



Delft University of Technology

Document Version

Final published version

Licence

CC BY

Citation (APA)

van Geene, L., Goncalves, J., Robinson, C., & Verma, T. (2026). Reclaiming data science for just geographies: A critical approach. *Big Data and Society*, 13(2). <https://doi.org/10.1177/20539517261426463>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Reclaiming data science for just geographies: A critical approach

Big Data & Society
 April–June: 1–17
 © The Author(s) 2026
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517261426463
journals.sagepub.com/home/bds



Laura van Geene¹ , Juliana Goncalves² , Caitlin Robinson³ and Trivik Verma⁴ 

Abstract

As data science increasingly shapes educational programmes, research agendas, and societal narratives, its practices have come under scrutiny for reinforcing historical inequalities, perpetuating biases, and neglecting critical engagement with issues of power, capital, and representation. Drawing upon critical social science theories including decoloniality, intersectionality, radical transdisciplinarity, and reflexivity, this paper narratively explores the limitations of conventional data science methods and pedagogy, advocating instead for a critical paradigm shift aimed at reclaiming data science for just geographies. We highlight the necessity for an approach that recognises data science as inherently subjective, deeply embedded in social and political contexts, and fundamentally shaped by historical legacies of colonialism and exclusion. By situating our experiences within universities in Western Europe, we illustrate how education and research can inadvertently perpetuate harmful structures when failing to critically engage with the positionalities and power dynamics inherent to data practices. Responding to these broader societal challenges, we propose a practical, iterative framework for critical data science that has emerged from our teaching methods and research experiences. This framework invites researchers and educators to continually reflect upon inclusivity, inequality, participation, power, and positionality throughout each stage of the data science process. Ultimately, our aim is to empower a generation of data scientists capable of interrogating dominant narratives, embracing diverse perspectives, and collaboratively working towards more equitable, just, and caring futures for all.

Keywords

Critical data science, intersectionality, decoloniality, radical transdisciplinarity, reflexivity, education

Introduction

Moving through a university, it is commonplace to hear slogans like data science for social good, artificial intelligence (AI) for healthcare, data and digitalisation, digital futures, or some version of these. Coupled with advances in computational infrastructures, data science and AI have generated excitement within higher education. Universities are going through a rapid, unplanned, but opportunistic transformation, incorporating data science, digital technologies, and open practices into research and teaching.

Programmes dedicated to data science are meant to respond to societal challenges – climate change, inequalities and ecological degradation among others. This is typically structured within discipline-specific data science programmes including health, social, geographic and engineering data science specialities. But these challenges are not disciplinary. How inequalities are reproduced across generations (Nijman and Wei, 2020), how climate breakdown is deeply intertwined with economic growth (Brand-Correa et al., 2022), or how energy transitions

reproduce extractive colonial legacies (Hamouchene, 2023): these are deeply interconnected challenges that require not purely technological solutions but deep political and social transformations (Boyd and Juhola, 2015; Goncalves et al., 2025a; Hickel and Sullivan, 2023; O'Rourke and Lollo, 2015). This is often overlooked in teaching of data science principles. For example, new technologies can be effective in supporting Alzheimer's patients with memory recollection. But AI

¹Faculty of Technology, Policy, and Management, Delft University of Technology, Delft, The Netherlands

²Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands

³School of Geographical Sciences, University of Bristol, Bristol, UK

⁴Geography and Environment, Loughborough University, Loughborough, UK

Corresponding author:

Trivik Verma, Geography and Environment, Loughborough University, Loughborough, UK. Email: t.verma@lboro.ac.uk



and healthcare study programmes rarely discuss who has access to healthcare or why certain communities struggle to find care. Technologies used in the classroom and for research also suffer from gendered and intersectional inequalities that are deeply resonant in the profession of AI and data science (Young et al., 2023).

Within quantitative disciplines in particular, a classroom teaches us to draw borders around our expertise. Ultimately, without building bridges between disciplines (Kitchin, 2014), without collective deliberation with society (Kallis, 2019), without engaging in social and democratic processes to build a future (Pearce, 2024), conventional data science practices risk perpetuating social challenges (Cugurullo, 2024). Universities also support and invest in open education practices. The widespread availability and success of Open Educational Resources like Massive Open Online Courses¹ highlights the global reach and impact of western perspectives, methodologies, and practices. However, despite being free to own, share and modify, these materials reinforce existing systems and structures of power. Materials overwhelmingly exemplify data sources from the United States or European countries where plenty of data supposedly exists compared to so-called Global South countries (Pasquinelli, 2020). In research, international programmes based on open-source data science pipelines that promote capacity building in the Global South, typically ignore local knowledge systems and trust popular narratives of development and growth. This, coupled with a disproportionately high number of citations for White senior male scientists (Liu et al., 2023) whose thinking may be widely adopted as open-source models (Cottineau et al., 2024), hinders engagement with and from diverse identities. Through such dominant forms of knowledge, pedagogy and research remain closed to genuine question, critique, and enrichment, with no space for alternative social realities, lived experiences, methodologies, and frameworks (Franklin et al., 2023).

In teaching, researching and using such methodologies at scale, we engage in a complex dynamic that has the potential to do societal harm. The political economy of education and labour, shaped and collectively delivered by universities globally, has laid out a production line of future researchers and practitioners who often lack opportunities for critical thinking. Increasingly, higher education is being commodified and its value defined in terms of labour market contributions (Wilkinson and Wilkinson, 2023). Moreover, the data science and AI sector was built on exploited human labour (Hao and Hernandez, n.d.) and extraction of services and resources from the Global South (Hao and Heidi, n.d.). Big Tech uses these methodologies to control narratives and infrastructure to make unprecedented and disproportionate profit (Sadowski, 2021; Sadowski and Bendor, 2019). For example, many organisations employ mapwashing techniques, where disingenuous uses of maps (and the data powering them) undermine participatory planning processes (Mattern, 2020). In turn, a cycle emerges, potentially driving universities to prioritise skills that sideline critical

thinking. Non-consensual data extraction and unequal forms of participation in decision-making, labour, and society are normalised, further perpetuating damages to vulnerable communities, now and in the future.

There is a need to recognise the role that data science plays in society and its influence on equity, justice, sustainability, participation and deliberation. Academic research and teaching can reproduce harm in subtle yet systemic ways. For instance, widely used ‘open’ datasets, such as crime statistics or social-media traces, often encode racialised, gendered, and class biases that are rarely interrogated in relation to the racialised or classed histories of their production (Brayne, 2017; Noble, 2016). Some established algorithms taught in the classroom can inadvertently shift attention from structural inequalities to individual behaviour, thereby normalising market-driven and exclusionary urban processes (Vybornova and Verma, 2025). Similarly, students analysing official mobility or housing data frequently engage with datasets that omit informal settlements, migrant communities, or gendered care work (De Madariaga, 2013), rendering them statistically invisible. Such omissions, though unintended, perpetuate epistemic harms and reinforce rather than challenge the social inequalities that contemporary data-science education and research often claim to address.

Although academic, governmental, and commercial domains of data science may seem distinct, they are in fact deeply interdependent. Governments routinely partner with Big Tech firms to manage data infrastructures and deliver digital public services, while universities align their teaching and research agendas with these same policy and market priorities. Through funding mechanisms, labour-market pressures, and collaborations branded as ‘transdisciplinary’, higher-education institutions act as key sites where data-science knowledge is produced and recirculated back into state and corporate systems, reinforcing the very logics they seek to study (Birhane and Guest, 2020; Couldry and Mejias, 2019). Focusing on education and research, therefore, allows us to interrogate how universities participate in, but may also contest, these broader assemblages of power shaping contemporary data practices.

Students and researchers interested in data science should have the opportunity to develop a critical perspective, engage with the processes by which existing structural inequalities and injustices manifest, and learn how to mitigate and address them. While the necessity of a critical perspective in data science is gathering attention (Goncalves et al., 2025b; Jamieson et al., 2023), the field itself is undergoing a period of epistemic uncertainty. Positioned at the intersection of statistics, computer science, and science and technology studies, data science lacks a stable identity, oscillating between being a technical discipline, an epistemic framework, and a socio-political practice. It continues to grapple with tensions between methodological rigour, computational efficiency, and ethical reflexivity (Boyd, 2021b). Under this identity crisis, universities increasingly include

ethics modules and operate research ethics boards to safeguard data privacy, consent, and storage. But these mechanisms tend to emphasise procedural compliance rather than critical reflection on the epistemic and methodological assumptions that shape data science itself. Students are rarely encouraged to question how their analytical choices, data transformations, or visualisations may reproduce bias or inequality. This omission is especially evident in shorter, technically focused data science courses that prioritise coding proficiency and methodological execution over reflexivity. Though critical or reflexive thinking may feature in social science programmes, it remains non-existent within most technical data science curricula. Consequently, ethics in data science education often stops at data collection, overlooking the broader social, political, and interpretive consequences of analytical practice.

