Critical Data Science

A student handbook

led by Laura van Geene

EPA122A - Spatial Data Science
MSc Engineering & Policy Analysis
Delft University of Technology
November 2023

Table of Contents

1. Introduction

The aim of this handbook is to help you develop a critical approach towards data science. This handbook encourages you to ask critical questions during your own data science practice, but also while reviewing the work of others. We begin by explaining the necessity of a critical view in data science, which is followed by a theoretical and practical guide on how to make data science work more critical.

1.1. Data Science, Subjectivity & Power

Objectivity is *"to aspire to knowledge that bears no trace of the knower - knowledge unmarked by prejudice or skill, fantasy or judgement, wishing or striving"* (Daston & Gallison, 2021, p. 17). Objectivity relies on the existence of an *objective reality,* a reality which *exists as it is independent of any conscious awareness* (Mulder, n.d). The scientific method, with evolving practices and methods, has been proposed for the *objective* pursuit of knowledge in that *objective reality*. Through this process, knowledge could be built upon scientific consensus. This way of thinking is especially prevalent in natural sciences such as physics and chemistry.

However, there is no consensus that this pure objectivity pursued in physics or chemistry is achievable at all. For decades, there has been a discussion if objective sciences, such as mathematics, are not just subjective attempts by humans to understand the world that have been so widely accepted that they seem purely objective (Goodman, 1979). Subjectivity becomes an even bigger theme as soon as science intersects with social experiences, which are complex, dynamic and fluid (Boyd, 2021b). Especially in urban contexts, where there is an intersection of a variety of histories, interests, biographies and mosaics of social structures that are invisibly and visibly permeating human and material activity. In contrast to the neat, measurable reality of the atom in physics or the molecule in chemistry, an urban spatial data scientist works with dynamic realities that are all fueled by subjective experiences.

But how does this subjectivity work through in the research process?

First, the relationship between the object of study and the researcher is inherently entangled. In data science, the objects of study are directly or indirectly generated and influenced by human behaviour, including their own. The line becomes especially blurry during interpretation of results, where different stakeholders will have different interpretations. Therefore, both the object of study and the researchers find themselves in a subjective plain, their understanding defined by their lived experiences (Takacs, 2003).

Second, data science plays an essential role in identifying patterns, trends and relationships within the data to provide evidence for causality. However, *finding evidence of causality is not the same as explaining the causality*. Data is rarely self-explanatory and requires participation and explanations from discipline experts, stakeholders, and theoretical frameworks based on previous studies (Gitelman & Jackson, 2013). The experts, stakeholders, and frameworks considered can differ per study and, therefore, the data scientist will collect, describe, interpret data according to frameworks and theories that have deemed to be relevant for that specific case but that are not universally applicable. Additionally, the causal deductions require triangulation of different perspectives, datasets, and theories (Kandt and Batty, 2021).

Third, the practice of data science is complicated by historical and existing power relations. For example, data has been historically used for colonisation, but this has more recently evolved in ‘data colonialism’ which combines historical extractive traditions with current day data science (Couldry & Mejias, 2019). Likewise, search algorithms can systematically enhance the visibility of some while distorting views of others. In this way, these digital technologies can become oppressive mechanisms in today’s increasingly digitised world (Noble, 2018). However, insights from (geospatial) data, AI, algorithms and machine learning are considered beyond human in many senses and are used to legitimise policies and decision making (Gitelman & Jackson, 2013). This is particularly relevant in Spatial Urban Data Science as spatial analysis tools have *“been critiqued particularly for affording scientist a disembodied view of the world often referred to as the* ***‘god trick’*** *referring to the ability to see the whole world while being distant from it*” (Ricker, 2017 p. 106).

Understanding that data scientists, our methods, analyses, and interpretations are not distant from the subjective reality around us, is at the heart of critical data science. Data science is not objective, not of any fault of its own, but because of the realities it deals with and takes place in. Harding (2013) argues that the objectivity of your study actually improves when it is considered within its social context and the pre-existing assumptions, beliefs, and prejudices that are layered beneath it. In this way, more knowledge systems are integrated in your research, making your research valid to a wider array of stakeholders (Harding, 2013). Exploring and acknowledging the subjectivities beneath your research is what makes you a ‘critical’ data scientist (Iliadis & Russo, 2016). With this handbook, we give you a first introduction to what a critical data science process could look like. We expect that the handbook will offer ideas to better navigate and participate in the dynamic, complex, and fluid reality in which spatial data science takes place. By properly accounting for this, we believe the objectivity, quality and, equally important, the social impact of your data science research will improve.

