, or missing a relationship that really exists. For a plain language explanation of this phenomenon, see https://www.scribbr.com/statistics/statistical-power/.

That's true even if the data is accurate, but we also have reason to believe it often is not. The US Census Bureau, for example, acknowledges that its data collection efforts frequently undercount communities of color, both in Maine and nationally. At the national level, the Census Bureau estimates that Black Americans were undercounted by 3% in 2020, while Latinos and Tribal members living on reservations were undercounted by five and six percent, respectively.[24] Within Maine, we also have data to show that the census — which is intended to be a full count of all residents — in some counties includes fewer people of color than are currently enrolled in Maine's public schools. Why do these inconsistencies exist?

Maine is a state undergoing rapid demographic changes, both in terms of movement within the state and migration into it from outside. This alone, however, may not account for all of the inconsistencies we see in who is where. In the previous section, we documented the complicated history of government-sponsored data collection for the marginalization of people of color, where in some cases, populations were targeted or tracked for political purposes. For people of color and members of sovereign tribal nations, concerns related to exposure of confidential data are particularly resonant, not just for the risk of data breaches but equally in the intentional sharing of data within and between state entities in ways that may later be weaponized against them (see Eugenics on Page 6).[25, 26]

This has led to a documented and wide-reaching distrust of data collection efforts among communities of color (see Figure 2).[27] A study by the US Census Bureau on participant hesitation found that five persistent factors appear to influence refusal to participate in the US Census:

- **Concerns about data privacy and confidentiality.**
- **Fear of repercussions.**
- **Distrust in all levels of government.**
- **Feeling that it doesn't matter if you are counted.**
- **Belief that completing the census might not benefit you personally.**[28]

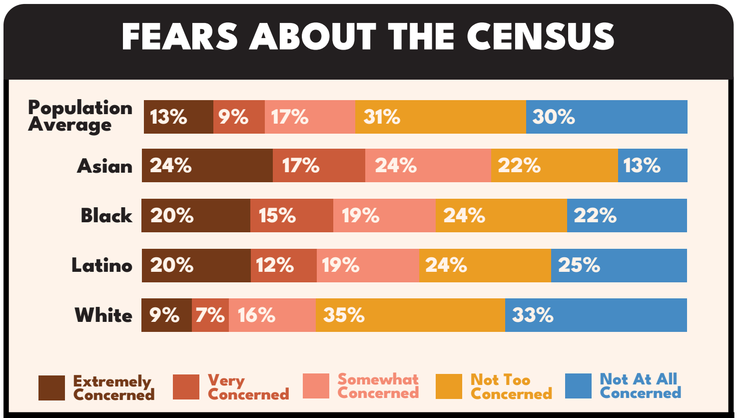


Figure 2. Concerns about data being used against communities of color lead to lower levels of participation in important data collection efforts. Adapted from US Census Bureau, CBAMS (2018).

In addition to fears about the disclosure of racial data, many individuals express concerns that their data will not get used, or not be used in ways favorable to their community. This is a well-founded concern. As stated by sociologist Dr. Anthony Ryan Hatch, "the operative assumption is that producing and analyzing [...] data is critical in the development of anti-racist action plans to close gaps in experiences and outcomes between racial and ethnic groups." The reality, however, is that data often isn't — or can't be — used to address these disparities, and results in "an endless search for more and more refined measurements of racism's harms, while the political and economic systems that comprise the fundamental causes of those harms are given a pass until all the data are counted."[29]

> "[The result is] an endless search for more and more refined measurements of racisms' harms, while the political and economic systems that [cause] those harms are given a pass until all the data are counted."

Big Data and Artificial Intelligence

This constant search for data is not limited to state efforts to better understand disparities. Concurrent with the rise of computer technology since the 1980s, there has been a dramatic shift in the social role that data plays in society, where a large number of our interactions and transactions have come to be defined as and by "data." Some experts refer to this as a process of datafication, or the mediation of social life by data-intensive systems and processes that track, collect, and analyze information about people as they move through their day.[30] While originally characteristic of the private sector, this practice has become increasingly prominent in the public sector as well.[31]

