



ARTICLE



<https://doi.org/10.1057/s41599-020-0495-3>

OPEN

Situated data analysis: a new method for analysing encoded power relationships in social media platforms and apps

Jill Walker Rettberg  ¹ 

This paper proposes situated data analysis as a new method for analysing social media platforms and digital apps. An analysis of the fitness tracking app Strava is used as a case study to develop and illustrate the method. Building upon Haraway's concept of situated knowledge and recent research on algorithmic bias, situated data analysis allows researchers to analyse how data is constructed, framed and processed for different audiences and purposes. Situated data analysis recognises that data is always partial and situated, and it gives scholars tools to analyse how it is situated, and what effects this may have. Situated data analysis examines *representations of data*, like data visualisations, which are meant for humans, and *operations with data*, which occur when personal or aggregate data is processed algorithmically by machines, for instance to predict behaviour patterns, adjust services or recommend content. The continuum between representational and operational uses of data is connected to different power relationships between platforms, users and society, ranging from normative disciplinary power and technologies of the self to *environmental power*, a concept that has begun to be developed in analyses of digital media as a power that is embedded in the environment, making certain actions easier or more difficult, and thus remaining external to the subject, in contrast to disciplinary power which is internalised. Situated data analysis can be applied to the aggregation, representation and operationalization of personal data in social media platforms like Facebook or YouTube, or by companies like Google or Amazon, and gives researchers more nuanced tools for analysing power relationships between companies, platforms and users.

¹University of Bergen, Bergen, Norway. email: jill.walker.rettberg@uib.no

Introduction

We are all data subjects, constantly sharing data. If you are reading this journal article on the web, the website is likely sending your data to third parties that will use it to tailor advertisements for you. Your phone sends your location to your service provider, which may anonymize it and sell it¹. Social media platforms analyse your likes and habits to try to make you view more ads or to manipulate you in other ways. Shoshana Zuboff calls this incessant collection and use of personal data “surveillance capitalism”, which she describes as the “new global architecture of behavioural modification” (2019, p. 2). Situated data analysis is a new method for analysing these architectures that emphasises how data is always situated, both in how it is constructed and how it is presented in different contexts. I analyse the manipulation and behavioural modification of platforms dealing with personal data as *environmentality*, a concept recently emerging as descriptive of digital media and developed by Jennifer Gabrys (2014), Erich Hörl (2018) and Mark Andrejevic (2019). Environmentality was originally suggested by Michel Foucault in a brief passage at the end of a lecture in 1979 (Foucault, 2008, pp. 259–260). Environmental power is encoded in our surroundings and the infrastructures and technologies we use. While disciplinary power relies on each individual internalising norms, environmental power alters the environment so as to promote certain behaviours and make undesired behaviours difficult or impossible.

Strava. The fitness tracking app Strava provides us with an interesting case because its users *want* it to record their data, and because this data is situated, analysed and displayed in a range of different ways. Strava is a fitness tracking app that uses location data from users’ phones to track outdoor exercise such as running or cycling. Users allow the app to track their location during a run, and afterwards, Strava displays a summary of how fast and how far the user ran, with a map showing the distance run, and statistics comparing performance to the user’s own previous performance and to that of other runners or cyclists who have run the same route. Strava produces personal data that aligns with the individual’s internalised norms, as well as aggregate data that is used to embed environmental power structures in cities and infrastructures, for instance by allowing city planners to position bike paths according to traffic measured by Strava. By analysing Strava, I develop a situated model of power at different levels of data collection and use that can also be applied to other platforms where the data is experienced very differently by the individual user to the way it is used at an aggregate level by companies and advertisers.

As several scholars of self-tracking have shown, users of self-tracking apps like Strava use them as what Foucault calls “technologies of the self”: as means for self-improvement (Foucault, 1988; Sanders, 2017; Depper and Howe, 2017; Kristensen and Ruckenstein, 2018). In the case of fitness apps like Strava, this self-improvement is desired and cultivated by the individual, and aligns with public health goals and societal ideals of fitness and beauty. In her analysis of the biopolitics of the quantified self movement, Btihaj Ajana notes that “biopower and biopolitics are not so much about explicit coercive discipline, but follow the neoliberal modality of free choice and the promise of reward” (Ajana, 2017, p. 6). Strava’s multiple levels of data, from the individual via the nearby to the global, thus become interfaces between individual self-discipline and normative societal pressures.

Each user’s data is also aggregated. This is made evident to users as Strava uses visualisations of the aggregate data in its marketing. Strava maintains a website with a Global Heatmap

showing where people run, cycle and swim as well as publishing annual reports with various graphs and charts showing how people in different cities tend to exercise at different times and so on. These data visualisations are intended for human audiences and are representations of the data that can be analysed using visual methods such as social semiotics (Aiello, 2020), but humans are not the only audience of the aggregated data (see Fig. 1). Publicly released, aggregated Strava data is processed algorithmically, by machines, for purposes such as city planning, public health research, data-based activism and developing navigation systems for automated vehicles and human cyclists or pedestrians. In these cases, the data is no longer primarily representational: it is *operational*. Operability is a concept I will return to later, but in brief it means that the data *does* something through computation rather than *showing* something to humans as in a visual representation (Hoel, 2018). Because Strava situates the same data in different ways, it allows us to analyse the different kinds of power relationships that are made possible by these different situations. When the data is operational rather than primarily representational, it shapes new *environmental* forms of power, and actually contributes to altering our physical and digital environment.

Situated data. Many data visualisations are presented as though they are neutral renditions of fact, with a “view from nowhere”. This is what Donna Haraway called “the God trick” in “Situated Knowledges” (1988, p. 581). Examining different representations or framings of data from a single app highlights the situatedness of the “knowledges” and corresponding power relationships that are presented through situated data and data visualisations.

I am introducing the term *situated data* as a tool to analyse how the same data can be situated in different ways and how this situation is integral to power relationships between users, platforms, and ideologies. A fundamental tenet of situated data analysis is the acknowledgement that data is always situated. It is not a given and is not objective, not even when we are dealing

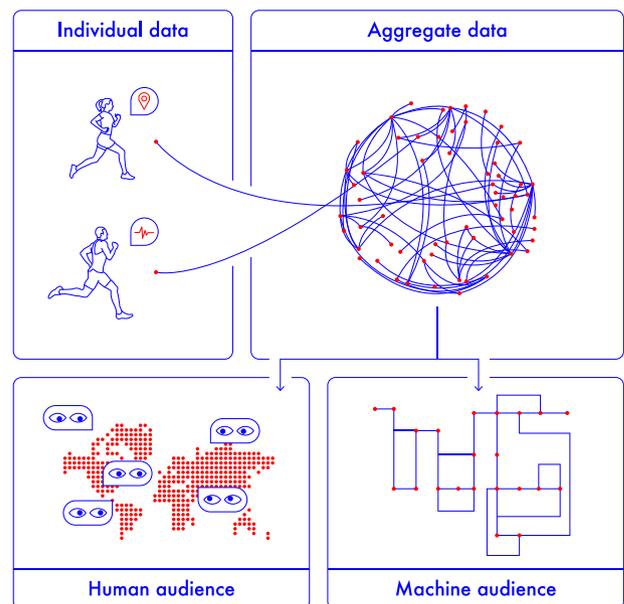


Fig. 1 Data situated for different audiences. Data from individual users of Strava is aggregated and processed for human audiences through data visualisations that are representational, but machine audiences do not need visual representations of the data. Diagram designed for this publication by Maren L. Knutsen at Overhaus.

with immense quantities of data. Data does not provide us with an all-knowing, all-seeing, God's eye view of data. The digital humanities have shown this again and again, from Lisa Gitelman's anthology *Raw Data is an Oxymoron* (2013) to the extensive work in feminist, Black, intersectional and postcolonial digital humanities (see for instance Brown et al., 2016; Losh and Wernimont, 2018; Risam, 2019a). Research on algorithmic bias shows that racial and gender biases are encoded into the datasets used for algorithmic prediction (Bolukbasi et al., 2016; Buolamwini and Gebru, 2018; Eubanks, 2018; O'Neil, 2017), and artists and scholars have demonstrated how datasets used in machine learning are flawed, misleading and biased (Crawford and Paglen, 2019; Kronman, 2020). Situated data analysis recognises that data is always partial and situated, and it gives scholars tools to analyse how it is situated, and what effects this may have.