We need a new, deeply normative paradigm (Kitchin, 2019). Critically redefining data science practices encourages us to adopt a pluriversal perspective (Escobar, 2018) to knowledge creation, and exchange and engage in a truly transdisciplinary way of working. Several multidisciplinary authors have proposed approaches to introduce criticality. In computer science, the work of Alicia E. Boyd (Boyd, 2021b) creates space to consider the relationship between existing inequalities and researcher positionality. Building on such scholarship, we present a framework for practising critical data science in education and research. The framework guides a future data scientist to consider inclusion of people or communities, existing inequalities, participation in the scientific process, unequal power dynamics, and positionality. We borrow from a non-exhaustive set of four critical theories (Decoloniality; Intersectionality; Transdisciplinarity; and Reflexivity), leaving the framework fluid and open to development. While the questions raised in this paper extend across disciplines busy with data science, our approach is rooted in long-standing debates in geographical thinking, which frames geography as a transdisciplinary and reflexive mode of inquiry concerned with uneven power relations and the politics of knowledge production (Sheppard, 2015). Geography brings a distinctive sensitivity to spatial relations, scale, and place that enriches critical data science. It reminds us that data are always rooted in place and generated, circulated, and acted upon within uneven geographies of power and infrastructure. Geography helps make visible the relational and situated aspects of data, showing how data are produced through dynamic social relations, power structures, and spatial processes that might otherwise remain hidden. Our aim is to provide a practical strategy for developing a truly open perspective to data science.

The importance of being critical

Observing university and industry cultures alike, the goal of a data scientist is to interpret large numbers of observations and to use those insights to make sense of the world (Ball

and Rague, 2022). Due to this focus on turning the numerical into the meaningful, data science is often viewed as objective (Iliadis and Russo, 2016). Objectivity relies on the existence of an objective reality, a reality existing of, as Daston and Galison (2021) write, ‘knowledge that bears no trace of the knower – knowledge unmarked by prejudice or skill, fantasy or judgement, wishing or striving’ (p. 17). Resultant insights are often considered beyond human intelligence, used to legitimise decisions (Cugurullo, 2024; Gitelman and Jackson, 2013).

Achieving objectivity in data science is impossible. Objects of a study are (in)directly influenced by diverse social processes (Boyd, 2021a; Cugurullo, 2024) and the understanding of those who practice data science is subjective, defined by their lived experiences (Takacs, 2003). Subsequently, resultant data is rarely self-explanatory and requires explanations from disciplinary experts, stakeholders, and theoretical frameworks based on previous studies (Gitelman and Jackson, 2013). Causal deductions require triangulation of different perspectives, datasets, and theories (Kandt and Batty, 2021) that are rarely part of a data science process.

The practice of data science is also affected by power. Historically, quantitative methods and mapping techniques were devised for colonisation (Arneil, 2020). This violent process of settlement has evolved into ‘data colonialism’ which combines historical extractive traditions with current day data science practices (Couldry and Mejias, 2019) to exploit data and labour for profit-making. Likewise, search algorithms can systematically enhance the visibility of some while distorting views of others, restructuring power relations and the distribution of capital for much of society. In this way, digital technologies can become oppressive mechanisms in an increasingly digitised world (Noble, 2016).

The idea that data science is subjective, and data are socio-technical constructs, should not topple any foundations of the field. Engaging with one’s own lived experiences, beliefs and norms promotes a process of reflection within research, exposing the contextual realities of the world which we often assume are outside the data science process (Jamieson et al., 2023). Harding (2012) argues that a study’s ability to properly describe reality improves when social context, and the pre-existing assumptions, beliefs, and prejudices layered beneath, are considered. In this way, more knowledge systems are integrated, making results meaningful to a wider array of stakeholders (Harding, 2012).

Critical data-science approaches invite practitioners to reflect on the influence of subjectivity on their results. A typical data science process, especially with the rise of Big Data, analyses records of phenomena that are easy to capture. Large datasets often exclude smaller data sources, marginalising important perspectives (Kitchin, 2013). Crime databases, for instance, easily record crimes that occur in public, making them a priority for response (Brayne, 2017). Moreover, these datasets often disproportionately focus on marginalised communities compared to the actual

crime rates in these communities (Beckett et al., 2005; Brayne, 2017). In this way, models and datasets that are used in everyday life have prejudices, bias and misunderstandings encoded in them (Brayne, 2017; O’Neil, 2016). It can be immensely dangerous to let these models and data define their own (distorted) reality, which their users in turn use to justify their results, especially when the outputs of models and data are falsely alluded to be objective to legitimise a particular perspective or voice (O’Neil, 2016). This objective action is seen in the disproportionate arrests made in black communities in the United States despite a relatively low crime rate (Beckett et al., 2005). It is easy to consider social and political motivations, much like in policing crime, outside of the data science box. Once biased records have found a home in a model, they become part of the black box and appear to be objective (Pasquale, 2015), underpinning policies that do anything but social good.

Data science – the traditional process of collecting, analysing, modelling and legitimising decisions of those in power – then needs to be inward looking. To confront, account for, and learn from our positions, we might understand how each step of the data science process may reveal and perpetuate our subjective influences and biases (Jamieson et al., 2023). As data scientists, we could engage with our own positionalities and implicit perspectives formally (Iliadis and Russo, 2016). By recognising the full spectrum of cultures and realities we are embedded in, we can design solutions and technologies that engage with this diversity more fully (Harding, 2012).

Critical data science

There are multiple visions for a critical data science process (Boyd, 2021a; Jamieson et al., 2023). According to Iliadis and Russo (2016) a critical data process includes questioning ‘the realities of shifting information infrastructures, multiple data subjects and their rights, deep information histories, work and power, and hybrid digital cultures that underpin’ (p.2) the process. Dalton and Thatcher (2014) envision a process that considers the historical developments leading to the realisation of ‘big data’, who is in control of the data, the motivations that drive the research, the subjects of the data and their knowledge systems, the role of data in the production of place and space, and the final use of the data. Kitchin and Lauriault (2014) suggest going further and consider the influence of the political economy, financing, and the subjectivities and communities of the people involved.

Our (the authors of this work) inherent research interests in urban inequalities mean that our data science practice has individually and collectively evolved to learn from various critical theories like intersectionality, reflexivity, decolonial theories (including feminist and critical race perspectives), and radical transdisciplinarity. Although much of qualitative social science formally engages with and celebrates

reflexive thinking (Jamieson et al., 2023), our research and teaching have followed a serendipitous process of imagining how the quantitative social sciences could learn from these critical theories.

With this ethos, we developed a master’s level spatial data science course over a five-year period. By formally engaging with critical theories, through reflexive thinking, and feedback from an enthusiastic and diverse student body, we evolved our practice of teaching data science. The course challenged the linear process of data science by carefully breaking it down into an iterative analysis where students were encouraged to engage with each step critically and with agency, seeing themselves as sources of knowledge.

Building on this experience, we integrated views of multiple authors including (Dalton and Thatcher, 2014; Harding, 2012; Iliadis and Russo, 2016; Jamieson et al., 2023; Kitchin, 2013; Kitchin and Lauriault, 2014) into one critical data science process based on four critical pillars. Reflexivity questions how maps construct reality, Intersectionality critiques the oversimplification of lived experiences, Decoloniality exposes the erasures and biases inherent in data collection, and Radical Transdisciplinarity calls for integrating multiple ways of knowing – including more-than-human perspectives. Alongside, we formulated several key questions relating to inequality, knowledge construction, power and positionality within data science.

We incorporate a visual representation that situates our four pillars within a critical cartography lens (Figure 1). The layer composition of the image challenges conventional spatial data practices by mirroring the multidimensionality of critical data science. It reinforces the need to move beyond reductionist, two-dimensional understandings of space to incorporate power relations, situated knowledge, lived experiences, and the interconnected human and more-than-human elements. We now discuss each pillar in turn.

The four pillars of critical data science

Decoloniality

Decolonial thinking questions how existing power dynamics and histories of oppression are reflected in data science (Couldry and Mejias, 2019). Decolonialisation is the process via which bureaucratic, cultural, linguistic and psychological divestment from colonial power occurs (Smith, 2021), challenging the dominance of Euro-American universalising approaches to understanding and theorising the world (Radcliffe, 2017). Such processes call for reclaiming mapping and data practices that centre Indigenous and community epistemologies as opposed to technocratic control (Rose-Redwood et al., 2020). Within data science, there is a risk that data is employed to support those already in power and to disadvantage the already disadvantaged (O’Neil, 2016). Moreover, digital spaces, like physical spaces, can also become spaces of extraction and