2. Critical Data Science

As we have explained above, science and the data science process are never completely objective. This also means that during the evolution of the data science field, certain ideologies, histories and philosophies have permeated the way in which researchers think about data and the way in which data science is practised (Iliadis & Russo, 2016). We have already highlighted three ways in which this is reflected in the research process. Subjectivity and the permeation of these ideologies, histories, and philosophies is not necessarily a bad thing. However, the danger lies in the fact that these models and data can define their own (distorted) reality, which the models and their users in turn use to justify their results. In this way, many models have prejudices, bias and misunderstandings encoded in them (O'Neil, 2016). Data science must be viewed with a critical perspective that scrutinises these implicit perspectives (Iliadis & Russo, 2016). With this guide, we provide support to implement *Critical Data Science* in practice.

There are multiple visions of critical data science processes. According to Iliadis & 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 data science process. Dalton & Thatcher (2014) envision a process that considers the historical developments to lead the realization of 'big data', who is in control of the data, the motivations that drive the research, the subjects of the data and their knowledges, the role of data in the production of place and space, and the final use of the data. Kitchin & Laurialt (2014) suggest to go beyond the vision of Dalton & Thatcher (2014) and also consider the influence of the political economy, financing, and the subjectivities and communities of the people involved.

So, knowing that important questions as above can be included, how does *Critical Data Science* help you handle the process of making your scholarship and practise critical? For this, we have defined four theoretical pillars that 'support' a critical data science process. Each pillar is a critical theory that can be used to investigate several of the critical aspects mentioned by Iliadis & Russo (2016), Dalton & Thatcher (2014) and Kitchin & Laurialt (2014).

3.1 Decoloniality

The first theoretical pillar of Critical Data Science is decoloniality. Within the data sciences, decolonial thinking questions how existing power dynamics and histories of oppression are reflected in the data and data science process (Couldry & Mejias, 2018). Questioning power relationships plays an important role in *Critical Data Science*, as with data there is the risk that it is objectified 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 exploitation and can be subject to "digital coloniality" (Mohammed et al., 2020). More specifically, the idea of data colonialism goes back to one argument: digital data infrastructure is designed to extract, circulate, and analyse data without having obtained the informed consent of those that produce the data. This relationship grows more oppressive as producers have a lesser say about the data which increasingly becomes equinamous with capital (Singh, 2021). The core factor in this, according to data colonialism scholars, is that the power is often held by large, also politically powerful Western corporations (Couldry & Mejias, 2018).

Decolonial thinking within data science should not be seen as a tool to problematize, but rather it is an invitation to critically examine and critique the politics of race and colonial ways of thinking that are visible in today's technology that exclude, limit, or discredit ways of thinking that are beyond the Western standard (Adams, 2021). Decolonial thinking is local; it is critically aware of the social mechanisms that reproduce racism and discrimination; it is critically aware of the power dynamics in today's world and their influence on the production of "good" knowledge (Adams, 2021).

3.2 Intersectionality

The second theory integrated in the Critical Data Science approach is intersectionality. This pillar reflects how data is a reflection of multiple overlapping realities experienced by people through life, and thus significant to understand in the process of transforming and analysing data (Lee et al., 2022). Misrepresentation of particular groups is a problem in contemporary society and in the data sciences. Within this context, intersectionality investigates how the representation of people with a marginalized position in society influences the data and the wider field of data science (Lee et al., 2022). Data and its analytical models are often created by a small, under-representative group of men (D'Ignazio & Klein, 2020; Tacheva, 2022), an expression of one sex and often one race influencing how the rest of us live and adapt. As a result of this disproportionate power to influence decisions in the world, the perspectives of other societal groups can be excluded from mainstream data science (D'Ignazio & Klein, 2020). Integrating these perspectives is vital to prevent unjust outcomes. Intersectional thought revolves around investigating structural oppression of people whose social identities, such as race, gender, ethnicity, and sex leads them to be marginalized in multiple ways, such as, but not only limited to, women of colour (Lee et al., 2022).