Today, facial recognition software, video monitoring of public spaces, forced engagement with online platforms and apps to access government services, mandatory data collection for health care access, the use of AI software in decision-making and management, and other forms of state activity intended to monitor and track individual citizens show how the state engages with data about its citizens at every level of operation.[32] Sometimes, these processes come with promises of greater public safety or more efficient government services. However, to those communities being monitored, increased collection of data may serve to intensify existing disparities.[33]

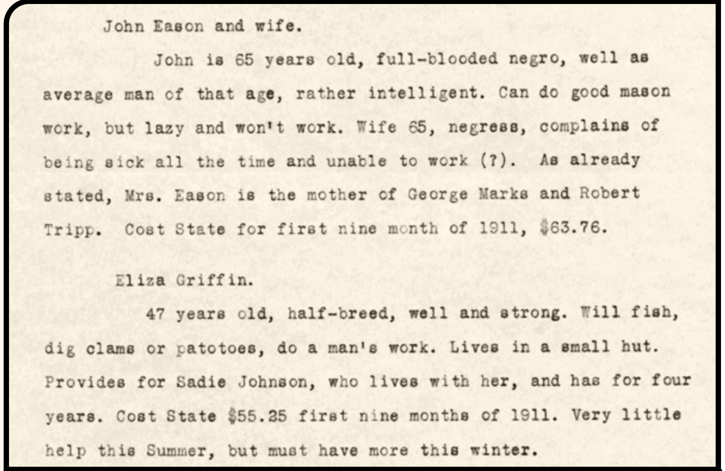
This is particularly resonant with the growing prevalence of AI in and adjacent to state decision-making processes. By mining already existing data, AI technology often serves to reinforce societal stereotypes and power structures in society.[*] [39, 40] While this concern is global, it is also local, as noted in the Governor's 2024 directive establishing the Maine Artificial Intelligence Task Force, which notes that AI applications in state government require significant scrutiny because:

> *"AI relies on collecting and interpreting large amounts of data from end users, which makes it susceptible to reinforcing biases, removing transparency from decision-making, and misusing private consumer information.*[†] *"*

Without the capacity among state data experts to critically examine AI outputs, we run significant risks of further entrenching inequalities. While a full analysis of AI is beyond the scope of this report, the Permanent Commission welcomes continued discussion and collaboration around this evolving topic.

EUGENICS AND ITS IMPACT IN MAINE

In 1883, English statistician Francis Galton advanced his now widely debunked pseudoscientific theory of eugenics, which articulated that the human population could be manipulated to improve hereditary quality, and in doing so, remove things like poverty, disability, mental illness, and criminal behavior from society.[34‡] Eugenics built on the logic of racial stratification that emerged alongside the rise in social statistics, and quickly gained prominence in the United States among certain segments of the scientific and business community. Its reach, however, was not limited to scientific study. Eugenicists in the United States helped to advance discriminatory legislation like the 1924 Johnson Reed Act that severely restricted immigration, and pushed state and federal entities to support policies that resulted in the violent separation of families, forced sterilization of minorities, and other human rights violations.[35]



John Eason and wife.

John is 65 years old, full-blooded negro, well as average man of that age, rather intelligent. Can do good mason work, but lazy and won't work. Wife 65, negress, complains of being sick all the time and unable to work (?). As already stated, Mrs. Eason is the mother of George Marks and Robert Tripp. Cost State for first nine month of 1911, $63.76.

Eliza Griffin.

47 years old, half-breed, well and strong. Will fish, dig clams or patotoes, do a man's work. Lives in a small hut. Provides for Sadie Johnson, who lives with her, and has for four years. Cost State $55.25 first nine months of 1911. Very little help this Summer, but must have more this winter.

Figure 3. The state of Maine used eugenics to justify detaining some of the residents of Malaga Island at the "Maine School for the Feeble Minded." Image from the Maine State Archives.