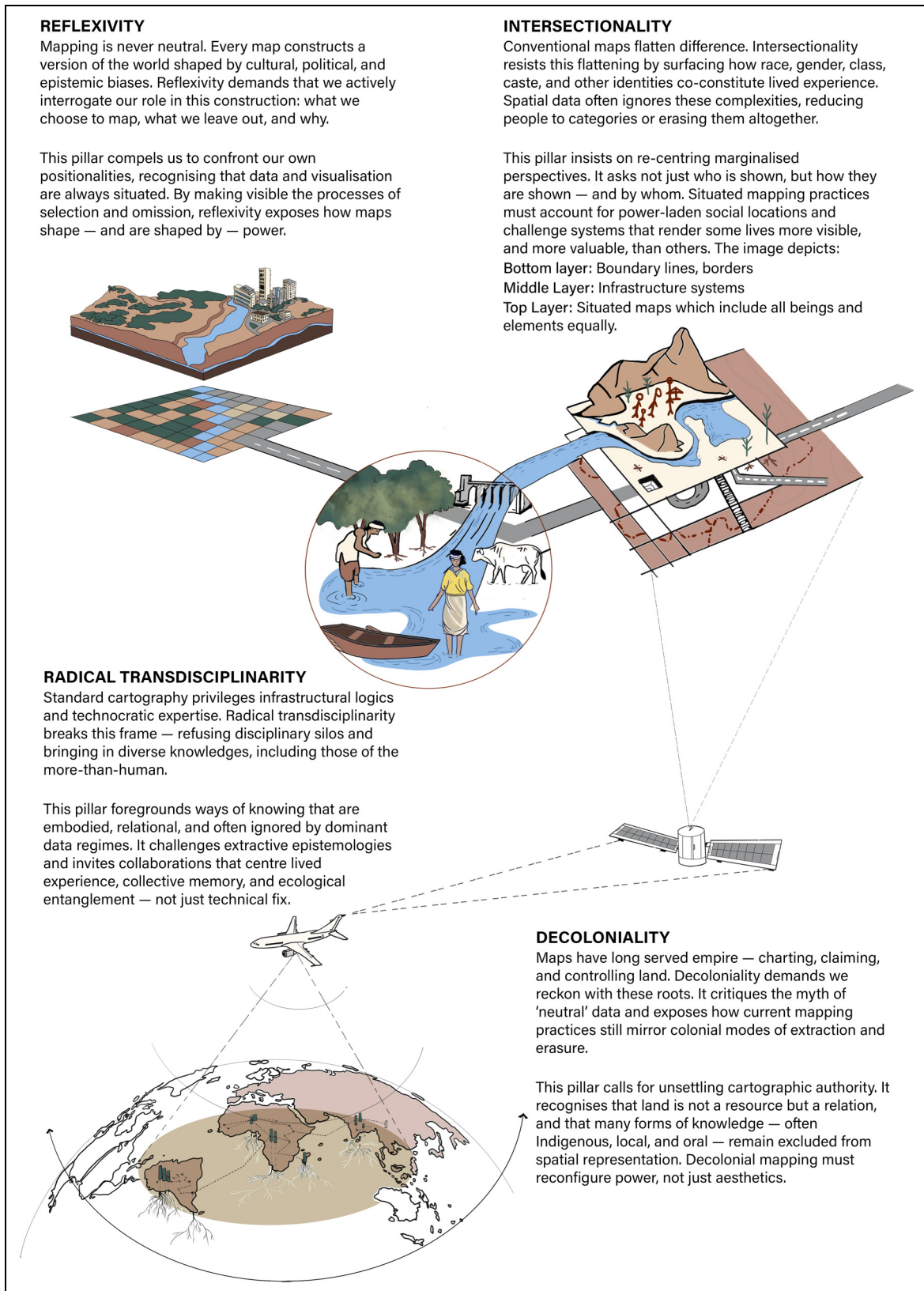


Figure 1. The four pillars of critical data science through a cartographic lens. While satellite and aerial imagery have colonial origins and often served extractive purposes, they are now widely available as open data. Critical data science encourages reclaiming these tools — using open imagery and remote-sensing data for community-led mapping, environmental justice, and spatial equity — turning historically dominant infrastructure into instruments of collective and situated knowledge.
- art by an artist-researcher, Namrata Narendra.

exploitation, subject to ‘digital coloniality’ (Mohamed et al., 2020). Digital infrastructure is designed to extract, circulate, and analyse data without obtaining meaningful consent of those that produce the data. For example, using coercive practices or bundled fine print that nobody reads, location, CCTV, call detail records, and trip patterns are often used in the classroom and research projects in domains such as mobility/urban analytics (Mattern, 2020). This relationship becomes more oppressive as producers have a lesser control over their data, increasingly viewed as capital (Singh, 2023). Instead, data is often held by large, politically powerful Western corporations (Couldry and Mejias, 2019), and is often not freely accessible (Richards and King, 2013; Rowe, 2021).

Decolonial thinking is an invitation to critically examine and critique the politics of race and colonial ways of thinking that are visible in today’s technology and data that exclude, limit, or discredit ways of thinking outside Western standards (Adams, 2021). Decolonial thinking is local; it is critically aware of the social mechanisms that reproduce racism and discrimination; it acknowledges power dynamics and their influence on the production of ‘good’ knowledge (Adams, 2021). Efforts to radically rethink and decolonise knowledge production are increasingly informing education and research in diverse disciplinary and higher education settings (Adébişi, 2024; Birhane and Guest, 2020; Laing, 2021; Noxolo, 2017). Therefore, this pillar is central to the data science process, localising knowledge in confronting histories and power but globalising impact by reconfiguring power dynamics and highlighting new ways of knowledge productive within the quantitative social sciences.

Intersectionality

Intersectionality questions the use of data and data science to reinforce existing structural and systemic oppression and marginalisation. Particular groups are boxed in simple categories in contemporary society, and in turn in data and data science. Data and analytical models are created disproportionately by a small, under-representative group of men (Dignazio and Klein, 2020; Tacheva, 2022). With disproportionate power to influence decisions, the perspectives of other societal groups can be excluded from, or distorted by, mainstream data science (Dignazio and Klein, 2020). Or worse, data science can be used to further oppress and dispossess specific groups, such as in the case of the Dutch government wrongfully accusing families of childcare benefit fraud (European Parliament, 2022), or the persecution of Uyghurs and other Turkic Muslim minorities in China (Oztig, 2023).

An intersectional perspective goes beyond revealing biases in data science. In particular, it critically challenges the structural oppression of people that face discrimination across race, gender, ethnicity, and/or class (Carastathis,

2014; Cho et al., 2013; Lee et al., 2022). The intersection between categories of discrimination leads to under-representation of these groups in society (Carastathis, 2014; Cho et al., 2013) and, therefore, also in data (Bowleg and Bauer 2016). Thus, intersectionality in data science explicitly considers people who are faced with multiple structural inequalities, and how such intersectional experiences can influence their representation in analysis.

Radical transdisciplinarity

Calls for transcending disciplinary boundaries have been around for years. While interdisciplinarity involves integrating knowledge from different disciplines, transdisciplinarity goes further by integrating non-disciplinary knowledge, often involving groups from private, public, and/or civil society. The choice of who is involved in the data science process greatly influences both the process and its outcomes. If the process excludes the perspectives and lived experiences of those being researched, the data may present a different narrative that differs from the situated struggles (Lee et al., 2022). This is increasingly harmful if those excluded have a history of being marginalised and excluded from the dominant narrative (Smith, 2021). Thus, transdisciplinarity has the potential to address issues of bias and misrecognition in datasets and models.

There are two approaches to transdisciplinarity. A pragmatic approach enables collaborations around a specific problem, where participants jointly define new research questions and methodologies (Hansson and Polk, 2018; Robinson and Tansey, 2006; Talwar et al., 2011). Active and meaningful involvement of (marginalised) communities in research improves the quality of project results by including bottom-up knowledge (Eckhardt et al., 2021; Hendricks et al., 2022). Moreover, such involvement empowers communities in decision-making, leading to more just and democratic processes (Hendricks et al., 2022). However, this approach falls short of challenging the structural and systemic dynamics that create these issues in the first place. A radical approach, in contrast, challenges the very meaning of knowledge, recognising and asserting other ways of knowing and living. In doing so, it enables the emergence of truly radical alternatives to existing societal issues and crises.

Through engaging with decoloniality and intersectionality, radical transdisciplinarity recognises that different knowledge systems are valid and bring new perspectives to education and research (Brier, 2000; de Vos et al., 2022; Lodge et al., 2017). Knowledge creation is approached as knowledge construction, where different knowledge systems are combined to understand the complexity of our world (Brier, 2000). This requires knowledge systems to be situated within existing social struggles (Goncalves et al., 2025a; Kitchin, 2014; Mattern, 2020), for example, through the collective documentation of social processes (Maharawal and

McElroy, 2018; Mattern, 2020; Tschakert et al., 2016), as well as actions that counter the increasing commodification of data and knowledge (Rose et al., 2021; Thatcher et al., 2016), such as community-owned data infrastructures, grassroots technologies, public digital infrastructures, and open access practices in governance, research and education (Cardullo and Kitchin, 2019; Goncalves et al., 2024; Maharawal and McElroy, 2018; Paris et al., 2024). These practices not only redistribute power in data science but also embed digital practices within broader struggles for epistemic justice.

Reflexivity

Reflexivity intersects the three pillars above and invites the researcher to reflect on their positionality throughout the data science process. Encouraged by an iterative and continuous process of knowledge production (Boyd, 2021a; Jamieson et al., 2023), the researcher could acknowledge that they are not distant and independent from the data (Ricker, 2017). A researcher might not always be aware of their influence on the research process. Reflexivity can help the researcher understand their own place in the research and how their framing of a topic influences their understanding (Woroniecki et al., 2019). It calls for recognising and addressing one's own privileges, influence on biases in data science, and the resultant impacts on society. For example, if there are potential harms that can arise from choices about how to aggregate and link data (Saltz, 2019). By taking a reflexive approach, the researcher can meaningfully engage with or shape the underlying motivation of the research and help prevent erasure, harm, or silencing underrepresented voices as much as possible (Ricker, 2017).