In the following I analyse four levels of *situated data* in Strava. The boundaries between levels can be fluid, and other divisions could have been made. In my analysis I emphasise the source of the data (individual user or aggregate users) and its intended audience (an individual user, humans in general, or machines). The four levels I identify are (1) personal data visualised for the individual user and as shared with friends and nearby users, (2) aggregate data visualised for humans, and (3) aggregate data as an operational dataset, intended either for human users manipulating data through a dashboard showing data visualisations, or (4) for machines that process the data to generate new information. Before discussing each of these layers, I will explain the concepts of operation and representation and of disciplinary and environmental power, which are important to the analysis. These are not two sets of binary opposites but represent a spectrum. After the analysis of Strava, the paper concludes with recommendations for how to perform a situated data analysis of other platforms.

Operation and representation

Strava uses data visualisations to communicate its data to human viewers, whether individual users looking at statistics and visualisations of their latest run, or city planners looking at aggregate data showing how cyclists move through the city. Data visualisations are a genre of visual communication that has always been closely tied to quantification. The connection between nation states' collection of statistical, demographic data and the "golden age of data visualisation" in the 19th century is well-established (Friendly, 2008), and data visualisations have often revealed significant patterns that were not evident before visualisation. In his massively influential book *Envisioning Information*, Edward Tufte formulates the fundamental challenge for pre-digital data visualisations thus: "The world is complex, dynamic, multidimensional; the paper is static, flat. How are we to represent the rich visual world of experience and measurement on mere flatland?" (Tufte, 1990). Today's visualisations of big data present different challenges. The abstraction that Tufte describes as occurring between the "complex, dynamic, multidimensional" world and a data visualisation today happens at the point where data is captured. Of course, this was also the case in the "golden age" of statistics, but in contrast to then, many if not most data visualisations today are automatically generated². Each individual user of Strava sees their own data represented in standardised, computationally generated visualisations.

Johanna Drucker writes that data visualisations can either be representations or knowledge generators: "A basic distinction can be made between visualisations that are *representations* of information already known, and those that are *knowledge generators* capable of creating new information through their use" (Drucker, 2014, p. 65). Drucker describes representations as static, while

knowledge generators "have a dynamic, open-ended relation to what they can provoke" (p. 65). Representational visualisations often leave out information and present statistics as abstracted "from circumstance" to avoid "troubling detail" (2014, p. 92). Of course, the troubling detail may often be important factors in the actual lives of the people whose data is the basis of the visualisation. Knowledge generators may likewise abstract the "troubling detail" (in fact, this is perhaps the very foundation of any visualisation) but for Drucker they are also combinatory, "taking a fixed set of values and allowing them to be recombined for different uses and purposes" (2014, p. 105).

Drucker's distinction has a lot in common with the distinction proposed by filmmaker Harun Farocki and developed by artist Trevor Paglen between operational images and representational images. In "Operational Images," Paglen explains that "Instead of simply representing things in the world, the machines and their images were starting to "do" things in the world" (Paglen, 2014). This distinction between the procedural or computational or operational level of digital media and the representational level that is seen by humans has been noted before, as Aud Sissel Hoel points out (2018), but Paglen and Farocki's work demonstrates that even images and artifacts we have not previously thought of as specifically digital are becoming operational. Hoel calls for a theory of "operative media" that goes beyond images (2018, p. 27), and in his book *Automated Media* (2019), Mark Andrejevic goes some distance in developing such a theory, by combining Foucault's idea of environmentality with the idea of operationality. While Andrejevic argues that the operationalism of automated media sets them in absolute opposition to representation, I would argue that the reality is more nuanced than this³. His connection of operationality with environmentality is very fruitful, though, and I build upon it in this paper.

An app like Strava will always have operational aspects, because the data is obviously computationally processed, but it is fair to say that the primary purpose of the data visualisations as they are shown to the individual user is representational. The user wants to see a representation of their run or bike ride. This is a form of numerical or data-visualised self-representation, a way of seeing ourselves through technology (Rettberg, 2014). The emphasis shifts from representation to operation when the data are aggregated and used for other purposes, such as city planning or generating automated route suggestions.

Discipline and environmentality

As Deborah Lupton puts it, "the boundaries between small and big data are porous" (Lupton, 2016, p. 63). Individuals may not really experience their data as being part of a global set. They see their personal data compared to the data of their friends on the app, and to the fastest cyclists or runners in their area. But the truly big data is global and gives the companies that own the data a great deal of potential knowledge and potential profit. This tension between the intimacy of personal data, and the sometimes surprising uses that emerge when millions of individuals' personal data is aggregated is at the root of much contemporary anxiety about big data. Clearly, big companies make a lot of money from aggregating personal data. Detailed information about our personal preferences and habits is not only used by the apps and platforms we intentionally use; it is also shared with and sold to adtech companies and data brokers. In early 2020 the Norwegian Consumer Council reported finding that ten apps they tested transmitted user data to "at least 135 different third parties involved in advertising and/or behavioural profiling" (2020, p. 5). Data shared ranged from location, age and gender to highly sensitive information about sexuality, religion and drug use, despite such sharing clearly being illegal under the General Data

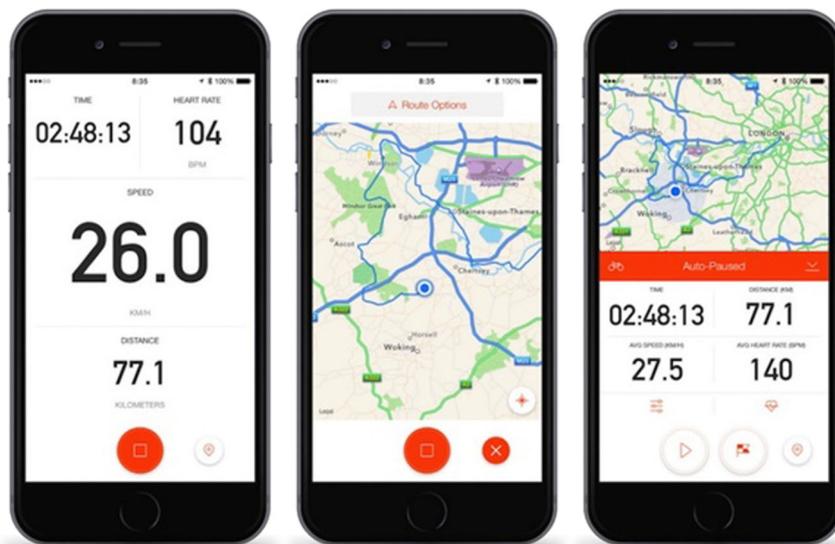


Fig. 2 Real-time data visualisations. Visualisations of an individual user's cycling data as shown to the user in the app Strava.

Protection Act (p. 6). The people who use these apps are usually neither aware of who their data is being shared with or what it is being used for.

The concepts of *environmentality* and *environmental power* can be productive lenses for understanding how aggregated personal data can involve a different kind of power in addition to the biopolitical and neoliberal control of bodies and population through numbers. Environmentality is a concept mentioned but barely developed by Foucault. At the end of a lecture he gave in 1979 he described environmentality as a new kind of power that “appears on the horizon (...) in which action is brought to bear on the rules of the game rather than on the players, and finally in which there is an environmental type of intervention instead of the internal subjugation of individuals” (Foucault, 2008, p. 260). While disciplinary societies use visible surveillance and other mechanisms to give their citizens a deep sense of internal discipline that keeps them in line, environmentality doesn't require individuals to have any normative sense of right and wrong. Instead, the environment is designed in a way that makes certain behaviours easy and others difficult. Foucault promises to return to this in his next lecture, but does not, however the idea has been taken up recently by several scholars. Jennifer Gabrys argues that “environmental technologies” in smart cities “mobilise urban citizens as operatives within the processing of urban environmental data” (Gabrys, 2014). So rather than the individual being governed, the individual becomes an “operative” in a larger system. Erich Hörl connects Foucault's *environmentality* to Skinner's behavioural engineering and Gibson's affordances, both lines of research that emphasise how the psychological or physical environment can be adapted to nudge people towards certain behaviours. This “radically relational and procedural conception of environment” (Hörl, 2018, p. 160) is also well-aligned with new materialist and posthumanist understandings of matter as “vibrant” (Bennett, 2010), technology as having the potential for non-conscious cognition (Hayles, 2017) and the need to understand how humans, things and technologies function together as assemblages.