An example of how this works in data science comes from Bowleg & Bauer (2016) who examined the results of a study on the effects of sex (male vs. female) and race (Black vs. White) on referral for specialist care in a hospital. When comparing these effects without looking at any potential interaction effects of sex and race, the data showed that White people and men had a higher chance of referral. However, when explicit attention was paid to potential interaction between sex and race, it was found that Black women had the lowest chance of getting a referral. Moreover, no other group's results were influenced that much by the interaction effect between sex and race (Bowleg & Bauer, 2016). Thus, intersectionality in data science is not only about bringing in the perspectives of women into data and analysis, but also about explicitly considering that there are groups that are faced with multiple structural inequalities and what additional influence this can have on their representation in and the analysis of the data.

3.3 Shared Knowledge Creation

The third theoretical pillar is Shared Knowledge Creation. Data science is about collecting, modifying, analysing, and visualising data to describe a certain situation. However, the choices of who to involve in the data science process can greatly influence the results and these choices need to be questioned thoroughly before making life-altering decisions for communities. For example, recently the Dutch authorities wrongly accused an estimated 26,000 parents of making fraudulent *childcare* benefit claims, requiring them to pay back the allowances they had received in their entirety. In many cases, this sum amounted to tens of thousands of euros, driving families into severe financial hardship. Missing the perspectives of the individuals that are living the actual situation that you are researching can make your data show a narrative that is different than that of the people on the ground (Lee et al., 2022). This becomes increasingly problematic when these people have a history of being excluded from the dominant narrative

(Smith, 2021). Both Dalton & Thatcher (2014) and Kitchin & Laurialt (2014) emphasise the importance of including the knowledge systems of the people involved in critical data science.

Participatory forms of research try to establish ways of co-learning between all parties involved to create a shared understanding and overlapping interests (de Vos et al., 2021). This requires a reflective way of thinking of the data scientists, a willingness to see from another point of view, and a recognition that different knowledge systems are all valid and can bring new insights to research (de Vos et al., 2021). This way of thinking relates to the concept of intersubjectivity: the idea that each person's reality is by definition unique and subjective and that, therefore, it is impossible to define one objective view of the world or for one person to describe how someone else views the world (Munroe, 2019). Shared actions, such joint participation in research, can help diversify one's perception of reality, but can also be an excellent way to unify the perspectives of different people and create common ground (Matusov, 1996).

Guyan (2022) gives an example of exclusive data collection in his book about the history and presence of queer data in the UK. Guyan (2022) argues that the underrepresentation of LGBTQ-people in data is due to *"a failure to ask inclusive questions about gender, sex and sexuality"* (p.48) that reflects assumptions of heteronormativity beneath data collection methods. The failure of properly including the perspectives of these societal groups has led them to be missing in data, but also erasing their voices in wide societal discussions (Guyan, 2022).

3.4 Reflexivity

The final pillar of Critical Data Science is reflexivity. It questions the position of the researcher themselves and encourages critical reflection on all the points above, but also on the researcher's own biases throughout the research process. Reflexive research sees the process as an iterative and continuous process of knowledge production (Boyd, 2021b) in which the researcher needs to acknowledge that they are not distant and independent from the data that they're working with (Ricker, 2017). O'Neil (2016) warned of the possibilities of encoding one's biases and prejudices in models. Reflexivity can help the researcher not only to be critical of the data science process that they are involved in, but also of their own place and their influence on the research. This means asking critical questions about, for example, the ethical considerations around access and ownership of collected data, but also potential harms that can arise from the aggregation, use, and linking of that data (Saltz & Dewar, 2019). For this very reason, the statical bureau of The Netherlands (CBS) does not freely allow researchers to access microdata of Dutch residents in its entirety. Achieving a balance between the risks and gains of using data is hard to achieve, since any action benefits some, while also harming others with often unexpected consequences (Hand, 2018). By taking a reflexive approach to data science, the researcher can flesh out the underlying motivation of the research and allow the researcher to prevent erasure, harm, and silencing of underrepresented voices as much as possible (Ricker, 2018).

Note:

While the four theoretical pillars discussed above are seen as the building blocks of the Critical Data Science approach suggested in this handbook, two important footnotes need to be made. First, our version of Critical Data Science is not a universal definition of how a critical approach in data science should look. Second, the selection of theoretical pillars discussed above is also not an exhaustive selection of theories that can be used in critical data science. Formulating universal definitions is in itself problematic, as this dismisses the experience of those that do not recognize themselves in the definition (Filebron & Trott, 2021), making the definition inherently exclusive. Critical data science is about critically questioning your own logic and assumptions during the data science process, which can and should preferably go beyond the four theoretical pillars covered above.