Navigating the process of critical data science

Drawing on the four critical pillars and our experience in practice, we propose five key questions to prompt the researcher to facilitate a Critical Data Science process. The questions relate to inclusion, inequality, participation, power, and positionality. In doing so, we build on existing frameworks that act as pivots and pirouettes (Ulmer, 2020), shaping our approach. Inclusion, inequality, and reflexivity are three core dimensions in the Quantitative Intersectional Data Science handbook by Alicia E. Boyd (2021b) and data representativeness, social justice and democratising access are at the core of the approach suggested by Nelson et al. (2022) for ethical and equitable spatial data science. Reflexivity, especially, finds central place in scholarship presented by Jamieson et al. (2023). In a departure from the algorithmic approach of quantitative sciences, Critical Data Science emphasises the process to be viewed as dynamic, depending on context. Hence, we do not assign a specific order of importance to the questions, as rankings

are inherently tied to power and are sources of bias (Lee et al., 2022). We invite the researcher to explore how these prompts could shape meaningful engagement with the data science process.

Who is (not) included in the data? (inclusion)

Data scientists interpret large amounts of information to gain new insights. Data practices often reflect historical and contemporary power dynamics (Boyd, 2021b; Couldry and Mejias, 2019; Lee et al., 2022). Who and what is included in a dataset influences the outcomes of research. Datasets are never complete – social groups are not entities awaiting representation but are actively produced and legitimised through the modelling choices, infrastructures that structure data, and policy levers, such as how neighbourhoods are delineated or variables selected for targeting policy action (Vybornova and Verma, 2025). For example, the census notoriously excludes those groups for which data is hard to collect (e.g., those without (a permanent) home) (Kearns, 2012; Schneider et al., 2016) or distorts second or later generation migrants into binaries of citizenship (Van Schie et al., 2020). Handling missing values is an important skill and should not be accounted for by algorithms that average out or exclude the realities of entire social groups (Lee et al., 2022). This happens frequently in studies about gender using datasets that underrepresent LGBTQ+ people (Guyan, 2022). Societal groups missing in data risk being erased across wider societal and political discussions (Guyan, 2022), especially those who have a history of being excluded (Smith, 2021). Critically reflecting on the inclusivity makes the researcher more aware of how different forms of oppression are reflected in their data (Lee et al., 2022). As the inclusivity of a dataset grows, it can also provide a more nuanced representation of the situation it is attempting to analyse and uncover hidden complexities (Guyan, 2022).

What role does inequality play in data science methods? (inequality)

Closely related to inclusion, this question specifically invites reflection on the analytical methods used to interpret data, rather than the data itself. Inequality can be encoded not only in what data represent but also in how they are modelled, correlated, and interpreted (Brayne, 2017; O'Neil, 2016). Methods such as exploratory (spatial) data analysis determine which variables are selected, how features are engineered, and which relationships are prioritised. When correlations are identified without theoretical or contextual grounding, variables such as ethnicity, income, or location can be misused as predictive features in decision-making pipelines. Such methodological biases may appear technically sound, but they reproduce harmful social assumptions. For example, in the Netherlands, when ethnicity

was incorporated into predictive models for detecting tax fraud, thousands of families were wrongfully penalised (Hadwick and Lan, 2021), leading to severe financial and social harm (European Parliament, 2022). These outcomes illustrate that methods themselves – through design choices, normalised workflows, and the absence of reflexivity – can entrench inequality. Reviewing analytical approaches with this awareness can reveal how prejudice becomes operationalised in code, guiding more just modes of analysis and feature selection (Boyd, 2021b). Students who learn these techniques without critical framing may later reproduce such biases in policy or commercial settings, further perpetuating structural harm.

Who is (not) involved in the data science process? (participation)

A diverse perspective is as instrumental in Critical Data Science as the use of a representative dataset. Research is always for the large part shaped by the perspective of its participants (Kusek and Smiley, 2014), especially when the interpretation of the results intertwines with the real-life situations that they describe (Boyd, 2021a). Diversity can be celebrated as helping to construct a more nuanced view of the world, bringing together different knowledge systems (Brier, 2000; de Vos et al., 2022; Harding, 2012). Research has shown that through active and meaningful involvement of marginalised communities in projects, the quality of the project outcomes improves by leveraging bottom-up knowledge (Eckhardt et al., 2021; Hendricks et al., 2022). Moreover, because of their involvement, communities are empowered in the decision-making process, leading to a more just and democratic process (Hendricks et al., 2022). Furthermore, meaningful participation teaches us to move away from the dominant paradigm to encompass perspectives and methodologies of researchers from the Global South, which are commonly excluded in the dominant Western ways of working (Davies and Standring, 2023). It is important to scrutinise the Western paradigms from this wider point of view (Davies and Standring, 2023). The extent and form of participation will vary across research contexts. Not all projects can or should involve direct engagement with data subjects, especially when confidentiality or data sensitivity requirements make this impractical or unethical. The framework therefore invites researchers to adapt participation thoughtfully within the ethical and practical constraints of their own contexts.

How does the data science process reflect existing power dynamics? (power)

The inclusion of certain perspectives in data, the composition of the research team, and the choice of method of analysis, all start and end with questions of power. With the growing

influence of data science in today's society, so-called 'informational power' is becoming increasingly important (Hoffmann, 2021). Informational power is the power to decide which information or data counts as valid, but also the power to decide whose perspective is valid to include. Often, informational power is grounded in historical patterns of privilege and oppression (Hoffmann, 2021; Potts and Brown, 2005; Singh, 2023). Such hidden power influences all parts of the data science process and is of relevance across all pillars of Critical Data Science. Consider, for example, energy inequality. Energy transitions aided by government subsidies exacerbate energy vulnerabilities if they are not tailored to differing capacities of households (Kraaijvanger et al., 2023). The intersections between different categories of oppression here might greatly influence specific outcomes (Carastathis, 2014), like risking the vulnerable households to pay twice for someone else's transition process (Kraaijvanger et al., 2023). Understanding of power by a researcher has a large influence on their behaviour in the research process, but also on their views about how data is assembled, whose experiences or realities are represented, what research questions are important (Woroniecki et al., 2019), and how to identify and engage with caring futures for the subjects of their study. It is important that researchers consider power dynamics, their own positions of power, and how these influence the research process (Boyd, 2021b; Woroniecki et al., 2019).

What is the researcher's own positionality and motivation? (positionality)

Positionality is a cross-cutting issue. Knowledge production is a constant and iterative process to which the researcher and their perspectives are highly connected (Boyd, 2021a; Ricker, 2017). A truly critical data science process invites us to critically examine our positionality at multiple stages. To prevent the inadvertent projection of our own biases on the research, reflection is not only emphasised for each step of the data science process, but we are also encouraged to reflect on how our own lived experiences influence the research questions and outcomes. Such reflexivity echoes long-standing arguments in feminist geography for situated knowledges and accountability in knowledge production (Rose, 1997).

A guide for critical data science

Our five key questions are integrated into an iterative and cyclic step-by-step approach that supports researchers and educators to critically examine the entire data science process. The most common and traditional process is divided into the steps of 'Focus of Analysis', 'Collect & Combine Data', 'Transform Data', 'Analyse Data', 'Interpret & Visualise Data', and 'Communicate Findings'. Each step encourages researchers to rigorously examine not only technical aspects of data handling but also the broader societal

implications of their work. By explicitly questioning aspects of inclusion, inequality, participation, power dynamics, and their own positionality, we expect that data scientists can guide themselves towards a shared and more responsible and ethical practice, challenging the dominant paradigms within data science, urging practitioners to reflect on whose voices are amplified or silenced through their choices, and aims at creating shared research learnings that genuinely contribute to social justice across diverse contexts.

Table 1 presents the suggested approach. While the table follows the familiar structure of a research workflow, we recognise that contemporary data-science practice often departs from this linear model. Data science is inherently exploratory and iterative, characterised by stages of data mining, feature engineering, model training, validation, and automation that continually reshape both questions and outcomes. These algorithmic phases amplify concerns around bias propagation, privacy, opacity, and accountability, especially when working with large-scale, unsampled data drawn from administrative, sensor, or platform sources. The framework therefore should not be read as a prescriptive sequence but as a reflexive scaffold that overlays both qualitative and computational workflows. Each stage – from defining the focus of analysis (how problems are framed) to analysing data (how models and interpretations are produced) – invites researchers to interrogate how inclusion, inequality, participation, power, and positionality manifest in automated systems. Thus, the Critical Data Science process expands beyond traditional research design to encompass the ethical and political questions that arise uniquely in algorithmic and data-intensive environments (Keßler and McKenzie, 2018; Wiltshire and Albanides, 2022).

To accompany the step-by-step guide in Table 1, we include a visual representation of the data science process grounded in a specific spatial context (Figure 2, Table 2). This image serves to anchor each stage of the process – from defining the focus of analysis to communicating findings – within the everyday geographies and lived realities that conventional data practices frequently abstract away. By situating the process within a neighbourhood facing flood risk in Indonesia, the image makes visible how structural inequalities, colonial legacies, and dominant spatial narratives are embedded within data practices. Visualisation choices themselves hold power: Different design decisions can emphasise, obscure, or politicise spatial patterns (Ricker et al., 2020). It challenges the notion of mapping as a neutral or technical exercise, instead showing how spatial knowledge is contested, relational, and shaped by power. In doing so, the visual reinforces the argument that a Critical Data Science practice can be attentive to context, grounded in the everyday, and capable of surfacing voices, knowledges, and ways of seeing that are routinely made invisible. This process can also be applied to a non-spatial practice of data science in a similar way.