Level 1: Personal data visualised for humans

Like many self-tracking apps, Strava is designed as an intimate companion that the user can confide in, presenting itself as a trusted diary-like tool (Rettberg, 2018a). The data you see in the app when you run is personal (Fig. 2). As you run, you see clean

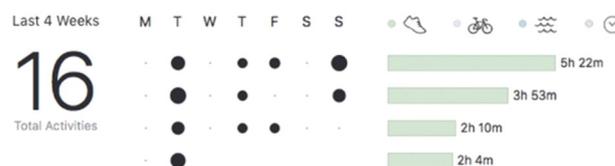


Fig. 3 Personal statistics. A section of a Strava user's dashboard, showing an overview of the frequency and length of their runs in the last 4 weeks. Reproduced with permission from Strava.

black numbers on a white background, with a red button that starts or stops the tracking. You can swipe to see a map of where you have run or pause to see more detail. As you amass data after running regularly, you can view visualisations of your running habits on the Strava website, seeing your habits stack up neatly as in Fig. 3, or perhaps being disappointed at a more haphazard pattern. While you run, Strava uses minimal visualisation, presenting you with numbers and facts instead: You have run for so many minutes, your speed is this. Afterwards, you see visualisations comparing your effort both to your own previous efforts, as in Fig. 3, and to results from your friends on the app and other local Strava users. For instance, when I log a run around the lake near our house, I pass through two “segments” that have been defined by other users, and Strava automatically compares my performance on these segments to my earlier runs. Last time I ran, Strava put a little bronze medal icon next to the map of my run to indicate that I had achieved my third fastest time on the 300 m stretch on one side of the lake. Clicking on this, I could see my personal record for this stretch, when I achieved it, and how many “efforts” I had made. A link to the Leaderboards encouraged me to compare myself to other Strava users who have run the same stretch. I could view the fastest women, the fastest men, the fastest users this year or today, and if I paid for a premium subscription, I could compare myself to even more specific groups, for instance to other users of my age or weight. By using the map of the run as a central visual and conceptual interface to my own and other users' data, Strava creates a sense of sociality that is local. The scope of my data sharing feels limited. I can only see data from my friends and from people who live or at least run and cycle near me.

At this level, Strava functions as a technology of the self. The user chooses to use it, chooses to log their data, and primarily sees

their own data. There is no emphasis on the fact that the individual user's data is also being shared widely and aggregated, although the comparisons with friends and with named users in the same geographic area establish a sense of immediate local community, both in terms of geography and personal relationships. The graphs, comparisons and badges help motivate users, as one of the Strava users in a study by Lupton et al. (2018) reported: "Then you start to rush because you want to beat your time. There's a little graph that they draw [on the Garmin website] and you never want to go slow" (p. 659). Visualisations like that shown in Fig. 3 give a sense of aesthetic satisfaction when users use Strava regularly, and provide disharmonious patterns for more irregular usage, reinforcing the discipline that the users are trying to develop. This of course connects to societal norms encouraging fitness and health, but also the societal norm that self-discipline and regular, healthy habits are good in themselves.

The Strava app does not invite users to think about their data as part of a global data-gathering infrastructure, indeed, it could be argued that it obfuscates this (Sareen et al., 2020). Rather than the grand scale of global or city-wide visualisations, individual cyclists deal with their own data, seen in the context of their connections on Strava and of the leader boards and the information about the fastest cyclists in their area. This *mundane data* (Pink et al., 2017) is presented to the individual user as their own. However, many Strava users are very aware of what data they share with friends and other users and use selective sharing to negotiate the relationship between themselves and other individuals, similarly to users of other social platforms (Kant, 2015). In a study of Strava users' communicative practices, Smith and Treem report an informant telling them about withholding specific kinds of data:

"A lot of guys don't share their heart rate and wattage, because of that, they'll just share, hey this was my segment time and this is what my elevation, just the really basic data to share, but the pros rarely post their heart rate and wattage" (male, mountain biker, 31) (Smith and Treem, 2017, p. 143).

The cyclists Smith and Treem interviewed spoke of using their own data as self-validation, but also imagining others looking at the data. As one informant told them, "There's an element of vanity to it, look at me, look at how fit I am." This rider went on to summarise, "[T]here's this subconscious thing in the back of your mind where you think someone is looking at your Strava profile" (p. 144). It is easy to make the connection from this to Foucault's analysis of disciplinary power and the idea of the panopticon, where citizens behave because they know they might be being watched at any moment. This is an example of disciplinary power rather than environmental power.

There is also a sense of pleasure in viewing one's own data. Like taking a selfie, using a self-tracking app allows individuals to reflect upon themselves, and to create a self-representation they can admire. In a study focusing on how cyclists *sense* their data, Lupton et al. found their informants spoke of their enjoyment of seeing the numbers:

Tony observed that he enjoyed reviewing the details that Strava gives him about his rides, including gradient graphs, speed graphs, heart rate fluctuation graphs, and "power zones," which provide a calculation based on combined data on heart rate, gradient, and speed. He also enjoyed looking at the dashboard on his Garmin bike computer during his trips so that he could monitor his data in real time (Lupton et al., 2018, p. 654).

This pure pleasure in having data is fascinating. Users were also aware of the inaccuracies and lacunas in the data they

collected. Strava sometimes gets the GPS tracking wrong, and it does not track everything, not even things that it could track and that could impact a cyclist's performance, such as the air temperature or how well the cyclist slept the night before (Lupton et al., 2018, p. 654).

Strava sets up very clear goals for its users: go faster to beat your own records; compete with others to get on the leaderboard for a particular segment. Lupton et al.'s informants described altering their cycling plans in order to do well at these tasks: "Like on the weekend, on Saturday, there was a bit of an easterly coming up in the afternoon when I headed out for a ride... and I structured my ride around being able to take advantage of that, and went after a few Strava segments that I hadn't done so well on in the past" (p. 656).

In summary, the data visualisations Strava users see of their own data and their friends' data support their use of Strava as a "technology of the self". The social aspects of Strava in particular support a biopolitical form of discipline rather than environmental control.

It is in the comparison of different contexts and situations that situated data analysis has its greatest effect, although it is quite possible that the method can be applied to just one level with productive results. The methods used for analysing each level of situated data will be different, both based on the approach chosen by the researcher and what kinds of access to data are possible on different platforms and at different levels. My analysis of the first level of situated data in Strava used semiotic analysis of the visual interfaces and data visualisations and drew upon published ethnographic work on Strava by other scholars. Interviewing users directly or using other ethnographic methods, such as observations or diaries would be an obvious alternative, or one could use the walkthrough method to analyse usage of an app with a particular focus on how data is gathered and situated for an individual user (Light et al., 2016). Another strategy might be to analyse how the platform situates data for an imagined or ideal user, and to push against that. Safiya Noble does this when discussing Google search results for "black girls", writing of the pain of seeing young black girls confronted with racist and sexist search results clearly not intended for them (Noble, 2018). Scholars in Black digital humanities do this when considering the effects of the digitisation of records of enslaved people (Johnson, 2018), and critical race scholars do it when analysing encoded racial stereotypes (Benjamin, 2019) or the ways DNA analyses reinscribe racial categories (M'charek et al., 2020). None of these methods of analysis will tell us everything, but as Haraway reminds us, acknowledging that our knowledge is partial and situated is the closest we will get to objectivity.