3. Critical Data Science in practice

3.1. Key questions of Critical Data Science

Each of the pillars introduced in Section 2 helps ask certain critical questions about the data science process. The pillars are complimentary, but also interrelated. To provide you with a strategy that you can apply in practice, we have translated the broad theory behind the approach in five key questions relating to inclusion, inequality, positionality[1], participation and power. The key questions will be shortly explained below.

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

An element of inclusivity comes back in all four pillars. Decolonial thinking encourages the researcher to think about inclusivity by considering whose framing of the research topic can (not) be seen in the data, especially in relation with existing power dynamics (Couldry & Mejias, 2018). Intersectionality in the data sciences revolves around considering how marginalization of certain communities influences their representation in the data (Lee et al., 2022). One of the main ideas related to shared knowledge creation is that the more diverse the research sample is, the better the answer will describe reality through the combination of individual views of the research topic (Matusov, 1996). A reflexive research approach encourages the researcher to think about how their own choices in the research process influence the inclusivity of the sample (Boyd, 2021a).

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

Although closely related to the key question about inclusion, in answering this question you should consider how data science methods can be used to contribute to existing inequalities. Not just within your research, but also in the larger data science field (Boyd, 2021a). Exploration of structural inequalities and questions of marginalization are key in the field of intersectionality (Boyd, 2021a), but also play an important role in decolonial thinking, where it has to do with larger societal processes such as privilege, oppression, and historical injustices (Adams, 2021). When it comes to shared knowledge creation, it is important to consider what stakeholders and knowledge has been structurally left out of research in the past (Lee et al., 2022). Again,

[1] Inclusion, Inequality, and Reflexivity are also the three core dimensions in the Quantitative Intersectional Data Science approach proposed by Boyd (2021a)

reflexivity comes through in questioning your own choices on the research and in turn structural inequalities (Boyd, 2021a).

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

This question concerns the people involved in the research. Having a research team with a diverse set of perspectives (as CUSP proudly is) is as important as having a diverse data set. Participation is important in decolonial research to ascertain informed consent during data collection since the 'data is only borrowed from the producer' (Harrington et al., 2021; Nash et al., 2022). Intersectionality encourages the involvement of people with multiple different social identities to make sure that interplay between different marginalisations is also accounted for (Boyd, 2021a). The key for both these theories is that the data science process cannot reinforce unfair systems by excluding certain people/perspectives. Shared knowledge creation per definition requires a diverse group of researchers. Reflexivity also encourages the researcher to reflect about the influence of including or excluding certain perspectives in their research (Boyd, 2021a).

How does the data reflect existing power dynamics? (Power)

Questions of power underlie all the four pillars and therefore are one of the key questions of the critical data science approach. All pillars directly or indirectly encourage the researcher to question the underlying power dynamics of the data science process. Decoloniality has to do with histories of injustice and oppression (Couldry & Mejias, 2018) and with modern practices of data colonialism (Singh, 2021). Intersectionality concerns the mechanisms that cause people with multiple different social identities to be more underrepresented than people that are (tragically) only marginalised in *one* way (D'Ignazio & Klein, 2020, Lee et al., 2022). Shared knowledge creation requires a willingness to recognize that all viewpoints are valid (de Vos et al., 2021). However, the power in knowledge creation processes often lies with the privileged and the wealthy (Potts & Brown, 2005) and researchers need to be critical of this. When it comes to reflexivity, it is important to consider if you as a researcher stand to gain from the research and how this influences your decisions in the research process (Boyd, 2021a).

What is your own positionality with the research? (Positionality)

A truly critical data science process requires not only consideration of all the above, but also critical reflection during the process to make sure that you, as a researcher, are being as critical as possible. Knowledge production is a constant and iterative process to which the researcher and their perspectives are highly connected (Ricker, 2017; Boyd, 2021b). To prevent the researcher from inadvertently projecting their biases on the research, reflection is required for each step of the data science process, but also the entire process as a whole. Examples of these questions could be: Why are you the best person to research this? Whose voices are being amplified and who is being silenced? Why are you involving certain people and why not others?

3.2. Step-by-step approach

To provide you with further support to set-up your own critical data science process, we have integrated the questions into the standard six-step data science process. Table 1 outlines some more considerations for each of the key questions of Critical Data Science. You can use this table to learn how to conduct the data science process (practice makes good), while also use it for day-to-day assignment questions on reflection. Use it as a guide, a friend or a manifesto!