Maine was not immune to these pressures. In 1912, the state, acting under a directive of the Governor's Executive Committee, forcibly removed 40 residents from the mixed-race community of Malaga Island. While many individuals left to surrounding towns, the state also used eugenic arguments to place several individuals against their will in the "Maine School for the Feeble Minded."[36] While the state continues to reckon with this history today, it is not the only instance in which eugenic logic shaped state decision-making. The eugenics movement also informed Maine's support for the federal Indian boarding school movement, which placed Indigenous children in government-funded boarding schools where they were stripped of their heritage, culture, language, and familial ties.[37] Even as the use of residential schools phased out, Maine Child Welfare Services continued to remove Wabanaki children from their homes and place them in the care of non-native families, often away from their tribes and without clear justifications. By 1972, more than 30% of Indigenous children in Maine were in foster care across the state. This practice continued for decades and was found in 2015 to constitute cultural genocide.[38]

‡ ***For more information, see Stern, A. (2015). Eugenic Nation: Faults and Frontiers of Better Breeding in Modern America. University of California Press.***

---

[*] *Recent attention to bias encoded into AI software is one great example of this reproduction.*

[†] *In December of 2024, Governor Mills signed an executive order establishing the Maine Artificial Intelligence Task Force, which will examine opportunities and implication sof the use of US in Maine. See https://www.maine.gov/governor/mills/official_documents/executive-orders/2024-12-order-establishing-maine-artificial-intelligence-task.*

Part II

Moving Toward Data Justice

If we need data to address racial inequalities, but the collection of that data may put people at risk, how do we take effective action? Fortunately, these questions are not new and a range of individuals, organizations, research institutes, and advocacy groups today exist that can shed light on how best to balance these considerations. In this section, we will walk through existing and emerging standards, principles, and ethical considerations that we at the Permanent Commission weigh in our work.

Research Ethics and the Rights of Participants

Aware of the potential dangers and risks of unregulated data collection efforts, the 1974 National Research Act created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, tasked with identifying basic ethical principles for work with human subjects. Two years later, the Commission published the Belmont Report, which outlined three key ideas that should inform all research, especially research conducted by the state or with the support of state funds. These principles include:

- **Respect for Persons:** the belief that individuals should be treated as autonomous agents in making choices to participate or not participate in data collection efforts, and that people with diminished autonomy are entitled to special protections. This includes individuals with limited faculties and decision-making abilities (such as children and people with mental disabilities), but also those limited by circumstance to consent freely (such as people who are incarcerated, economically or educationally disadvantaged, pregnant, and other vulnerable populations, including members of racial and ethnic minority populations).
- **Beneficence:** the belief that those collecting data have an obligation to maximize benefits and minimize harms for those who participate, and that such principles likewise extend to the entire field of research as an institution.
- **Justice:** the belief that for research to be ethical, those who bear the burden of participation ought also to receive its benefits, and moreover, that no one should be unduly excluded from the opportunity to participate and thereby receive those benefits.

We see these ethical principles as a helpful starting point for ensuring that the data we collect does not replicate structures of harm. We find the principals within the Belmont Report to be foundational to our work, while also recognizing the limitations of an ethical approach to data collection that focuses solely on individual agents, particularly when collaborating with Indigenous research partners.* As such, we next turn to broader data justice frameworks that we use to help to guide our efforts to support and uplift racial, Indigenous, and tribal communities in Maine.

Racial Equity and Data Justice Frameworks

While the Belmont Report has been guiding ethical research for half a century, more recent efforts to inform ethical data collection have centered explicitly on questions of racial equity. These frameworks, often oriented toward the work of state, non-profit, and private organizations, articulate the need to consider how choices at every stage of the "data lifecycle" may influence outcomes for communities of color.[41] Resources often center efforts to reduce burdens on communities about whom data is being collected, improve the infrastructure necessary for data sharing among government agencies, and outline processes for ensuring the protection and rights of participants. Complimenting data equity frameworks are efforts centering principles of data sovereignty, which advocate for impacted communities – particularly sovereign tribal nations – to control their own data. Table 1 breaks down some of the prominent models currently used in the North American context.†

In our work at the Permanent Commission, we recognize the value of each of these frameworks in the unique perspectives they offer. We also recognize that data equity encompasses one branch of a larger project toward reconciling the state's need for data with the empowerment of communities. Here, it is the intersection between these frameworks and broader questions about data — how it is governed, how it is collected, and how it is operationalized — that we find to be of significant interest in our work. Beyond the resources provided in Table 1, the emerging field of data justice offers us additional tools for thinking through these questions. Data justice explores how the benefits and harms brought about by data collection efforts, much

---

* *For more information and guidance on working with tribes to collect data, we strongly recommend exploring the First Nations Principles of OCAP. https://fnigc.ca/ocap-training/.*

† *The models outlined here focus specifically on state data governance efforts, but a wealth of other resources on digital data justice can be found at: www.coalitioncommunitiescolor.org/.*

like disparities themselves, fall in predictable patterns across communities.[44] As such, this framework suggests that efforts to address racial disparities in society must engage consciously with data systems by recognizing the power of data to shape and reshape lived experiences; by tangibly transferring power and ownership of data over to impacted communities; and by changing how, when, and by whom data is collected, shared, and used. In the section that follows, we outline the core ideas at the center of this emerging discourse, and how we at the Commission have begun to think about their application and operationalization in our work.