Level 2: Aggregate data visualised for humans

Strava presents an example of data willingly gathered by individuals for their own benefit, but where that data is also used at an aggregate level that the individuals have less control over. This entails particular ethical issues. Roopika Risam recently showed that when data visualisations showing the flow of migrants into Europe are created from datasets that lack participant involvement, they tend to dehumanise migrants, framing them as a problem. Notably, the "othering" of the migrants increased when visualisations used existing data, but in cases where the data visualisers collaborated with migrants about the data and its visual representation, the visualisations were less dehumanising and expressed more empathy with the migrants (Risam, 2019b). Strava users are hardly full collaborators in the way their data is collected, aggregated and visualised, but they do have choices in terms of what data they choose to record and what data they choose to mark as private. Strava's data visualisations of global

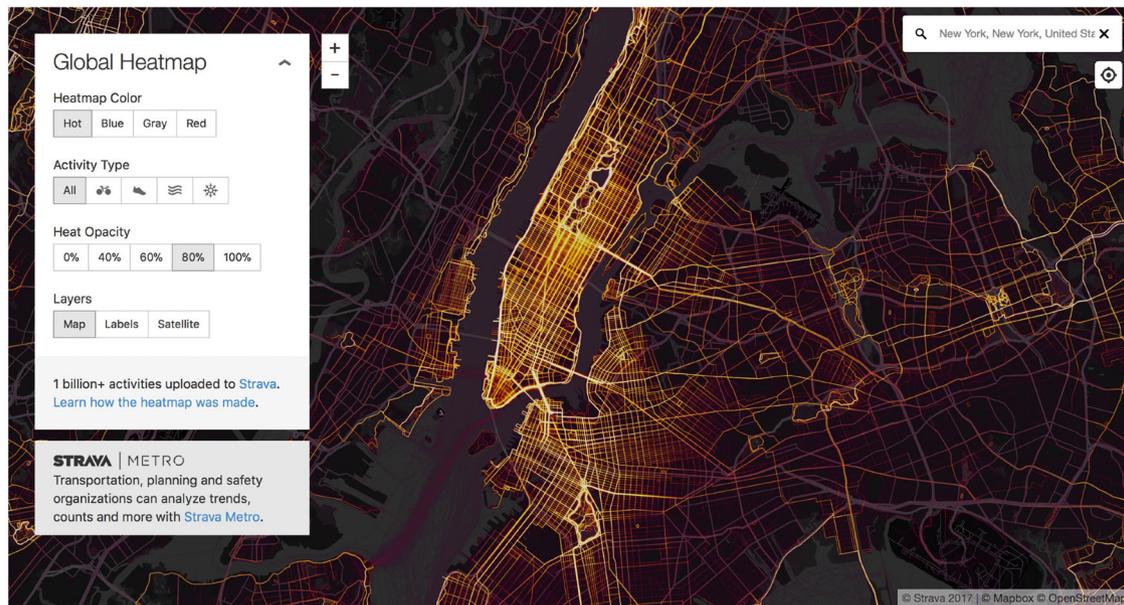


Fig. 4 Strava Global Heatmap. The view of Manhattan visualises the paths taken by cyclists and runners as golden lines on a map of the city.

data have the app users as a main intended audience, providing a sense of community, much as a range of other personal data tracking services provide in their annual visualisations of user activity (Rettberg, 2009).

The data visualisation of aggregate data that Strava promotes the most is the Global Heatmap⁴, a publicly available website with a world map showing the paths of runners and cyclists as glowing yellow lines on a dark background. Brighter lines mean that more Strava users have used those routes. Part of the Global Heatmap's appeal is the thrill of recognition combined with a slight newness. We recognise the map of Manhattan shown in Fig. 4, but at the same time it is different from the maps we are used to seeing, with its strong emphasis on the central parts of the city, and the strong visibility of the irregular shapes of paths in Central Park. The relative darkness of Harlem, just North of Central Park, may be due to demographic imbalance in who uses Strava and thus whose movements are visualised rather than actual running and cycling activity. As Alex Taylor writes of a similar map of journeys made on London's bikesharing network, when we visualise "the most common cycle journeys or where the flows are most dense" we generate "visualisations that seem to conveniently remind us where the wealth flows in London" (Taylor, 2016, p. 201).

I will analyse the data at this level in two ways: looking at the data visualisations as representations and looking at how the individual user is situated as an anonymous member of a community but can also be resituated and deanonymized from the data.

Aggregate data as representations. Data visualisations are representations of *data*, but data is itself a representation, a measurement, a proxy for that which the visualisation purports to represent. Representations always situate that which they represent in a particular way. They are constructions (Hall, 1997, p. 25) that situate reality or fiction in a particular way. Figuring out what the data actually represents and what is left out is an important step in situated data analysis.

One of the ways data is constructed is by using proxies. We cannot measure everything, so datasets often consist of proxies. A proxy is something that we *can* measure, or that we have data

about, that we assume directly corresponds to the thing that we are trying to analyse or represent. Proxies do not always tell the full story. For instance, when Sendhil Mullainathan and Zaid Obermeyer used machine learning to analyse the health records of patients who had had a stroke, the system found that based on the data available, having had prior accidental injuries or acute sinusitis or having been screened for colon cancer were among the statistically valid predictors for later having a stroke (Mullainathan and Obermeyer, 2017). As it turns out, having a bad cold with acute sinusitis does not actually cause you to have a stroke later on. Strokes are not always diagnosed, and many people who experience a minor stroke do not go to hospital and may never realise that they had a stroke, so, "whether a person decides to seek care can be as pivotal as actual stroke in determining whether they are diagnosed with stroke" (p. 477). The problem is that health records are incomplete, because patients do not always go to the doctor, and even if they do, not every aspect of their health will be diagnosed. Health records are interpretations of reality, rather than pure reflections of it, or, in Mullainathan and Obermeyer's more medical terminology, "Many decisions and judgments intervene in the assignment of a diagnosis code—or indeed, any piece of medical data, including obtaining and interpreting test results, admission and re-admission of a patient to the hospital, etc." (Mullainathan and Obermeyer, 2017). In this case, the existence of data about an injury predicts the existence of data about a stroke, because "both are proxies for heavy utilisation patients" (p. 478). Johanna Drucker suggests that perhaps we should use the term *capta* instead of data, to emphasise that data is captured from reality and is not equivalent to reality (2011). There is no such thing as raw data, as Lisa Gitelman points out (2013).

What are the data and the proxies that are being visualised in Strava's Global Heatmap? Strava's Global Heatmap visualises data generated by its users as they track their runs, bike rides and other activities. The most obvious limitation is that it does not include data about runners and cyclists who are not using Strava's software. It also does not include data about Strava users who have marked their data as private, or about Strava users' movements inside of areas that they have marked as private. Sometimes the GPS measurements are off, resulting in faulty paths. Sometimes a Strava user forgets to wear their tracker and



Fig. 5 Night Lights 2012 Map. Strava's Global Heatmap reminds us of images of the Earth as seen from space. Even this "God's eye view" of Earth, from NASA's Earth at Night collection, is an example of situated data: it is a visualisation of filtered data gathered over a six month period.

goes for a bike ride that is not tracked. Other times, perhaps a tracker is forgotten on a bus and movements are tracked that were not cycled or run. Other limitations are noted by geographers who want to use aggregated data from the package Strava calls Strava Metro, which is discussed below, in Level 3. The geographers cannot analyse individual riders' habits or trajectories, do not have access to the age, gender or socio-economic status of riders, which would be useful for designing better cycling infrastructure, and GPS accuracy is not perfect (Romanillos et al., 2016). The demographic imbalance is also noted as a problem, as Strava users are predominantly male and affluent (Griffin and Jiao, 2015). Strava's data is a particular, situated representation of reality, with clear limitations and errors.

The visual rhetoric of a data visualisation also makes certain claims about truth. The most obvious design choice in the Strava Global Heatmap is to make the exercise paths show up as lights on a black map of the Earth. Maps that show various features of technological adaptation are frequently shown in this style, with individual users showing up as lights on a dark globe. The image also reminds us of photos of Earth as seen at night from space, with cities visible by their lights (Fig. 5).

It is important to remember that even an apparently photographic image, such as that of the Earth shown in Fig. 5 is actually a data visualisation, and not something that could be seen by a human, even if a human were positioned in space. First of all, it is a composite image assembled from images taken by an unmanned satellite 6 months apart. Second, the image is not a simple representation of light that is visible to the human eye. NASA's description explains that this series of images used the visible infrared imaging radiometer suite (VIIRS), which "detects light in a range of wavelengths from green to near-infrared." They also used "filtering techniques to observe dim signals such as city lights, gas flares, auroras, wildfires, and reflected moonlight," and because the image is a composite, Australian wildfires that happened at different times are all visible simultaneously (NASA, 2012). In other words, the image of the Earth in Fig. 5 is already a

situated data visualisation that carefully shows certain kinds of data, in this case city lights, as a proxy for dense and technologically supported human habitation. The image excludes other data, such as natural lights or information about any human habitation with fewer electrical lights, as you might see in a refugee camp or poorer community. Refugee camps might show up on a night-time image taken from space, but at a far lesser rate than more affluent communities do. Based on studying nighttime images from space over time, a group from Refugees Deeply estimated that "a 10 percent increase in the refugee population is associated with a 3.6 percent increase in the nighttime lights index within 10 km (6.2 mi) of the camp" (Alix-Garcia et al., 2017). We could say that NASA's image of Earth is a representation of an operational process of visualisation.