Figure 2 should be read as an illustrative example of a visual language that is used metaphorically to represent

how spatial knowledge is produced through multiple, coexisting epistemologies: the coexistence of official spatial plans and infrastructure, and lived, and situated knowledges within the same analytical frame. Lived experience is incorporated in maps by changing the processes through which data are produced, transformed, and interpreted. Situated knowledge may take the form of oral histories, everyday navigational practices, customary land relations, or community-defined boundaries of risk, all of which can be meaningfully incorporated within conventional two-dimensional cartography. Such reflexive visualisation practices retain context, make visible whose knowledge counts at each stage of the data science process, and resists the erasure produced by techniques in Big Data such as aggregation and standardisation. Figure 2, therefore, functions as a critical device intended to challenge dominant assumptions of neutrality and objectivity in mapping by making explicit how spatial representations are shaped by power, positionality, and lived realities, and how alternative mapping practices can elicit these dimensions with care.

Drawing on accumulated teaching experience of the authors, the examples of Table 2 show how students were encouraged to move beyond critique alone by redefining problems, integrating alternative and situated knowledge sources, and reconsidering analytical and visualisation choices at different stages of the data science process. These shifts point toward substantively different ways of understanding flood risk, including the identification of vulnerabilities that are often distorted by conventional datasets and modelling practices. Instead of presenting a universal or prescriptive solution, the table illustrates how Critical Data Science can create space for more situated and reflexive analytical discussions within the classroom.

Critical thinking is central to data science

Current approaches to data science typically ignore social realities and cannot properly account for the subjectivities underlying the data science process, leading to potentially exclusionary and unjust questions, practices, and outcomes. In response, we present a new critical paradigm for data science research and education, ‘Critical Data Science’. Critical Data Science builds on four main theories of decoloniality, intersectionality, radical transdisciplinarity, and reflexivity. These theories encourage questions about data inclusivity, inequality in data science methods, knowledge creation, power dynamics, and researcher positionality. Using these questions as a basis, we propose a detailed step-by-step approach that outlines specific considerations for each step of the research process.

Critical Data Science encourages researchers and students to question and go beyond the dominant, Western paradigm in data science, instead amplifying perspectives and methods commonly excluded. The theories incorporated in our framework give more space to these non-Western

Table 1. A structured, iterative approach to critical data science.

Data science process	Inclusion	Inequality	Participation	Power	Positionality
Focus of Analysis: <i>Theories, processes & stakeholders that drive the analysis</i>	Investigate exclusive practices of past and present research. Consider how these affect the diversity of people represented in the data.	Consider the tools reliably usable for the research. Research method limitations, particularly their influence on structural inequalities.	Develop a participatory modelling approach that involves stakeholders not typically included. Discuss multiple framings of the topic.	Investigate power relations and injustices within sampled populations and data.	Critically reflect on your own position: Why this topic? Is your involvement suitable? Possible biases? Potential harm or erasure to communities?
Collect and Combine Data: <i>Data collection, dataset creation, ethics</i>	Examine who is central to data collection, especially data collected or extracted using excluded groups.	Assess structural inequalities reflected by collection methods including proxies, aggregated data and sampling biases. Consider alternative sources to recognise and empower neglected groups.	Verify all relational* partners' involvement and seek input from affected communities.	Recognise data ownership by participants. Implement ethical data-collection practices. Explain purpose and usage. Obtain informed consent.	Critically reflect on data-collection choices: Why these data sources, relational partners, and sampling methods? Whose voices are amplified or silenced and why?
Transform data: <i>Completeness, missing data, consistency, pluralism & accuracy of collected data</i>	Identify missing data (diversity in variables) and whose data is absent (diversity in sources).	Assess impacts of data cleaning on representation of perspectives including aggregating, imputing and slicing. Explore alternatives for cleaning data.	Maintain transparency in transformation decisions; Explore alternatives and implications collaboratively.	Evaluate if data cleaning, including normalisation or standardisation, reinforces dominant narratives or interpretations.	Critically reflect on data-cleaning methods: Why these methods? Whose voices are amplified or silenced, and why?
Analyse Data: <i>Representation, dominant narratives & uncertainty</i>	Reflect on representation and exclusion in analysis and its impact on learnings.	Test sensitivity of methods to biased data. Examine structural differences within samples of population.	Analyse and reflect on intermediate results with stakeholders. Share preliminary results and decisions with those with lived experience of a problem. Clarify decisions about visuals.	Contextualise analysis close to original data environment. Reflect on power dynamics and implications for findings.	Critically reflect on data analysis: How might your own proximity to data shape the analysis? Do stakeholder-representation choices matter? Whose voices are amplified or silenced, and why?
Interpret & Visualise Data: <i>Representativeness, reflect on complexity & accuracy</i>	Consider representation of viewpoints of all stakeholders in interpretation and visualisation.	Link interpretation to questions of inequality and marginalisation; Assess choices explicitly.	Acknowledge different interpretations of data. Incorporate affected communities' viewpoints in shared analysis and interpretation.	Examine if visualisations or interpretations reinforce dominant frameworks.	Critically reevaluate research motivations and framing: Is your storytelling necessary and just? Are you reinforcing dominant narratives? Whose voices are amplified or silenced, and why?

*the use of the term relational partners refers to the structuring of research through ongoing relationships of care, accountability, and power-sharing, rather than transactional or extractive forms of engagement.

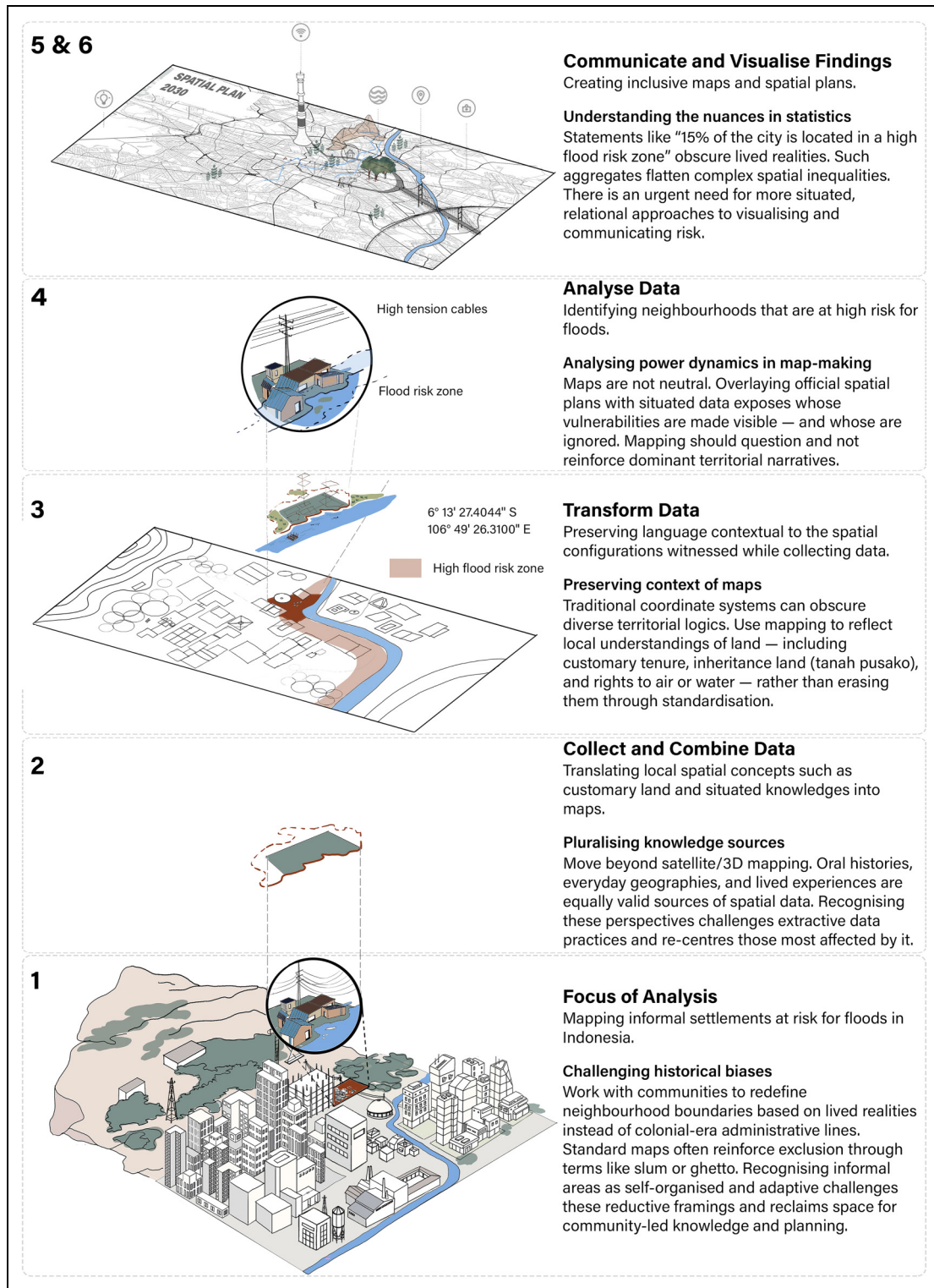


Figure 2. Visualising the critical data science process through a spatial lens. Aggregate datasets can hide uneven risks and lived experiences. Situated approaches combine quantitative models with community knowledge — oral histories (of those who tell them and as such cannot be expropriated), testimonies, and participatory maps — to reveal how risks and inequalities intersect. Though such methods could risk over-localisation or privacy issues, they can equally foreground whose realities count in spatial planning. Situated data are locally grounded observations produced collaboratively, not extracted without consent (like Big Data). While terms like slum or ghetto can stigmatise, co-produced and reflexive maps can also empower and support vulnerable communities.
- art by an artist-researcher, Namrata Narendra.