FRAMEWORK	FRAMEWORK DESCRIPTION
Actionable Intelligence for Social Policy (AISP): Centering Racial Equity Throughout Data Integration[*]	AISP vision is a framework of ethical data use with a racial equity lens that supports power sharing and building across agencies and community members. AISP achieves this vision through normalizing, organizing, and operationalizing racial equity across the data lifecycle. It encourages power sharing agreements, social license for data access and use, and small-scale data sharing and integration. It discourages broad access to raw datasets, use of unregulated predictive algorithms, and linked data systems that automatically share data with those displaying patterns of institutionalized racism (e.g., law enforcement).[42] The AISP draws heavily from the GARE (Government Alliance on Race and Equity) racial equity toolkit and the broader work of the AISP's data sharing project at the University of Pennsylvania.
Transform Health: Health Data Governance Principles[†]	A significant amount of data collected by government bodies is intended to inform public health policies. The Health Data Governance Principles emerged from a series of regional and global workshops, and center a human rights approach to data governance through three primary ideas: • Data practices should not cause harm (protection). • Should create tangible benefits (promoting health value and outcomes). • Should prioritize equity at their core.[‡]
Global Indigenous Data Alliance: FAIR and CARE Principles[§]	The FAIR and CARE data principles represent two separate but intertwined approaches centered in Indigenous Data Governance that are powerful tools for ensuring tribal data sovereignty. FAIR principles dictate that data should be: • Findable. • Accessible. • Interoperable. • Reusable (data sharing). CARE principles outline that data should provide: • Collective benefit. • Authority of control to Indigenous communities. • Responsibility and Ethics around data control and use.[43] The FAIR and CARE principles draw heavily from the 2007 United Nations Declaration on the Rights of Indigenous Peoples.
Black Health Equity Working Group: Engagement, Governance, Access, and Protection (EGAP): A Data Governance Framework for Health Data Collected from Black Communities[¶]	The EGAP framework is a data sovereignty framework for Black communities that emerged in Ontario during the COVID-19 pandemic. The framework outlines the need for: • Genuine, accessible consultation with communities about data (Engagement). • Community decision-making structures to guide data collection and use (Governance). • Rights of communities to access their data and determine who else can access it (Access). • Safeguarding individual rights with regard to data (Protection). EGAP stresses that collection of race-based data should never be the goal in and of itself, but should always be used to dismantle structural racism and advance equity.

Table 1. Actionable frameworks for centering racial equity in data governance processes.

[*] *https://aisp.upenn.edu/wp-content/uploads/2022/07/AISP-Toolkit_5.27.20.pdf*

[†] *https://healthdataprinciples.org/*

[‡] *For more, see the National Academies of Science, Engineering, and Medicine's 2024 report, "Rethinking Race and Ethnicity in Biomedical Research."*

[§] *https://www.gida-global.org/care*

[¶] *https://www.communitycommons.org/entities/ac48dd7f-d24b-492f-a0d9-f5e82fc298b4*

Part III.

Operationalizing Data Justice

At the core of the data justice framework are six foundational principles: the redistribution of power over data; a dedicated focus on equity; improved access to data and its benefits for involved communities; community participation in data collection, analysis, and application; recognition of the reciprocal relationship between data collection and identity formation, and improved access to knowledge in its varied forms.[45]

Acknowledging the importance of these principles in working with racial data is an important starting point for advancing ethical and just practices. Operationalizing them throughout the data lifecycle is a more complicated task. While this is an ongoing and iterative learning process for us at the Permanent Commission, we have so far attempted to apply these principles to our work through four distinct practices:

DATA JUSTICE FRAMEWORKS

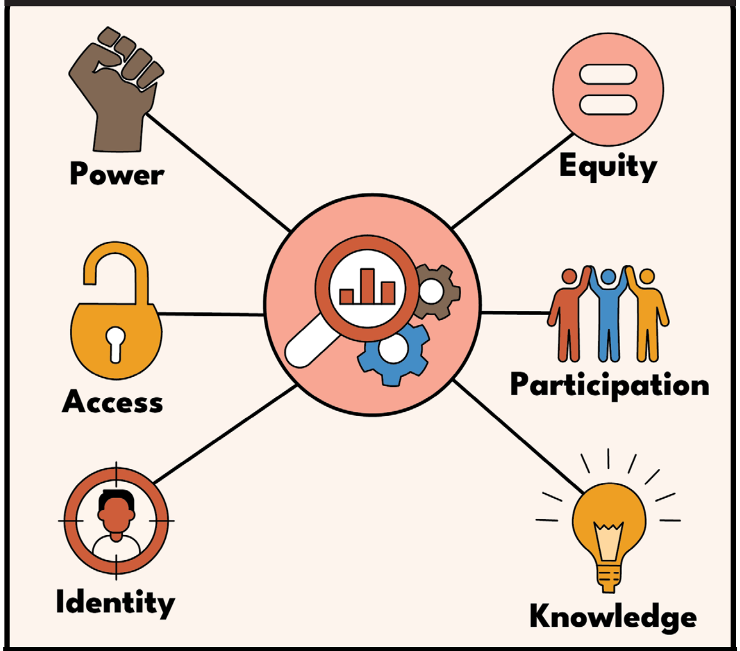


Figure 4. Data Justice frameworks draw together ideas about racial equity and broader themes of governance and autonomous control. Adapted from Leslie et al. 2022.

1 First, we recognize that systemic inequalities are the result of systems, institutions, policies, and practices — not the behaviors or attitudes of individuals. As such, **our research begins with a critical analysis of institutional structures themselves, and looks to impacted communities for sustainable solutions and alternatives rather than proof of harm.**

2 Second, when data collection is necessary, **we ensure that the data we collect is 1) used, 2) for good, 3) only for good, and 4) only with the informed consent of those involved.** In some cases, this means making a choice to collect racial data to fill in gaps in our knowledge. In other cases, this means making an intentional choice not to collect racial data if doing so threatens the autonomy or well-being of communities, or if we do not have a clear understanding of how we or others intend to use that information.

3 As we collect data, we acknowledge that choices made throughout the data lifecycle impact both the quality and meaning of findings, which can lead to actionable research or reinforce baseless stereotypes. As often as possible, we partner with impacted communities in data planning, collection, and analysis efforts, so that we can ground our findings and ensure that our research has impact. This also includes — where possible — **training and funding communities to collect data themselves so that they can share their findings with us, thereby shifting the balance of power over data back to the communities where it originates.**

4 Finally, while we recognize the value in consistent and translatable data related to race and other demographics, we seek — to the extent possible — to complement quantitative analysis with qualitative data. In particular, **our work looks to elevate methods that bring richness and living description into our understanding, especially through the use of stories, oral histories, personal narratives, and art.**