Strava's Global Heatmap clearly references the many composite images we have seen of the Earth at night. It also, perhaps inadvertently, reproduces colonial, Eurocentric maps of the world where Africa was shown as a "dark continent". Indeed, so do NASA's visualisations, although since NASA is explicitly emphasising city lights as seen in the dark night, their choice of colour scheme is at least based on the actual colours of the data they are visualising.

As Martin Engebretsen notes, the viewers of a visualisation will tend to assume that all the visual elements, including colour choices, *mean* something (Engebretsen, 2017). The Global Heatmap replicates the aesthetic of heatmaps, coming from infrared vision, perhaps suggesting the Strava app as equivalent to a God's eye view from space that can see more than any human eye. This matches their emphasis of the sheer *size* of their dataset in press releases (Rettberg, 2018b), and fantasies about big data's framelessness (Andrejevic, 2019, p. 113) and lack of a point of view (Feldman, 2019). Situated data analysis instead emphasises the situatedness of data even when it is presented as immense, all-encompassing and objective. We can see this situatedness in the Global Heatmap's perhaps inadvertent but telling reproduction of colonial maps, visualising Strava use as light against a background of darkness.

Re-individualising aggregate data. The Global Heatmap situates the aggregate data from all its users as communal. The data is not situated as belonging to an individual user, but as coming from a global community of users who merge together, almost forming a superorganism of runners and cyclists lighting up the globe. We imagine a shared humanity collectively creating glowing lines of light through the world. Moving through the map visualisations is a very sensuous experience, and the visualisations are beautiful, perhaps even sublime (Rettberg, 2018b).

Big data that comes from people may be anonymised with the intent to only reveal global patterns, but often the identity of individuals can still be reconstructed. The case of the reidentification of individuals from AOL's massive data dump of anonymized searches in 2006 is a frequently cited example of this (Barbaro and Zeller, 2006), and Strava's global data has also been traced back to individuals. A couple of months after Strava released its 2017 Global Heatmap, Nathan Ruser, a 20-year-old Australian student of international security and Middle Eastern studies, posted a series of tweets showing how the Heatmap precisely revealed the locations of US army bases in the Middle East by showing the exercise paths of Strava users (Ruser, 2018). Looking at the global map you can easily see that most Strava users come from specific cultures. Europe and North America are densely criss-crossed with tracks, while the Middle East and Africa are mostly black on the map. But if you zoom-in on the Middle East, you'll see a few lit-up dots. The Norwegian broadcasting company NRK used Ruser's idea to identify the individual profiles of European soldiers whose running routes were shown on Strava's heatmap, although they had to use third party software and feed Strava false GPS coordinates to access the information (Lied and Svendsen, 2018). Responding to Ruser's tweet about the army bases, twitter user Brian Haugli also noted that the global map allows you to see the actual homes of individual Strava users, by zooming in to see individual houses and looking for houses that have short lines connecting them to the thicker running paths (Haugli, 2018). By default, Strava users automatically share all their location data, but it is possible to set all data to private, or to set a "Privacy Zone", so that no data from immediately around your home is shared. Obviously, many soldiers and individuals have not done this, and are probably unaware that their private data is showing up in global visualisations like this. This is an example of how the way that data sharing is situated to the individual user or data subject at one level (e.g. in the app) can be at odds with the way the data is displayed at a different level (e.g. on the Global Heatmap).

In summary, we see that Strava's Global Heatmap is primarily representational: it is a data visualisation of aggregate data meant for a human audience. The data visualisation is a visual representation of the data, and the data itself is a constructed and thus situated representation of the reality of users' actual exercise patterns. The Heatmap is a rhetorical statement with inherent biases, such as the emphasis of areas frequented by wealthy, Western users. In this way the Heatmap enacts power much as other forms of representation do: by showing the world from a particular point of view it makes implicit claims about what is important. It also inadvertently puts individuals at risk by disclosing their data in ways that are not made clear at other levels. Although the Heatmap is presented as a representation of immensely rich data, it is situated and limited as all other data.

Methods that can be used for analysing how data is situated at a level where aggregate data is represented to humans can include semiotic, visual and rhetorical analysis, explorations of the data's provenance and omissions, and could also include ethnographic observations or audience studies. I also used media reports and social media discussions where people have tried to deanonymize Strava users based on the publicly available aggregate data. This

kind of reverse engineering could be used as a method in itself, although one should consider potential ethical and legal issues if attempting it. In the following we will see how the same data can be used in a more operational mode, and how that leads to more environmental modes of power.

Level 3: Aggregate data as an operational dataset

Less visible to ordinary users is "Strava Metro", a programme where Strava sells aggregate user data to cities with the slogan "Better data, better cities."⁵ Although little if any information is provided to users about this programme, the website presents it as a community effort that users are willing participants in: "Our community actively contributes because they know logging their commutes is a vote for safer infrastructure." Customers can purchase access to the data with a web dashboard (shown in Fig. 6) allowing them to quickly see where and when Strava users are cycling and running in a specific city.

Here user data is completely anonymized, and individual users can no longer see themselves in the data as they can in the app. The dashboard still provides *representations* of the data, but the representations are not the end goal as in the earlier examples. This is *operational* more than representational. This data is no longer functioning as a technology of the self. It is still involved in biopolitics and population management, but no longer in a way that is directly experienced by the individual cyclists and runners. Instead, this data supports an *environmental* power, as city planners use data from Strava users to redesign the city to encourage or alter the usage patterns they observe.

Uses of Strava's datasets do not always consider the difference between Strava's data and actual cycling, as in an analysis of cycling in Johannesburg that found only 20% of rides were commutes, but that neglected to consider that Strava users were not a representative sample of actual cyclists in Johannesburg (Selala and Musakwa, 2016). This paper cites, without acknowledgement, on p. 588, a statement that I assume, based on how often it is cited elsewhere online, must at one time have been published on Strava's own website, "Strava Metro is a data service providing "ground truth" on where people ride and run." But of course, this "ground truth" is not the whole truth. Ground truth is an important term in machine learning, but while other fields, such as meteorology, can compare their models to the "ground truth" of weather measurements made on site, the ground truth in machine learning is usually at least one step removed from the "real world". The ground truth for a machine-learning algorithm is the training data set with its labelled data (or in this case data from Strava users), which is itself often biased and non-representative⁶. Most cyclists probably do not use Strava, and the cyclists who do use Strava are more likely to be affluent and male than those who do not use the app. Especially in a city like Johannesburg, this biased data set can lead to very inequitable analyses.

Other studies do recognise that Strava data is demographically biased. For instance, a robust study on Victoria in Canada that used Strava data alongside geological survey data on slopes, population data, city data on pavement width, parked cars, and bicycle lanes, and more, found that there was a linear relationship between crowdsourced data from Strava and manual counts of cyclists in the same locations and at the same times: for each Strava user logged, there were 51 actual cyclists on the street (Jestico et al., 2016). This was despite the data showing that 77% of Strava users were men, so not a representative sample of the population. Similar findings were made in Glasgow, where a volunteer pro-cycling advocacy group called GoBike! compared Strava data from Glasgow to roadside surveys of cyclists conducted by the city, and found that despite 82% of users in