Note:
The questions below can be answered in any particular order. We do not assign a certain ranking or order. Rankings themselves are human constructs and decisions behind ranking are inherently tied to questions of power and researcher bias (Lee et al., 2022).

Table 1: Detailed outline of Critical Data Science Process.

Data Science Process	**Inclusion** *Who is (not) included in the data?*	**Inequality** *What role does inequality play in data science methods?*	**Participation** *Who is (not) involved in the data science process?*	**Power** *How does the data reflect existing power dynamics?*	**Positionality** *What is your own positionality with the research?*
Focus of Analysis *Theories, processes & stakeholders that drive the analysis*	Investigation of exclusive practices of past and present relating to the research focus, and how these affect the diversity of the people represented in the data (Boyd, 2021a; Lee et al., 2022).	What tools can be reliably used to explore the research topic? Research the limitations of the methods, particularly their influence to structural inequalities (Boyd, 2021a).	Use a participatory modelling approach and include stakeholders that may not have otherwise been involved in the design of the research process and discuss how to include topics/perspectives that are not commonly researched (Lee et al., 2022). Discuss the possibility of multiple framings of the research topic (Delbosc, 2023).	Investigation of where and with whom power was distributed in the situations referenced with the data (Lee et al., 2022). Also investigate potential histories of injustices and oppression of the sampling population (Harrington et al., 2021).	Critically reflect on your own position to the research (Boyd, 2021a): 1. Why are you doing research about this specific topic? Why are you specifically involved in this research? What makes you suitable for this research? 2. What is the story that you are trying to tell with this research? Consider biases: do you already have ideas about how this story should go? 3. Is there potential that you cause harm or erasure with your research about this topic?

Collect & Combine Data *Contexts and power relations that lie beneath data collection & creation of data sets*	Who is at the center of this data collection process? Pay particular attention to data collection with/from neglected and historically excluded groups (Boyd, 2021a; Lee et al., 2022).	Are structural inequalities reflected in the data collection process? (Boyd, 2021a). For example, is the method suggesting the use of proxies or aggregated data? How does the sampling method influence this? The latter includes considerations of past developments that have caused these structural inequalities (Harrington et al., 2021). Are there alternative data sources that might improve visibility of certain groups?	Are all relevant stakeholders included in the data collection? Obtain data as close to the source as possible. Actively seek the help of those that are living the experience/topic of interest (Lee et al., 2022).	Acknowledge that the data does not belong to the collector, but is borrowed from the people that produce the data (Harrington et al., 2021; Nash et al., 2022). Have you obtained informed consent from those that produce the data by explaining the need, goal, and use of their data?	Critically reflect on your data collection choices (Boyd, 2021a): 1. Why are you collecting data from these specific sources? Why are you using this specific sampling method? 2. Why are you including these specific stakeholders? 3. How are you silencing certain voices by excluding them in the data collection and your data set? And why? 4. How are you amplifying certain voices by including them in the data collection and your data set? And why?
Transform Data *Completeness, Missing data, Consistency, Pluralism & Accuracy of collected data*	Do not only consider what data is missing from the dataset, but also whose data is missing (diversity in variables, but also diversity in sources).	Are you erasing or magnifying someone's perspective by cleaning the data (aggregating, replacing missing value, or slicing)? (Boyd, 2021a). Did the (joint) distribution of the data change after cleaning? If so, explore the impacts of a different cleaning approach.	Ensure transparency of data cleaning choices. Collaboratively discuss the impact of these decisions and alternative ways of transforming the data.	Are the data cleaning techniques (normalization, replacement of missing values) reinforcing a dominant framing of what the data should show? (Boyd, 2021a).	Critically reflect on your data cleaning choices? 1. Why are you using these specific data cleaning methods? 2. How are you silencing certain voices in your data cleaning process? And why? 3. How are you amplifying certain voices in your data cleaning process? And why?