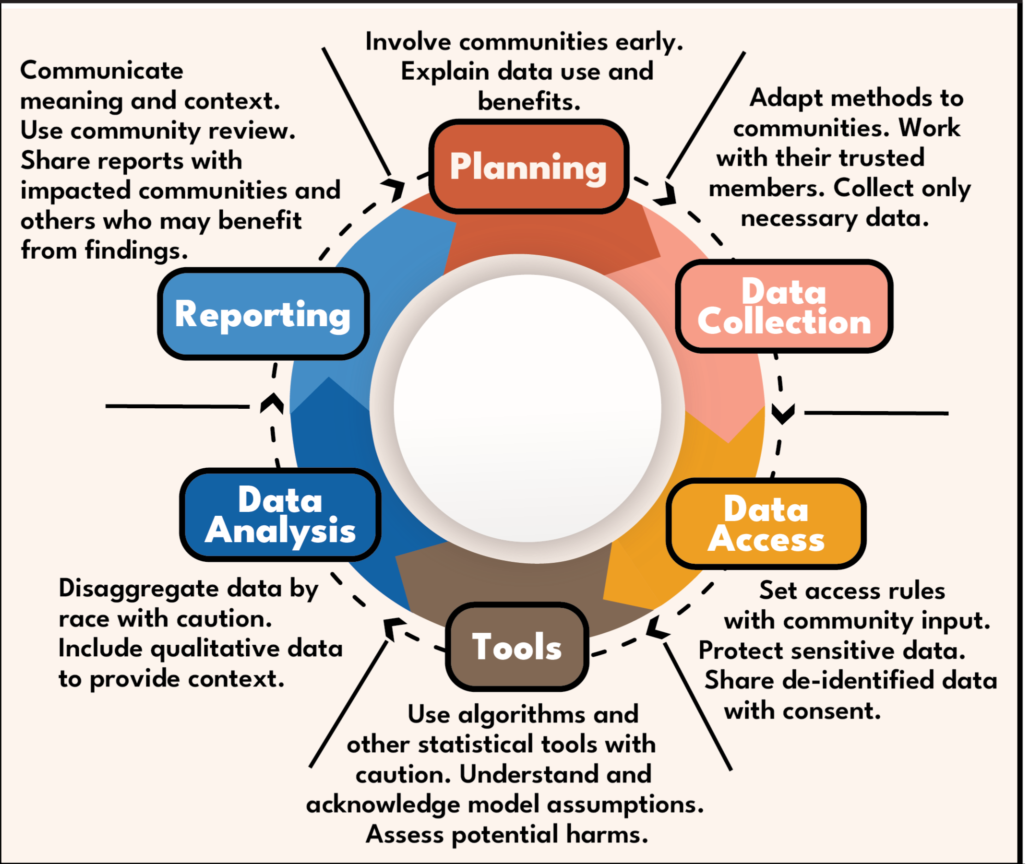
Questions for Further Consideration

Today, our motivation to collect more data about people in Maine stems from a genuine desire to fix problems deeply embedded in our social systems. These are noble and important pursuits. As we look to systemically address race- and class-based disparities across our state, a data justice framework opens critical lines of questioning so that we can orient our research efforts toward sustainable, positive change along the way. It also, notably, opens the door to more questions than answers. As we reach toward more just forms of engagement with data in our own work, we continue to explore the following questions, and invite conversation with others working in and alongside state government:

- Given the complexity of data systems and the technical skills required for collection and analysis, how do we more effectively balance power between technical experts and citizens in governing how data gets used?
- How do we ensure that citizen engagement is not just perfunctory, but provided with the resources to be sustainable and meaningful over the long-term?
- How do we create infrastructure for data sharing that doesn't jeopardize the safety and security of individuals and meaningfully honors the principles of prior and informed consent?
- What do we do in scenarios where federal requirements or even our own legal mandates require that we undertake data collection efforts that may negatively impact communities we serve?
- Under what conditions must the state gather and store data on its residents? Under what conditions is the state able to support communities to collect, steward, analyze, and present their own research findings?
- How do we elevate principles of data justice in the context of AI, where existing bias is replicated and often presented uncritically?
- How do we advance principles and practices of data sovereignty in communities lacking centralized authority or organization (e.g., among Maine's Black residents)? Who can we imagine empowering as data stewards in cases where centralized authority does not exist?
- How do we balance a genuine need for qualitative data (e.g., stories from marginalized people) with the risks of exposure of personal information and retraumatization?

We do not have answers to these questions today, but we believe that they lie in authentic dialogue within the state and with the communities most impacted under our current systems. At the Permanent Commission, we are committed to building a better Maine for all Mainers, and see data justice as a crucial part of that ongoing process. While we cannot undo the harm done in the past, we can — and must — design new processes for the future that center equity and justice at their core so that all Mainers can realize the dream of "life — the way it should be."

JUSTICE IN THE DATA LIFECYCLE



Involve communities early. Explain data use and benefits.

Adapt methods to communities. Work with their trusted members. Collect only necessary data.

Set access rules with community input. Protect sensitive data. Share de-identified data with consent.

Use algorithms and other statistical tools with caution. Understand and acknowledge model assumptions. Assess potential harms.

Disaggregate data by race with caution. Include qualitative data to provide context.

Communicate meaning and context. Use community review. Share reports with impacted communities and others who may benefit from findings.

Figure 5. The Permanent Commission seeks to operationalize principles of data justice at every step through the data lifecycle.

Glossary

Data: pieces of measurable or observable information that can be aggregated to understand patterns and meanings in our environment, human behavior, or social systems.