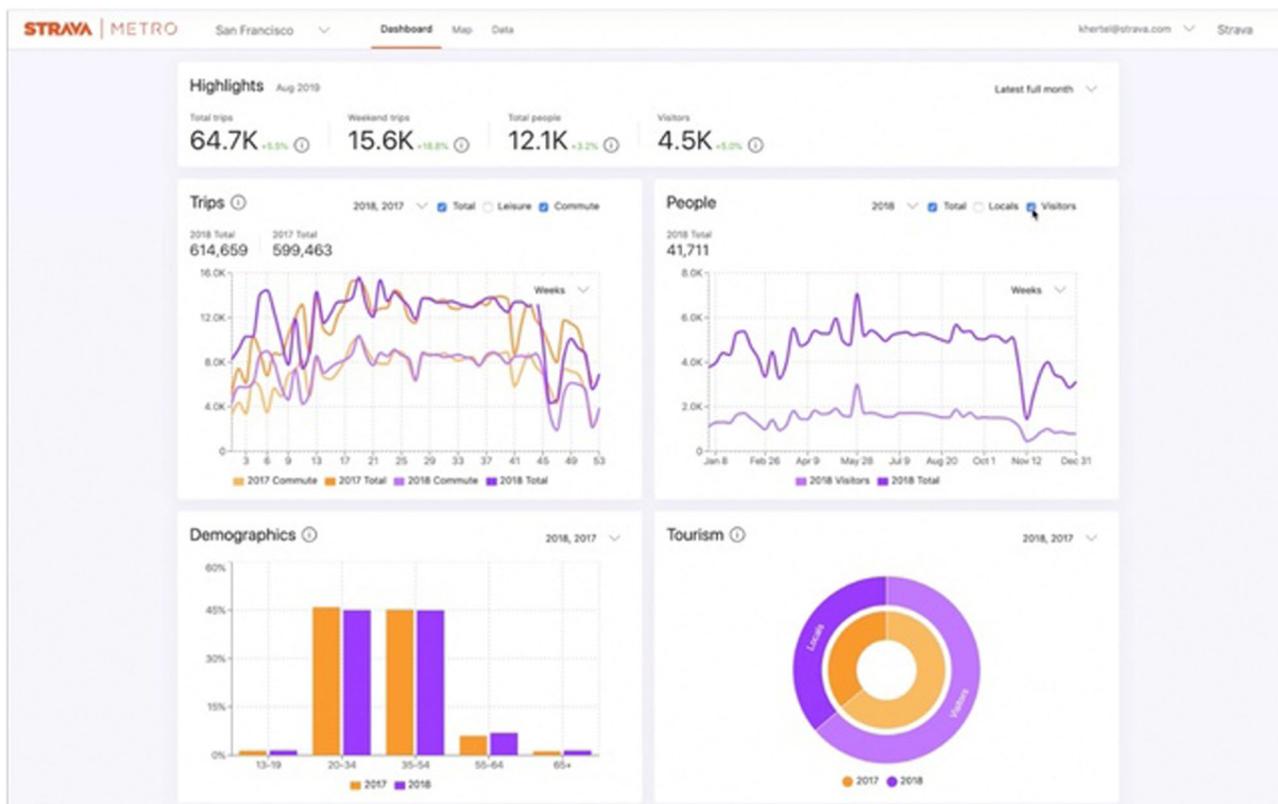


Fig. 6 A screenshot of the Strava Metro dashboard for San Francisco, as shown in a demo video for the product (Strava, 2019).

Glasgow being male, Strava data closely matched the manual count in the roadside surveys (Downie, 2015).

To analyse the situatedness of data at a mostly operational level like this can be more difficult than analysing the representational level, which is directly accessible to us. My approach here was to look at published studies by people who have used the data, and to view the promotional material made public about the system. If it is possible to gain direct access to the system, another method might be to test it out, perhaps applying the walkthrough method (Light et al., 2016) or using critical code studies (Marino, 2020) or other experimental methods to analyse how the system works. In the case of Strava, researchers might be able to purchase access to Strava Metro data and compare it to data presentation in other levels of the analysis. A third approach might be to interview people who use or develop the system, or the people and institutions who purchase the data to understand how they use it.

Level 4: Aggregate data for a machine audience

At level 3, Strava Metro data is presented as operational data that is processed by humans working with machines. Level 4 of my analysis examines how Strava data is algorithmically processed by machines with little human involvement. At this level there is no need for human-readable representations such as data visualisations and the data leans heavily to the operational end of the spectrum. The example I will focus on is the processing of “trajectory data”, that is, data about how people or vehicles move through space. Trajectory data from apps like Strava can be used as data sets for systems that can calculate the best route for a self-driving car to take from point A to B, given changing traffic patterns at different times of the day, or an algorithm could predict future traffic and use that to recommend altered traffic light signalling, different routes for public transport, or the improvement of bicycle lanes along certain roads.

In an overview of trajectory data mining, Yu Zheng describes three common ways of visualising numerical data about the trajectories of people or vehicles through space. First, trajectories can be projected onto an existing map, treating the map as a network graph in itself, where intersections are nodes and roads are edges (in a network graph, an edge is the line between two nodes). Then the data about actual movements on the roads is used to weight the existing edges, so for instance, the thickness or colour of the roads would be altered according to the speed or volume of traffic on them. Second, the trajectories could also be visualised in a landmark graph, where the emphasis is on the trajectories themselves rather than on the conventional map. Here, paths that are frequently taken become landmarks and the trajectories themselves are seen as the network that generate a visual mapping of the area. Third, the map could be divided into regions, and arrows could show how traffic moves between different regions.

Zheng is not very interested in what the network graphs look like. His goal is not visualisation, but computation. Transforming trajectories “into other data structures (...) enriches the methodologies that can be used to discover knowledge from trajectories,” Zheng explains (2015, p. 24). He does include images of two of these kinds of network graph, just as he includes mathematical formulas to describe the trajectories in their non-graph format. For Zheng, as for the computers that process this data, it is a pure coincidence that the network graphs are also visually interesting. These almost coincidental visualisations are “images made by machines for other machines”, as Paglen described operational images (Paglen, 2014).

At this level, the data is primarily operational, and it is not directly experienced by human users. Instead, the data is used to change the environment, for instance by resulting in analyses that lead to the physical environment being changed or changing municipal allocations to various kinds of infrastructure, or by

impacting the route guidance individual humans may receive when planning a route across the city. Perhaps a driver would be routed away from roads that were heavily used by cyclists, or a runner would be given route suggestions that slightly increased the distance run in order to encourage greater health benefits. This kind of power is environmental, not disciplinary, and can change behaviour although it may not even be noticed by the people affected by it. It is also a power not directly wielded by Strava, but one that might better be understood as enacted through an assemblage, following Jane Bennett (2010) and other feminist new materialists and posthumanist scholars, which consists of individual Strava users, the data collected by them, the algorithms that aggregate the data, the APIs that make it reusable, the algorithms that are used to process it, the apps that provide an interface for new users (for instance navigational apps) and the cities or other parties that use it.

The increased operability of the data at this level naturally makes it less visible to humans, as it is meant for machines to read, not for us. That means its effects can be harder to spot using traditional social science and humanities methods. My solution has been to read technical papers to understand how the data can be used. This allows me to speculate about the environmental effects such uses of data might have, for instance on urban design and on the technical infrastructures we use. Other approaches might be to look for patents filed by the company, as these often include descriptions of how data can be used (see Andrejevic, 2019, pp. 117–118 for an example), or interviewing stakeholders such as developers, urban designers, public health officials, cyclists, professional drivers or residents or homeless people affected by changes in physical and digital infrastructures. Obviously, no single analysis could do all these things, but an awareness of other possible ways of studying situated data is central to situated data analysis, because this is a way of acknowledging and reflecting upon how the analysis itself is also necessarily situated.

Conclusion

I have analysed four levels of situated data in Strava: (1) personal data visualised for the individual user and shared with friends and nearby users, (2) aggregate data visualised for humans, and (3) aggregate data as an operational dataset, intended either for human users manipulating data through a dashboard showing data visualisations, or (4) for machines that process the data to generate new information. All these levels combine the representational with the operational, with operability becoming increasingly important the further the data is removed from the individual user. Representation is present not only in the visualisation of the data for human users, but also in the construction of the data itself.

As the data is situated in a more operational context, the power relations also move towards the environmental. This relates to the visibility of technologies of the self, and to a society where self-discipline is normalised partly through a knowledge that others may be watching us at any time. We use data visualisations of our personal data, like those we see in the Strava app, as self-reflections that allow us to improve ourselves. We are aware of what aspects of our personal data we share with others, and we compare our own metrics with those of others that are made visible to us. Strava Metro, on the other hand, is de-individualised and more operational than representational. Here the users are not situated as subjects seeing themselves and each other but as *sensors* feeding data into a larger system. In such an operational setting, power becomes environmental. Analysing the data allows cities to change the environment, and thus change the behaviour of residents.

Strava's levels of situated data are more openly accessible than many other systems, but it is reasonable to assume that we can see similar patterns elsewhere as well. At the individual level, YouTube or users of TikTok or Instagram or Twitter learn how to self-brand and create content that is liked and watched by others, thus internalising a particular set of norms, though less as a deliberate choice than for cyclists tracking their rides on Strava⁷. Data about each user's performance on social media platforms is visualised to them through hearts showing likes, or follower counts and comments, which also make the gaze of others visible to them. At the aggregate level data about users' behaviour is used to promote and recommend certain videos above others, and on many platforms, data is sold to third parties. Like with Strava, this aggregate data becomes operational, and is used to extend an environmental power, to "change the rules of the game" and thus alter the behaviour of the people whose data was collected.