Analyse Data *Representation of all, what is dominant, what is uncertain*	Taking all the above into account, who is (not) represented in your analysis? How does the inclusivity of your analysis influence your results?	How sensitive is your data analysis method to biased data? Consider and test for structural differences and similarities within your sampled population (Boyd, 2021a).	Place emphasis on collaborative analysis by encouraging transparency in sharing intermediate results. Create a feedback back to those living the experience about preliminary results (including decision about visualisation).	Analyse the data as close to the setting that you collected it from (Adams et al., 2022), meaning that your analysis includes considerations of the context in which the data is situated (e.g., oppressive practices, structural inequalities, etc.) (Harrington et al., 2021; Adams et al., 2022).	Acknowledge that you are not distant from the data that you are analyzing (Ricker, 2021): 1. How representative is your analysis? 2. Why are you including certain stakeholders in your analysis? 3. Can you find any personal biases towards the analysis? 4. How are you silencing certain voices in your analysis? And why? 5. How are you amplifying certain voices in your analysis? And why?

Interpret & Visualise Data *Consideration of all the above to formulate answer to research question*	Are your findings interpreted in a way that considers the diversity of stakeholders involved? Whose view of the problem/solution to the research problem is being represented with your visualisation? And whose view is being excluded?	Link your choices for interpretation and visualisation of the results to questions of unequal opportunities, marginalization, and vulnerability (Boyd, 2021a).	Acknowledge the existence of different interpretations of the data (Lee et al., 2022; Delbosc, 2023). Combine these perspectives, including the perspectives of those living the experience, in a shared interpretation of the research topic.	Is your interpretation reinforcing a dominant framing of the research topic? Pay explicit attention to past and present power mechanisms that might be of influence here. Is your chosen visualisation reinforcing a certain, subjective idea of what the data should show? (Hill, 2017). This relates both to the chosen visualisation (e.g., maps or diagrams), but also the choices of what to visualise.	Acknowledge that you are not distant from the data as you are interpreting it (Ricker, 2021): 1. Why are you interpreting and visualising your results in this way? 2. Can you find any personal biases towards the interpretation? 3. How are you silencing certain voices in your interpretation? And why? 4. How are you amplifying certain voices in your interpretation? And why?
Communicate Findings *Transparency and accessibility of the results*	Consider the explainability and transparency of your findings (Nash et al, 2022): are they accessible for the general public, but also for those that will be the most affected?	Critical and inclusive research does not automatically decrease injustices (Guyan, 2022): Is there potential for your findings to be used in such a way that they enhance structural inequalities or cause other harms?	Consider the explainability of your visualisations and findings (Nash et al, 2022): is it agreed upon and understandable for all those involved?	Explore ways of using your findings to challenge unjust systems of power and empower those in disadvantaged positions (D'Ignazio & Klein, 2020; Lee et al., 2022; Singh et al., 2021).	Reconsider your research goal and motivation: 1. Is it necessary for you to tell this story? 2. Are you doing the people at risk of harm justice by communicating your findings in this way? 3. Are you reinforcing a certain dominant/desired framing with your communicating? 4. Whose voice are you excluding/silencing when

					you communicate your findings? And how?

References

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46(1–2), 176–197. https://doi.org/10.1080/03080188.2020.1840225

Adams, G., Ratele, K., Suffla, S., Reddy, G., & The Readsura Decolonial Editorial Collective includes (in random order). (2022). Psychology as a site for decolonial analysis. *Journal of Social Issues, 78*(2), 255–277. https://doi.org/10.1111/josi.12524

Bowleg, L., & Bauer, G. R. (2016). Invited reflection. *Psychology of Women Quarterly*, *40*(3), 337–341. https://doi.org/10.1177/0361684316654282

Boyd, A. E. (2021a). Quantitative intersectional data (QUINTA): a #metoo case study [PhD thesis]. DePaul University.

Boyd, A. E. (2021b). Intersectionality and reflexivity—decolonizing methodologies for the data science process. *Patterns*, *2*(12), 100386. https://doi.org/10.1016/j.patter.2021.100386

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, *20*(4), 336-349. https://doi.org/10.1177/1527476418796632

Couldry, N., & Mejias, U. A. (2021). The decolonial turn in data and technology research: what is at stake and where is it heading? *Information, Communication & Society*, *26*(4), 786–802. https://doi.org/10.1080/1369118x.2021.1986102

Dalton, C., & 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*.

Daston, L., & Galison, P. (2021). *Objectivity*. Princeton University Press

Delbosc, A. (2022). There is no such thing as unbiased research – is there anything we can do about that? *Transport Reviews, 0*(0), 1–4. https://doi.org/10.1080/01441647.2022.2146939

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press. https://doi.org/10.7551/mitpress/11805.001.0001

Fileborn, B., & Trott, V. (2021). “It ain’t a compliment”: Feminist data Visualisation and Digital Street Harassment Advocacy. *Convergence, 28*(1), 127–149. https://doi.org/10.1177/13548565211045536

Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *Raw Data is an Oxymoron* (pp. 1-14). MIT Press.