Datafication: the process through which societies and the people that comprise them have become increasingly dehumanized as their thoughts, actions, ideas, stories, and bodies become conceptualized as data points.

Data Justice: fairness in the way people are made visible, represented, and treated through the process of data collection and analysis.

Data Sovereignty: the right of communities, especially (though not exclusively) sovereign tribal nations, to govern and control their own data.

Race: the social construction and categorization of people based on perceived shared physical traits that result in the maintenance of a sociopolitical hierarchy.

Racial Equity: a process of eliminating racial disparities and measurably improving outcomes for all people by changing policies, practices, systems, and structures that marginalize people of color.

Social Statistics: a branch of statistics that is oriented toward the study of human beings and their general composition in society.

State: for our purposes, the personification of a government or organized political community, regardless of its level of operation (local, state, or federal).

Acknowledgements

The Permanent Commission on the Status of Racial, Indigenous, and Tribal Populations would like to express a special thank you to Hunter Cropsey and Dr. Anthony Ryan Hatch for their valuable insights and contributions throughout the process of drafting and revising this report.

References

1 Hatch, A. (2024). Personal Communication.

2 Cumberland District Public Health Council (2020). Cumberland District Public Health Council Statement on racial disparities and COVID-19. Available at: https://rb.gy/i1qg4g.

3 MaineHousing. (2023). 2023 Point In Time Count. https://rb.gy/x0frf.

4 Office For Civil Rights, US Department Of Education. (2024). Maine. civilrightsdata.ed.gov/profile/us/me?surveyyear=2020.

5 America’s Health Rankings. (2022). Homeownership Racial Disparity. https://rb.gy/cthszx.

6 Prison Policy Institute. (2023). Maine Profile. Accessed October 10, 2023. https://www.prisonpolicy.org/profiles/me.html.

7 Taylor, L. (2017). What Is Data Justice? The Case For Connecting Digital Rights And Freedoms Globally. *Big Data & Society*, 4(2).

8 Rawls, J. (1971). *A Theory Of Justice*. Harvard University Press, Cambridge, MA.

9 Dencik, L., Hintz, A., Redden, J., & Trere, E. (2022). *Data Justice*. Sage Publications.

10 Latour, B. (1998). Essays On Science And Society: From The World Of Science To The World Of Research? *Science* (280)5361, 208-209.

11 American Psychological Association (2023). *APA Dictionary Of Psychology: Race*. https://rb.gy/o92nzs.

12 American Sociological Association. (2003). Statement Of The American Sociological Association On The Importance Of Collecting Data And Doing Social Scientific Research On Race. https://www.asanet.org/wp-content/uploads/race_statement.pdf.

13 Smedley, A., & Smedley, B. (2005). Race As Biology Is Fiction, Racism As A Social Problem Is Real: Anthropological And Historical Perspectives On The Social Construction Of Race. *American Psychologist*, 60(1), 16.

14 Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification And Its Consequences.* Mit Press, Cambridge, Ma.

15 US Census Bureau (2021). Census In The Constitution. Census.gov. Retrieved December 28, 2023, From https://rb.gy/akaw0d.

16 Mezey, N. (2003). Erasure And Recognition: The Census Race And The National Imagination. *Northwestern University Law Review*, 97, P. 1705.

17 Chetty, R., Jackson, M., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R., ... & Wernerfelt, N. (2022). Social Capital I: Measurement And Associations With Economic Mobility. *Nature*, 608(7921), 108-121.

18 Omi, M., & Winant, H. (2014). *Racial Formation In The United States*. Routledge, Oxfordshire.

19 Lee, S. M. (1993). Racial Classifications In The US Census: 1890–1990. *Ethnic And Racial Studies,* 16(1), 75–94.

20 Mezey, N. (2003). Erasure And Recognition: The Census, Race, And The National Imagination.

21 Lee, S. M. (1993). Racial Classifications In The US Census: 1890–1990.

22 Mezey, N. (2003). Erasure And Recognition: The Census Race And The National Imagination. P. 1701.

23 Permanent Commission On The Status Of Racial, Indigenous, And Tribal Populations. (2024). *State Of Racial Disparities* Report. https://www.pcritp.me/sites/pcritp.me/files/inline-files/state_of_disparities.pdf.

24 US Census Bureau (2022). Census Bureau Releases Estimates Of Undercount And Overcount In The 2020 Census. https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html.