Situated data analysis requires thinking about how data gathered by a platform or app is situated in different ways. It can be used when analysing a single aspect of a platform, or for a comparative analysis of the different levels, as in this paper. If focusing on a single level, a situated data analysis should consider how the same data might be situated differently, even if the analysis does not examine each level in depth.

To apply a situated data model to other platforms than Strava, a scholar would first identify the different levels of situated data in the platform, based on the source of the data (individual user or aggregate users) and its intended audience (an individual user, humans in general, or machines). At each level, an analysis should ask whether the data is presented as data visualisations or other forms of *representation* for a human audience, or whether they are instead *operational* and processed by machines. To understand power relationships, a situated data analysis should consider what power relationships each level emphasises or makes possible: disciplinary, environmental, or perhaps other power constellations. Figuring out what the data actually represents and what is left out is also an important step in situated data analysis. Depending on the focus of the analysis, other questions might include what kinds of knowledge are made visible, how the data subjects (the Strava users in my example) are presented or framed, and who is given access to the data. Is the data presented as neutral, objective, partial, comprehensive, flawed? Approaching these questions through the lens of situated data analysis allows for a more nuanced understanding of how massive platforms use data in different ways in different contexts.

Situated data analysis is explicitly indebted to Donna Haraway's insistence that knowledge is always situated, and that an omniscient "view from nowhere" is impossible. The main point of a situated data analysis is thus to examine how the data is situated and what that means. Situatedness has been used in other scholars' approaches to digital media, but not in the Harawayian sense. For instance, Carolin Gerlitz, Anne Helmond, David B. Nieborg, and Fernando N. van der Vlist discuss the *infrastructural situatedness of apps*, which "involves attending to apps' relations to one another and to other things, practices, systems, and structures". They continue by noting that "not only are apps situated in a technical or material sense, but also they are situated in an economic and institutional sense and can be related to other objects, devices, systems, infrastructures, clouds, and environments, both by humans and nonhuman actors" (Gerlitz et al., 2019). This is clearly relevant to a situated data analysis, but the object of analysis in a situated data analysis is the data rather than the app. Even more importantly, while Harawayian situatedness can encompass relationality, as in Gerlitz et al.'s approach, it also contains a strong argument against the idea of "disembodied scientific objectivity" (Haraway, 1988, p. 576), which has been echoed in José van Dijck's critique of dataism, which van Dijck

defines as the false but “widespread belief in the objective quantification and potential tracking of all kinds of human behaviour and sociality through online media technologies” (Dijck, 2014, p. 198). This perception of quantification leading to objectivity is arguably one of the main rhetorical claims of data visualisations as a genre (Rettberg, 2020b, pp. 38–39). Situated data analysis allows us to analyse the rhetorical claims made by the particular ways data is situated, both in terms of its relationships, and here Gerlitz et al.’s approach could be useful, but also in terms of how it is framed and what claims are made about it.

A related but not identical notion of situatedness can be found in science and technology research building upon Lucy Suchman’s work (1987, 2007), but here it is not the data itself that is situated in particular ways, but rather the people interpreting the data (e.g. Almklov et al., 2014) or other modes of situational awareness between humans and technology (Stacey and Suchman, 2012). These understandings of situatedness could be added to a situated data analysis.

The situated data model could also be expanded to include other aspects of social platforms. A related model for understanding platforms is the app walkthrough method (Light et al., 2016), which provides a methodology for analysing a user’s experience of and interactions with an app. In a sense, the situated data model is a walkthrough method for data, but as mentioned previously, the walkthrough method can also be applied at specific levels of a situated data analysis. André Brock’s critical technocultural discourse analysis (CTDA) could also be used alongside a situated data analysis. Brock specifically intends CTDA to “prioritise the epistemological standpoint of underrepresented groups of technology users” (Brock, 2019, p. 6), and by analysing digital media from a perspective that is different from that which the platforms expect users to inhabit, the way the data is situated for a specific expected audience becomes very clear. Approaches like platform studies and infrastructure studies (Bogost and Montfort, 2009; Plantin et al., 2018) are also relevant, especially as these engage specifically with data and data flows (Helmond, 2015).

We live in a society with increasingly complex layers of data collection and use. It is often hard to parse the contradictions in our own usage of massive platforms like Google or Facebook or even something as relatively single purpose as a fitness tracking app. At one level, these platforms are useful to us and even pleasurable and satisfying. At the same time, the knowledge that our data is also being used in completely different ways is worrying. I hope that situated data analysis will be a tool that allows us to more easily tease apart and understand how these different ways of situating data can lead to very different experiences and effects of power.

Received: 31 January 2020; Accepted: 6 May 2020;

Published online: 17 June 2020

Notes

- 1 This became particularly visible during the COVID crisis, when local municipalities began using such systems to monitor the movement of residents. For example, in Bergen, Norway, the city purchased a system from Everbridge which shows city administrators a live feed of where people are, anonymized by only showing groups of at least 20 people, and also allows the municipality to send out targeted SMSes to people in certain areas (Valhammar, 2020). Such systems are also used commercially.
- 2 Computational journalism provides some of the best examples of data visualisations that are not automated, but carefully researched and hand-crafted, making clear arguments based on data. For discussions of a range of data visualisations in journalism and other communication practices (see for instance, Engebretsen and Kennedy, 2020; Weber et al., 2018).
- 3 Andrejevic writes that “The code of the machine is purely operational—it does not refer to an absent referent but collapses the signifier into the signified: this is the logic of code and of the operational image/symbol” (p. 128). As I have noted elsewhere, this

ignores the wealth of existing research on the biases in data and how data is a constructed representation rather than the actual world (Rettberg, 2020a).

- 4 Strava’s Global Heatmap can be accessed at <https://www.strava.com/heatmap>. To zoom in, a free account is required.
- 5 Information about Strava Metro can be found at <https://metro.strava.com>.
- 6 See Linda Kronman’s study of how artists work to reveal and highlight biases in the construction of datasets for machine learning for a related discussion of how the biases of a dataset create a machine learning system’s *Umwelt*, or understanding of the world (2020). The discussion of a machine Umwelt, which is drawn from N. Katherine Hayles’ cybersemiotics, an extension of biosemiotics (2019), seems a promising direction to explore in connection with environmentalism and environmental power.
- 7 Although as Sophie Bishop has shown (2020), some YouTubers do very consciously set out to learn the rules and norms that the YouTube algorithms are encouraging.