Table 2. Articulating how specific data science issues can be understood through teaching practices in the classroom.

Stage of practice	Issues in data science	Classroom activity (<i>what students could do differently</i>)	Outcome (<i>changes in the analysis</i>)
Focus of analysis	Problem framing, variable selection, optimisation goals. Models risk reducing complex social–environmental processes to technical categories (e.g., ‘high-risk zones’).	Examine how ‘flood risk’ is defined in government data versus community narratives. Identify whose knowledge informs model boundaries and discuss alternative framings of vulnerability: reframe risk as a socio-spatial condition shaped by tenure insecurity, informal infrastructure, and everyday mobility constraints.	Flood risk treated not as a bounded hazard area but as a relational process unevenly experienced across households and neighbourhoods.
Collect and combine data	Dependence on satellite or sensor feeds; limited coverage of informal areas; ethical use of volunteered or community data.	Compare official spatial data with participatory mapping of informal settlements. Discuss consent, ownership, and representation when integrating local oral histories or customary land boundaries into GIS datasets.	Reveals discrepancies between officially mapped ‘safe’ zones and areas residents identify as repeatedly flooded or inaccessible during monsoon events, especially in informal settlements.
Transform data	Spatial discretisation, aggregation, coordinate systems, and translation of qualitative observations into quantitative layers.	Test how reprojecting or reclassifying community maps into standard coordinate systems alters visibility of local tenure types (e.g., <i>tanah pusako</i>) and access to water or air rights.	Identifies how reprojection and reclassification distorts vulnerability by erasing informal tenure types and masking fine-grained exposure along small waterways and alleys.
Analyse and model data	Algorithm choice (e.g., flood-susceptibility prediction); uncertainty propagation; power embedded in overlaying official and situated maps.	Overlay official flood-risk models with community-generated spatial data. Evaluate whose vulnerabilities are amplified or distorted. Debate ethical implications of automated ‘risk targeting’.	Modelled risk systematically underestimates exposure for households lacking formal infrastructure and positions risk in formally recognised zones, prompting discussion of ethical limits of automated risk targeting.
Communicate and visualise findings	Statistical aggregation hides inequality; cartographic conventions privilege state narratives; reproducibility and transparency.	Create two map versions of flood exposure: one aggregated (‘15% of the city’) and one relational (linking place, tenure type, and lived experience). Reflect on how visual choices shape interpretation and accountability.	Makes explicit how conventional cartographic choices legitimise state-centric narratives of risk, while situated visualisations highlight whose vulnerabilities are prioritised or ignored in planning decisions.

Read table in association with Figure 2 using the example of ‘Flooding in Indonesia’.

perspectives, such as those stemming from decolonial thinking, and methods such as critical cartography, participatory mapping, feminist HCI, Indigenous GIS, and the use of oral histories. Following this, data science should not be viewed as a purely technical, objective tool, but as a subjective approach inherently tied to its social, historical and political contexts. In this way, it aims to produce outcomes that better reflect the complexities of the realities that we seek to understand and could be seen as more just and fairer. And by making space for those perspectives and methods that do not have a place in common data science practices, results could be valid in a wider array of contexts and to more stakeholders (Harding, 2012), thus valuing community-building. The critical data scientist is also better able to go beyond their own limited framing of how the world works,

seeking to interpret data and analysis in a way that is also recognisable to those with different worldviews.

The critical approach could also reduce several ethical concerns in current data science methods. Increased access and transparency of data increases privacy risks (Rowe, 2021), requiring a balance between improvements to analytical power and safeguarding of privacy (Nelson et al., 2022). Additionally, multiple authors argue for increased attention to ascertain meaningful consent during data collection since the ‘data is only borrowed from the producer’ (Goncalves et al., 2024; Hand, 2018; Harrington et al., 2019; Nash et al., 2022). Questions of consent, unequal power, access, and inequality are intrinsic to the approach proposed from the very beginning, helping to create an ethically sound

data science practice that goes beyond performative and superficial practice.

The critical data science method is centred around radical transdisciplinarity, which challenges the meaning and methods of current-day data science and enables truly radical thinking about existing societal issues and crises. This radical perspective becomes increasingly important in a world that is plagued by challenges that transcend one specific discipline. For example, climate change is deeply connected with Western cultures of excessive growth (Brand-Correa et al., 2022; Goncalves et al., 2025a), making it both a problem of climate science as well as a socio-economic challenge. These are also challenges for which data science is increasingly employed, highlighting the importance of a data scientist that is able to handle this non-disciplinary context. Our approach stimulates the consideration of such interconnections by, for example, making an explicit link between representation in current-day datasets and historical exclusion. The critical data scientist strives to operate in a transdisciplinary context, empowering them to meaningfully respond to global societal challenges.

While this approach was developed in the context of and for use in data science education and research, it does not limit its application solely to this discipline. Considerations of inclusion, inequality, participation, power, and positionality are equally important beyond the data sciences. Calls for participatory, inclusive, and decolonial research can be found across disciplines (see, for example, Thambinathan and Kinsella, 2021 and Watharow and Wayland, 2022). We encourage researchers from other disciplines to use our approach as a basis for developing their own critical frameworks for their research, such as qualitative approaches or studies using a mixed-methods approach.

As the world is digitising, the influence of data on our ideals of equity, sustainability, justice grows (Hoffmann, 2021). At the same time, these ideals become more important as more decisions become driven by data. The scale and pace of developments in AI, for example, further complicate the ability of researchers to critically scrutinise these technologies, their outputs, and their implications for societal decision-making processes, especially for the most marginalised who are least likely to be engaged in their use (Mohamed et al., 2020). This is particularly pertinent given the dominance of western corporations in the ownership and development of new AI technologies. The use of AI will necessitate a similarly critical response that is currently beyond the scope of this paper. For example, given the computational and resource intensity of AI (Hao, 2019), how should researchers decide whether analyses are justifiable in an era of a warming climate? Here, emerging debates about frugal AI are likely to become important as part of a Critical Data Science framework, balancing researcher responsibilities to achieve data, social, and environmental justice (Dencik et al., 2017).

While the theoretical pillars and questions presented are the building blocks of a Critical Data Science approach, a few important footnotes remain. First, this approach should not be seen as a definitive prescription. Formulating universal solutions is in itself problematic, as this dismisses the experience of those that do not recognise themselves in it (Fileborn and Trott, 2022). This approach was developed as a starting point for critical views on data science and may be seen as a living document that can be expanded to realise other knowledge systems, contexts, and practices.

Second, the development of the Critical Data Science approach (and the writing of this paper) is affected by the positionality of the research team. As the lead author, the perspective of Laura van Geene was the most influential in the selection of theories and formulation of key questions. As a young, White, Dutch environmental scientist, her perspective is, in part, informed by the colonisation of knowledge within the Dutch education system. Due to her background and previous education, she has been awarded many privileges and, therefore believes, can never fully understand the lived experiences of those facing discrimination and oppression. Juliana Goncalves is a Brazilian, cis-gender woman. As an academic from the Global South working in a Dutch university, her positionality is shaped by both privilege and marginalisation. Coming from a middle-class background in Brazil, she benefits from social and economic privileges within her home country, where she is perceived as a White person. However, upon relocating to Europe, this racial categorisation shifts, and she finds herself navigating a more complex racial identity, where whiteness is defined differently and she is no longer seen as White. This dual experience informs her academic work and professional attitude, as she constantly negotiates the intersections of race, class, and geography, balancing the privilege she holds within her home context with the challenges of being 'othered' in Europe. Caitlin Robinson is a White, British, cis-gender woman, who was benefited from public funding to enable her to access higher education at British institutions. She is mindful that she has a lot to learn, especially when considering processes of decolonising research, and that this process may be led by Indigenous and decolonial scholars. Trivik Verma is of Indian heritage and has lived in Europe for more than a decade. He has received multiple public funds to support his education. This work was conceived after witnessing how Western universities dominate education and policy globally. His roots in Indian cultures, participation in India's service economy, and the rapid pace of urbanisation in India underpinning Western technical prowess has shaped his thinking. Disappointingly, he has found limited space in debates centred around data science for scholars of colour, especially for non-male genders. Although our team has a deeper understanding of the world collectively than any of us could hope alone, we are mindful that we are not fully representative of communities that have been




excluded from the promises of data science, not least the lower income groups of South and South-East Asia, Africa, and South America who power the tech industry in the West. We are aware that our positionalities might have led us to unintentionally exclude perspectives from this process.

We offer a vision for Critical Data Science that acknowledges the injustices of society and illuminates a path forward for recognition and reparation, transforming data practices into tools for justice and inclusion. We hope that researchers and students reclaim data science for collective learning and care, and in doing so set precedence for a rigorous but also compassionate and socially responsive science.