Goodman, N. D. (1979). Mathematics as an objective science. *The American Mathematical Monthly, 86*(7), 540-551.

Guyan, K. (2022). *Queer data: Using gender, sex and sexuality data for action*. Bloomsbury Publishing. https://doi.org/10.5040/9781350230767

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big data, 6*(3), 176–190. https://doi.org/10.1089/big.2018.0083

Harding, S. (2013). Rethinking standpoint epistemology: What is “strong objectivity”?. In L. Alcoff, E. Potter (Eds.), *Feminist epistemologies* (pp. 49-82). Routledge.

Harrington, C., Erete, S., & Piper, A. M. (2019). Deconstructing Community-Based collaborative design. *Proceedings of the ACM on human-computer interaction, 3*(CSCW), 1–25. https://doi.org/10.1145/3359318

Hill, R. L. (2017). The political potential of numbers: data visualisation in the abortion debate. *Women, Gender & Research , 26*(1), 83-96. https://tidsskrift.dk/KKF/article/view/109789/159113

Iliadis, A., & Russo, F. (2016). Critical Data Studies: An Introduction. *Big Data & Society*, *3*(2), 205395171667423. https://doi.org/10.1177/2053951716674238

Kitchin, R. & Lauriault, T.P. (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. *The Programmable City Working Paper 2.*

Lee, V. R., Pimentel, D. R., Bhargava, R., & D'Ignazio, C. (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. https://doi.org/10.1111/bjet.13251

Matusov, E. (1996). Intersubjectivity without agreement. *Mind, Culture, and Activity*, *3*(1), 25–45. https://doi.org/10.1207/s15327884mca0301_4

Mohamed, S., Png, M., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Mulder, D. H. (n.d.). Objectivity. Retrieved October 23, 2023, from *Internet Encyclopedia of Philosophy*: https://iep.utm.edu/objectiv/

Munroe, P. T. (2019). Intersubjectivity. *The Blackwell Encyclopedia of Sociology*, 1–3. https://doi.org/10.1002/9781405165518.wbeosi067.pub2

Nash, K., Trott, V., & Allen, W. (2022). The politics of data visualisation and policy making. *Convergence, 28*(1), 3–12. https://doi.org/10.1177/13548565221079156

Noble, S. U. (2018). *Algorithms of oppression : how search engines reinforce racism*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

O'Neil. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy* (First edition.). Crown.

Potts, K., & Brown, L. (2005). Becoming an Anti-Oppressive Researcher. In L. Brown and S. Strega (Eds.). *Research as resistance critical, indigenous and anti-oppressive approaches* (pp. 243-266). Canadian Scholars' Press. ISBN: 9781423745150

Ricker, B. (2017). Reflexivity, Positionality and Rigor in the Context of Big Data Research. In J. Thatcher, A. Shears, and J. Eckert (Eds.), *Thinking Big Data in Geography: New Regimes, New Research* (pp. 96-118). University of Iowa Press. https://ssrn.com/abstract=2911652

Saltz, J. S., & Dewar, N. I. (2019). Data Science Ethical Considerations: A systematic literature review and proposed project framework. *Ethics and Information Technology, 21*(3), 197–208. https://doi.org/10.1007/s10676-019-09502-5

Singh, R. (2021). The decolonial turn is on the road to contingency. *Information, Communication & Society*, *26*(4), 803–806. https://doi.org/10.1080/1369118x.2021.1986104

Smith, L. T. (2021). *Decolonizing methodologies*. Zed Books. https://doi.org/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), 205395172211129. https://doi.org/10.1177/20539517221112901

Takacs, D. (2003). How Does Your Positionality Bias Your Epistemology?, *Thought & Action 27*. http://repository.uchastings.edu/faculty_scholarship/1264

de Vos, A., Preiser, R., & Masterson, V. A. (2021). Participatory data collection. In Biggs, R., de Vos, A., Preiser, R., Clements, H., Maciejewski, K., Schlüter, M. (Eds), *The Routledge Handbook of Research Methods for Social-Ecological Systems* (pp. 199-134).