25 James, R., Tsosie, R., Sahota, P., Parker, M., Dillard, D., Sylvester, I., & Burke, W. (2014). Exploring Pathways To Trust: A Tribal Perspective On Data Sharing. *Genetics In Medicine,* 16(11), 820-826.

26 Actionable Intelligence For Social Policy. (2022). *A Toolkit For Centering Racial Equity Throughout Data Integration*. https://aisp.upenn.edu/centering-equity/.

27 Scharff, D., Mathews, K., Jackson, P., Hoffsuemmer, J., Martin, E., & Edwards, D. (2010). More Than Tuskegee: Understanding Mistrust About Research Participation. *Journal Of Health Care For The Poor And Underserved*, 21(3), 879-897.

28 Stanford, J. (2020). Why Some Americans Don’t Trust The Census. *The Conversation*. https://theconversation.com/why-some-americans-dont-trust-the-census-130109.

29 Hatch, A.R. (2022). The Data Will Not Save Us: Afropessimism And Racial Antimatter In The COVID-19 Pandemic. *Big Data & Society,* 9(1), 2.

30 Dencik, L., Jansen, F., & Metcalfe, P. (2018). *A Conceptual Framework For Approaching Social Justice In An Age Of Datafication.* Data Justice Project, Cardiff University, Cardiff.

31 Hintz, A., Dencik, L., Redden, J., Trere, E., Brand, J., & Warne, H. (2022). *Civic Participation In The Datafied Society: Towards Democratic Auditing?* Data Justice Lab, Cardiff University, Cardiff.

32 Crooks, R. (2022). Seeking Liberation: Surveillance, Datafication, And Race. *Surveillance & Society*, 20(4).

33 Ibid.

34 Galton, F. (1904). Eugenics: Its Definition, Scope, And Aims. *American Journal Of Sociology*, 10(1), 1-25.

35 Diamond, A. (2020). The 1924 Law That Slammed The Door On Immigrants And The Politicians Who Pushed It Back Open. *Smithsonian Magazine.* https://www.smithsonianmag.com/history/1924-law-slammed-door-immigrants-and-politicians-who-pushed-it-back-open-180974910/.

36 Maine Coast Heritage Trust. (2009). Malaga Island: An Overview Of Its Cultural And Natural History. https://www.mcht.org/story/malaga-island-history/.

37 Newland, B. (2024). *Federal Indian Boarding School Initiative Investigative Report Vol. II.* US Department Of The Interior. https://www.bia.gov/sites/default/files/media_document/doi_federal_indian_boarding_school_initiative_investigative_report_vii_final_508_compliant.pdf.

38 Maine Wabanaki-State Child Welfare Truth & Reconciliation Commission. (2015, June). *Beyond The Mandate: Continuing The Conversation.* https://d3n8a8pro7vhmx.cloudfront.net/mainewabanakireach/pages/17/attachments/original/1468974047/trc-report-expanded_july2015.pdf?1468974047.

39 Vázquez, A. F. D. C., & Garrido-Merchán, E. C. (2024). A Taxonomy Of The Biases Of The Images Created By Generative Artificial Intelligence. *arXiv* (2407.01556). https://arxiv.org/abs/2407.01556.

40 Ananya. (2024). AI Image Generators Often Give Racist And Sexist Results: Can They Be Fixed? *Nature*. https://www.nature.com/articles/d41586-024-00674-9.

41 Dencik, L., Hintz, A., Redden, J., & Trere, E. (2022). Data Justice. *Sage Publications*, Los Angeles, CA.

42 Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Berkowitz, E., Et Al. (2020). A Toolkit For Centering Racial Equity Throughout Data Integration. Actionable Intelligence For Social Policy, University Of Pennsylvania. https://aisp.upenn.edu/.

43 Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S. & Hudson, M. (2020). The Care Principles For Indigenous Data Governance. *Data Science Journal*, 19, 43.

44 Crooks, R. (2022). Seeking Liberation: Surveillance, Datafication, And Race. *Surveillance & Society*, 20(4).

45 Leslie, D., Katell, M., Aitken, M., Singh, J., Briggs, M., Powell, R. & Jayadeva, S. (2022). *Data Justice In Practice: A Guide For Policymakers.* https://ssrn.com/abstract=4080050.