Acknowledgements

The authors would like to thank Javier Trescoli Garcia, Nachiket Kondhalkar, and Philip Müller for their help in developing teaching materials for the course that is the passive object of this article, as well as Sterre ter Haar for help and encouragement during the writing process.

ORCID iDs

Laura van Geene  <https://orcid.org/0009-0007-7108-4028>
 Juliana Goncalves  <https://orcid.org/0000-0001-6888-3362>
 Trivik Verma  <https://orcid.org/0000-0003-2761-5155>

Author contributions

The images in this work are designed by artist-researcher Namrata Narendra, who engaged deeply with our research process to translate theoretical and conceptual insights into a visual and spatial representation. Using their engagement with critical cartography, the illustrations serve as an interpretative tool, making abstract concepts more tangible while simultaneously challenging dominant data mapping paradigms. Their work exemplifies how art and research can intersect to question power, representation, and inclusion in data science. Juliana Gonçalves and Trivik Verma designed the study and secured the funding. Trivik Verma designed and delivered the teaching material this article draws on. It is available in an open-source format. Laura van Geene led the design of process workflow and distilled teaching practices into a framework for data science in the classroom, as a teaching assistant. Laura, Juliana and Trivik analysed the teaching material, and Laura drafted the manuscript. Caitlin Robinson engaged with the teaching material and designed a framework for decolonised data science – the foundation of our process. Caitlin and Juliana reviewed and revised the manuscript for publication. All authors contributed to the writing of the publication.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful for the funding provided by the Open Education Programme at Delft University of Technology. Caitlin Robinson's time is funded by a United Kingdom Research and Innovation

(UKRI) Future Leaders Fellowship Mapping Ambient Vulnerabilities (MR/V021672/2).

Declaration of conflicting interests

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

Note

1. A massive open online course is a method for delivering learning material to any person who wants to take a course, with lower barriers for accessing expensive educational programs. This is done using digital technologies in an online environment.

References

- Adams R (2021) Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews* 46(1-2): 176–197.
- Adébişi F (2024) *Decolonisation and Legal Knowledge: Reflections on Power and Possibility*. Bristol, UK: Bristol University Press. doi:10.46692/9781529219401.
- Arneil B (2020) Origins: Colonies and statistics. *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 53(4): 735–754.
- Ball R and Rague B (2022) Introduction to data science. In: Ball R and Rague B (eds) *The Beginner's Guide to Data Science*. Cham: Springer International Publishing, 1–13. doi:10.1007/978-3-031-07865-11.
- Beckett K, Nyrop K, Pfingst L, et al. (2005) Drug use, drug possession arrests, and the question of race: Lessons from Seattle. *Social Problems* 52(3): 419–441.
- Birhane A and Guest O (2020) Towards decolonising computational sciences. arXiv preprint arXiv:2009.14258.
- Bowleg L and Bauer G (2016) Invited reflection: Quantifying intersectionality. *Psychology of Women Quarterly* 40(3): 337–341.
- Boyd A (2021a) Intersectionality and reflexivity—decolonizing methodologies for the data science process. *Patterns* 2(12): 100386.
- Boyd AE (2021b, June) *Quantitative intersectional data (quinta): A #metoo case study*. Doctoral dissertation. DePaul University.
- Boyd E and Juhola S (2015) Adaptive climate change governance for urban resilience. *Urban Studies* 52(7): 1234–1264.
- Brand-Correa L, Brook A, Büchs M, et al. (2022) Economics for people and planet—moving beyond the neoclassical paradigm. *The Lancet Planetary Health* 6(4): e371–e379.
- Brayne S (2017) Big data surveillance: The case of policing. *American Sociological Review* 82(5): 977–1008.
- Brier S (2000) Trans-scientific frameworks of knowing: Complementarity views of the different types of human knowledge. *Systems Research and Behavioral Science* 17(5): 433–458.
- Carastathis A (2014) The concept of intersectionality in feminist theory. *Philosophy Compass* 9(5): 304–314.
- Cardullo P and Kitchin R (2019) Smart urbanism and smart citizenship: The neoliberal logic of 'citizen-focused' smart cities in Europe. *Environment and Planning C: Politics and Space* 37(5): 813–830.

- Cho S, Crenshaw KW and McCall L (2013) Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society* 38(4): 785–810.
- Cottineau C, Batty M, Benenson I, et al. (2024) The role of analytical models and their circulation in urban studies and policy. *Urban Studies* 61(12): 2370–2398. doi:10.1177/00420980241237410.
- Couldry N and Mejias UA (2019) Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20(4): 336–349.
- Cugurullo F (2024) The obscure politics of artificial intelligence: A marxian socio-technical critique of the AI alignment problem thesis. *AI and Ethics* 5(1): 1189–1201. doi:10.1007/s43681-024-00476-9.
- Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? seven points for a critical approach to 'big data'. *Society and Space* 29. <https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care>
- Daston L and Galison P (2021) *Objectivity*. Princeton, NJ: Princeton University Press.
- Davies JS and Standing A (2023) Beyond the consolations of professionalism: Resisting alienation at the neoliberal university. *Soundings* 84(84-85): 13–28.
- De Madariaga IS (2013) From women in transport to gender in transport: Challenging conceptual frameworks for improved policymaking. *Journal of International Affairs* 67(1): 43–65.
- Dencik L, Hintz A and Cable J (2017) Towards data justice. *Data Politics* 167. doi:10.14763/2022.1.1615%0A.
- de Vos A, Preiser R, Masterson V (2022) Participatory data collection. In: Biggs R, de Vos A, Preiser R, et al. (eds) *The Routledge Handbook of Research Methods for Social-Ecological Systems*. London, UK: Routledge, 199. doi:10.4324/9781003021339.
- Dignazio C and Klein LF (2020) *Data Feminism*. Cambridge, Massachusetts: MIT press. doi:10.7551/mitpress/11805.001.0001.
- Eckhardt J, Kaletka C, Krüger D, et al. (2021) Ecosystems of co-creation. *Frontiers in Sociology* 6: 642289. doi:10.3389/fsoc.2021.642289.
- Escobar A (2018) *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Durham, NC: Duke University Press. doi:10.1215/9780822371816.
- European Parliament (2022, June 28) The dutch childcare benefit scandal, institutional racism and algorithms. European Parliament. Available at: <https://www.europarl.europa.eu/doceo/document/O-9-2022-000028EN.html> (accessed 9 August 2024).
- Fileborn B and Trott V (2022) "It ain't a compliment": Feminist data visualisation and digital street harassment advocacy. *Convergence* 28(1): 127–149.
- Franklin RS, Delmelle EC, Andris C, et al. (2023) Making space in geographical analysis. *Geographical Analysis* 55(2): 325–341.
- Gitelman L and Jackson V (2013) Introduction. In: Gitelman L (ed) *Raw Data is an Oxy-Moron*. Cambridge, Massachusetts: The MIT Press, 1–14.
- Goncalves J, Narendra N and Verma T (2025a) Everything about climate change is disproportionate: Principles for spatial justice in urban climate action. *Geo: Geography and Environment* 12(2): e70024.
- Goncalves J, Slingerland G, Garcia Alvarez M, et al. (2025b). Citizen Voices for Just Cities: Rethinking Participation, Technology and Knowledge Production in Urban Research, Education and Practice. Zenodo. 10.5281/zenodo.17396521
- Goncalves JE, Ioannou I and Verma T (2024) No one-size-fits-all: Multi-actor perspectives on public participation and digital participatory platforms. *Philosophical Transactions A* 382(2285): 20240111.
- Guyan K (2022) *Queer Data: Using Gender, Sex and Sexuality Data for Action*. Bloomsbury Academic, 27–48.
- Hadwick D and Lan S (2021) Lessons to be learned from the Dutch childcare allowance scandal: A comparative review of algorithmic governance by tax administrations in The Netherlands, France and Germany. *World Tax Journal* 13(4): 609–645.
- Hamouchene H (2023) *Dismantling Green Colonialism: Energy and Climate Justice in the Arab Regio*. London, UK: Pluto Press.
- Hand DJ (2018) Aspects of data ethics in a changing world: Where are we now? *Big Data* 6(3): 176–190.
- Hansson S and Polk M (2018) Assessing the impact of transdisciplinary research: The usefulness of relevance, credibility, and legitimacy for understanding the link between process and impact. *Research Evaluation* 27(2): 132–144.
- Hao K (2019) Training a single AI model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review* 75: 103.
- Hao K and Heidi S (n.d.) South Africa's private surveillance machine is fueling a digital apartheid — MIT Technology Review. MIT Technology Review.
- Hao K and Hernández AP (n.d.) How the AI industry profits from catastrophe. MIT Technology Review.
- Harding S (2012) Rethinking standpoint epistemology: What is "strong objectivity"? In: Alcoff L and Potter E (eds) *Feminist Epistemologies*. New York, NY: Routledge, 49–82.
- Harrington C, Erete S and Piper AM (2019) Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human Computer Interaction* 3(CSCW): 1–25.
- Hendricks MD, Meyer MA and Wilson SM (2022) Moving up the ladder in rising waters: Community science in infrastructure and hazard mitigation planning as a pathway to community control and flood disaster resilience. *Citizen Science: Theory and Practice* 7(1). doi:10.5334/cstp.462.
- Hickel J and Sullivan D (2023) Capitalism, global poverty, and the case for democratic socialism. *Monthly Review* 75(3): 99–113.
- Hoffmann AL (2021) Terms of inclusion: Data, discourse, violence. *New Media & Society* 23(12): 3539–3556.
- Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 2053951716674238.
- Jamieson MK, Govaart GH and Pownall M (2023) Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass* 17(4): e12735.

- Kallis G (2019) *Limits. Why Malthus was Wrong and Why Environmentalists Should Care*. Stanford, CA: Stanford University Press.
- Kandt J and Batty M (2021) Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities* 109: 102992.
- Kearns B (2012) Down for the count: Overcoming the census bureau's neglect of the homeless. *Stan. JCR & CL* 8: 155.
- Keßler C and McKenzie G (2018) A geoprivacy manifesto. *Transactions in GIS* 22(1): 3–19.
- Kitchin R (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.
- Kitchin R (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 2053951714528481.
- Kitchin R (2019) Toward a genuinely humanizing smart urbanism. In: *The Right to the Smart City*. Leeds, UK: Emerald Publishing Limited, 193–204.
- Kitchin R and Lauriault TP (2014, July) Towards critical data studies: Charting and unpacking data assemblages and their work (Working Paper No. 2). <https://ssrn.com/abstract=2474112>
- Kraaijvanger CW, Verma T, Doorn N, et al. (2023) Does the sun shine for all? Revealing socio-spatial inequalities in the transition to solar energy in The Hague, The Netherlands. *Energy Research & Social Science* 104: 103245.
- Kusek WA and Smiley SL (2014) Navigating the city: Gender and positionality in cultural geography research. *Journal of Cultural Geography* 31(2): 152–165.
- Laing AF (2021) Decolonising pedagogies in undergraduate geography: Student perspectives on a decolonial movements module. *Journal of Geography in Higher Education* 45(1): 1–19.
- Lee VR, Pimentel DR, Bhargava R, et al. (2022) Taking data feminism to school: A synthesis and review of pre-collegiate data science education projects. *British Journal of Educational Technology* 53(5): 1096–1113.
- Liu F, Rahwan T and AlShebli B (2023) Non-white scientists appear on fewer editorial boards, spend more time under review, and receive fewer citations. *Proceedings of the National Academy of Sciences* 120(13): e2215324120.
- Lodge JM, Alhadad SS, Lewis MJ, et al. (2017) Inferring learning from big data: The importance of a transdisciplinary and multi-dimensional approach. *Technology, Knowledge and Learning* 22: 385–400.
- Maharawal MM and McElroy E (2018) The anti-eviction mapping project: Counter mapping and oral history toward bay area housing justice. *Annals of the American Association of Geographers* 108(2): 380–389.
- Mattern S (2020) Post-it note city. *Places Journal*. doi:10.22269/200211.
- Mohamed S, Png M and Isaac W (2020) Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33(4): 659–684.
- Nash K, Trott V and Allen W (2022) The politics of data visualisation and policy making. *Convergence* 28(1): 3–12.
- Nelson T, Goodchild M and Wright D (2022) Accelerating ethics, empathy, and equity in geographic information science. *Proceedings of the National Academy of Sciences* 119(19): e2119967119.
- Nijman J and Wei YD (2020) Urban inequalities in the 21st century economy. *Applied Geography* 117: 102188.
- Noble S (2016) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press. doi:10.18574/nyu/9781479833641.001.0001.
- Noxolo P (2017) Decolonial theory in a time of the re-colonisation of UK research. *Transactions of the Institute of British Geographers* 42(3): 342–344.
- O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Crown.
- O'Rourke D and Lollo N (2015) Transforming consumption: From decoupling, to behavior change, to system changes for sustainable consumption. *Annual Review of Environment and Resources* 40(1): 233–259.
- Oztig LI (2023) Big data-mediated repression: A novel form of preemptive repression in China's Xinjiang region. *Contemporary Politics* 29(5): 576–597.
- Paris BS, Cath C and West SM (2024) Radical infrastructure: Building beyond the failures of past imaginaries for networked communication. *New Media & Society* 26(11): 6366–6393.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, Massachusetts: Harvard University Press. doi:10.4159/harvard.9780674736061.
- Pasquinelli M (2020) How a machine learns and fails—a grammar of error for artificial intelligence spheres.
- Pearce BJ (2024) Changing hearts instead of changing minds – another take on climate action. *Global Social Challenges Journal* 3(3): 1–12. doi:10.1332/27523349Y2024D000000006.
- Potts K and Brown L (2005) Becoming an anti-oppressive researcher. In: Brown L and Strega S (eds) *Research as Resistance: Critical, Indigenous and Anti-Oppressive Approaches*. Toronto, ON, Canada: Canadian scholars' Press, 243–266.
- Radcliffe SA (2017) Decolonising geographical knowledges. *Transactions of the Institute of British Geographers* 42(3): 329–333.
- Richards NM and King JH (2013) Three paradoxes of big data. *Stanford Law Review Online* 66: 41.
- Ricker B (2017) Reflexivity, positionality and rigor in the context of big data research. In: Thatcher J, Shears A and Eckert J (eds) *Thinking big Data in Geography: New Regimes, new Research*. Lincoln, Nebraska: University of Nebraska Press, 96–118. <https://ssrn.com/abstract=2911652>.
- Ricker B, Kraak M-J and Engelhardt Y (2020) The power of visualization choices: Different images of patterns in space.
- Robinson J and Tansey J (2006) Co-production, emergent properties and strong interactive social research: The Georgia basin futures project. *Science and Public Policy* 33(2): 151–160.
- Rose G (1997) Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography* 21(3): 305–320.
- Rose G, Raghuram P, Watson S, et al. (2021) Platform urbanism, smartphone applications and valuing data in a smart city. *Transactions of the Institute of British Geographers* 46(1): 59–72.

- Rose-Redwood R, Blu Barnd N, Lucchesi AH, Dias S and Patrick W (2020) Decolonizing the map: Recentring indigenous mappings. *Cartographica: The International Journal for Geographic Information and Geovisualization* 55(3): 151–162.
- Rowe F (2021) Big Data and Human Geography. doi:10.31235/osf.io/phz3e.
- Sadowski J (2021) Who owns the future city? Phases of technological urbanism and shifts in sovereignty. *Urban Studies* 58(8): 1732–1744.
- Sadowski J and Bendor R (2019) Selling smartness: Corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values* 44(3): 540–563.
- Saltz JS and Dewar N (2019) Data science ethical considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology* 21: 197–208. <https://doi.org/10.1007/s10676-019-09502-5>
- Schneider M, Brisson D and Burnes D (2016) Do we really know how many are homeless?: An analysis of the point-in-time homelessness count. *Families in Society* 97(4): 321–329.
- Sheppard E (2015) Thinking geographically: Globalizing capitalism and beyond. *Annals of the Association of American Geographers* 105(6): 1113–1134.
- Singh R (2023) The decolonial turn is on the road to contingency. Information. *Communication & Society* 26(4): 803–806.
- Smith L (2021) *Decolonizing Methodologies: Research and Indigenous Peoples*. London, UK: Zed Books. doi:10.5040/9781350225282.
- Tacheva Z (2022) Taking a critical look at the critical turn in data science: From “data feminism” to transnational feminist data science. *Big Data & Society* 9(2): 20539517221112901.
- Takacs D (2003) How does your positionality bias your epistemology? *Thought & Action*: 27–38.
- Talwar S, Wiek A and Robinson J (2011) User engagement in sustainability research. *Science and Public Policy* 38(5): 379–390.
- Thambinathan V and Kinsella EA (2021) Decolonizing methodologies in qualitative research: Creating spaces for transformative praxis. *International Journal of Qualitative Methods* 20: 16094069211014766.
- Thatcher J, O’Sullivan D and Mahmoudi D (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space* 34(6): 990–1006.
- Tschakert P, Tuana N, Westskog H, et al. (2016) Tchange: The role of values and visioning in transformation science. *Current Opinion in Environmental Sustainability* 20: 21–25.
- Ulmer JB (2020) Pivots and pirouettes: Carefully turning traditions. *Qualitative Inquiry* 26(5): 454–457.
- Van Schie G, Smit A and López Coombs N (2020) Racing through the Dutch governmental data assemblage: A post-colonial data studies approach. *Global Perspectives* 1(1): 12779.
- Vybornova A and Verma T (2025) You don’t have to live next to me: Towards demobilizing individualistic bias in computational approaches to urban segregation. arXiv preprint arXiv:2505.01830.
- Watharow A and Wayland S (2022) Making qualitative research inclusive: Methodological insights in disability research. *International Journal of Qualitative Methods* 21: 16094069221095316.
- Wilkinson LC and Wilkinson MD (2023) Value for money and the commodification of higher education: Frontline narratives. *Teaching in Higher Education* 28(2): 406–422.
- Wiltshire D and Alvanides S (2022) Ensuring the ethical use of big data: Lessons from secure data access. *Heliyon* 8(2). doi:10.1016/j.heliyon.2022.e08981.
- Woroniecki S, Krüger R, Rau A-L, et al. (2019) The framing of power in climate change adaptation research. *WIREs Climate Change* 10(6): e617.
- Young E, Wajcman J and Sprejer L (2023) Mind the gender gap: Inequalities in the emergent professions of artificial intelligence (AI) and data science. *New Technology, Work and Employment* 38(3): 391–414.