References

- Aiello G (2020) Inventorizing, situating, transforming: social semiotics and data visualization. In: Engebretsen M, Kennedy H (eds) *Data visualization in society*. Amsterdam University Press, Amsterdam, pp. 49–62
- Ajana B (2017) Digital health and the biopolitics of the quantified self. *Digital Health* 3. <https://doi.org/10.1177/2055207616689509>
- Alix-Garcia J, Walker S, Bartlett A (2017) Nighttime lights illuminate positive impacts of refugee camps. *Refugees Deeply*. <https://www.newsdeeply.com/refugees/community/2017/12/08/nighttime-lights-illuminate-positive-impacts-of-refugee-camps>
- Almklov P, Østerlie T, NTNU Social Research, Haavik T, NTNU Social Research (2014) Situated with infrastructures: interactivity and entanglement in sensor data interpretation. *J Assoc Inf Syst* 15(5):263–286. <https://doi.org/10.17705/1jais.00361>
- Andrejevic M (2019) *Automated media*. Routledge, London and New York
- Barbaro M, Zeller TJ (2006) A face is exposed for AOL searcher no. 4417749. *N Y Times*. <http://www.nytimes.com/2006/08/09/technology/09aol.html>
- Benjamin R (2019) *Race after technology: abolitionist tools for the New Jim Code*. Polity, Cambridge
- Bennett J (2010) *Vibrant matter: a political economy of things*. Duke UP, Durham and London
- Bishop S (2020) Algorithmic experts: selling algorithmic lore on YouTube. *Soc Media+Soc* 6(1). <https://doi.org/10.1177/2056305119897323>
- Bogost I, Montfort N (2009). *Platform studies: frequently questioned answers*. In: Penny S, Tenhaaf N (eds) *Proceedings of digital arts and culture 2009*. University of California, Irvine
- Bolukbasi T, Chang K-W, Zou J, Saligrama V, Kalai A (2016) Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: Lee D, Sugiyama M, Luxburg UV, Guyon I, Garnett R (eds) *Advances in Neural Information Processing Systems* 29. Curran Associates, New York. <https://dl.acm.org/doi/10.5555/3157382.3157584>
- Brock AL (2019) *Distributed blackness: African American cybercultures*. New York University Press, New York
- Brown NM, Mendenhall R, Black ML, Moer MV, Zerai A, Flynn K (2016) Mechanized margin to digitized center: black feminism’s contributions to combatting erasure within the digital humanities. *Int J Humanities Arts Comput* 10(1):110–125. <https://doi.org/10.3366/ijhac.2016.0163>
- Buolamwini J, Gebru T (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. In: Friedler SA, Wilson C (eds) *Proceedings of the 1st conference on fairness, accountability and transparency*. PMLR, New York, pp. 77–91
- Crawford K, Paglen T (2019) *Excavating AI: the politics of images in machine learning training sets*. The AI Now Institute
- Depper A, Howe PD (2017) Are we fit yet? English adolescent girls’ experiences of health and fitness apps. *Health Sociol Rev* 26(1):98–112. <https://doi.org/10.1080/14461242.2016.1196599>
- van Dijck J (2014) Datafication, dataism and dataveillance: big Data between scientific paradigm and ideology. *Surveill Soc* 12(2):197–208
- Downie B (2015) Evidence-based assessment of cycle traffic in Glasgow [Slides]. GoBike!, Glasgow
- Drucker J (2011). *Humanities approaches to graphical display*. *Digit Humanit* Q 5 (1). <http://www.digitalhumanities.org/dhqv/vol/5/1/000091/000091.html>
- Drucker J (2014) *Graphesis: visual forms of knowledge production*. Harvard University Press, Cambridge
- Engebretsen M (2017) *Levende diagrammer og zoombare kart*. *Nor medietidsskr* 24(2):1–27. <https://doi.org/10.18261/issn.0805-9535-2017-02-02>
- Engebretsen M, Kennedy H (2020) *Data visualization in society*. Amsterdam University Press, Amsterdam
- Eubanks V (2018) *Automating inequality: how high-tech tools profile, police, and punish the poor*. Picador, New York
- Feldman A (2019) Of the pointless view: from the ecotechnology to the echo-theology of omnivoyant war. In: Liljefors M, Noll G, Steuer D (eds) *War and algorithm*. Littlefield, London and New York, pp. 165–190

- Foucault M (1988) *Technologies of the self: a seminar with Michel Foucault*. University of Massachusetts Press, Amherst
- Foucault M (2008) The birth of biopolitics: lectures at the Collège de France, 1978–79 (M. Senellart, Ed.; trans. Burchell G). Palgrave Macmillan, Basingstoke
- Friendly M (2008) A brief history of data visualization. In: Chen C, Härdle W, Unwin A (eds) *Handbook of data visualization*. Springer, Berlin, pp. 15–56
- Gabrys J (2014) Programming environments: environmental and citizen sensing in the smart city. *Soc Space* 32(1):30–48. <https://doi.org/10.1068/d16812>
- Gerlitz C, Helmon A, Nieborg DB, van der Vlist FN (2019) Apps and infrastructures—a research agenda. *Comput Cult.* <http://computationalculture.net/apps-and-infrastructures-a-research-agenda>
- Gitelman L (Ed.) (2013) *Raw data is an oxymoron*. MIT Press, Cambridge
- Griffin GP, Jiao J (2015) Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *J Transp Health* 2(2):238–247. <https://doi.org/10.1016/j.jth.2014.12.001>
- Hall S (1997) The work of representation. In: representation: cultural representations and signifying practices. The Open University, Milton Keynes
- Haraway D (1988) Situated knowledges: the science question in feminism and the privilege of partial perspective. *Fem Stud* 14(3):575. <https://doi.org/10.2307/3178066>
- Haugli B (2018) You can see individuals that are using Strava by zooming it to houses that have a short line. Strava gives the ability to set up privacy zones, but it's not on by default. <https://twitter.com/BrianHaugli/status/957448126809493504>
- Hayles NK (2017) *Unthought: the power of the cognitive unconscious*. University of Chicago Press, Chicago
- Hayles NK (2019) Can computers create meanings? A cyber/bio/semiotic perspective. *Crit Inq* 46(1):32–55. <https://doi.org/10.1086/705303>
- Helmond A (2015) The platformization of the web: making web data platform ready. *Soc Media+Soc* 1(2). <https://doi.org/10.1177/2056305115603080>
- Hoel AS (2018) Operative images: inroads to a new paradigm of media theory. In: Feiersinger L, Friedrich K, Queisner M (eds) *Image-action-space: situating the screen in visual practice*. De Gruyter, Berlin and Boston, pp. 11–28
- Hörl E (2018) The environmentalitarian situation (trans: Schott NF). *Cult Politics* 14(2):153–173. <https://doi.org/10.1215/17432197-6609046>
- Jestic B, Nelson T, Winters M (2016) Mapping ridership using crowdsourced cycling data. *J Transp Geogr* 52:90–97. <https://doi.org/10.1016/j.jtrangeo.2016.03.006>
- Johnson JM (2018) Markup bodies: black [life] studies and slavery [death] studies at the digital crossroads. *Soc Text* 36(4 (137)):57–79. <https://doi.org/10.1215/01642472-7145658>
- Kant T (2015) Spotify has added an event to your past: (re)writing the self through Facebook's autostating apps. *Fibrecult J* 25. <https://doi.org/10.15307/fcj.25.180.2015>
- Kristensen DB, Ruckenstein M (2018) Co-evolving with self-tracking technologies. *New Media Soc* 20(10):3624–3640. <https://doi.org/10.1177/1461444818755650>
- Kronman L (2020) Intuition machines: cognizers in complex human-technical assemblages. *APRJA* 9(1). <http://aprja.net>
- Lied H, Svendsen C (2018) How soldiers from Norway, Denmark and USA disclose who they are and where they exercise in war zones. NRK. https://www.nrk.no/urix/how-soldiers-from-norway_-denmark-and-usa-disclose-who-they-are-and-where-they-exercise-in-war-zones-1.13892695
- Light B, Burgess J, Duguay S (2016) The walkthrough method: an approach to the study of apps. *New Media Soc* 20(3). <https://doi.org/10.1177/1461444816675438>
- Losh EM, Wernimont J (eds) (2018) *Bodies of information: intersectional feminism and digital humanities*. University of Minnesota Press, Minneapolis
- Lupton D (2016) You are your data: self-tracking practices and concepts of data. In: Selke S (ed.) *Lifelogging*. Springer, Wiesbaden, pp. 61–79
- Lupton D, Pink S, LaBond CH, Sumartojo S (2018) Digital traces in context: personal data contexts, data sense and self-tracking cycling. *Int J Commun* 12. <https://ijoc.org/index.php/ijoc/article/view/5925/2258>
- Marino MC (2020) *Critical code studies: initial methods*. MIT Press, Cambridge
- M'charek A, Toom V, Jong L (2020). The trouble with race in forensic identification. *Sci Technol Hum Values* 1–25. <https://doi.org/10.1177/0162243919899467>
- Mullainathan S, Obermeyer Z (2017) Does machine learning automate moral hazard and error? *Am Econ Rev* 107(5):476–480. <https://doi.org/10.1257/aer.p20171084>
- NASA (2012) *Night Lights 2012 Map*. <https://earthobservatory.nasa.gov/images/79765/night-lights-2012-map>
- Noble SU (2018) *Algorithms of oppression: how search engines reinforce racism*. NYU Press, New York
- Norwegian Consumer Council (Forbrukerrådet) (2020). Out of control: how consumers are exploited by the online advertising industry (p. 186). Norwegian Consumer Council (Forbrukerrådet). <https://www.forbrukerradet.no/out-of-control/>
- O'Neil C (2017) *Weapons of math destruction, how big data increases inequality and threatens democracy*. Penguin, London
- Paglen T (2014) *Operational images*. E-Flux J 59. <https://www.e-flux.com/journal/59/61130/operational-images/>
- Pink S, Sumartojo S, Lupton D, Bond CHL (2017) Mundane data: the routines, contingencies and accomplishments of digital living. *Big Data Soc* 4(1). <https://doi.org/10.1177/2053951717700924>
- Plantin J-C, Lagoze C, Edwards PN, Sandvig C (2018) Infrastructure studies meet platform studies in the age of Google and Facebook. *News Media Soc* 20(1): 293–310. <https://doi.org/10.1177/1461444816661553>
- Retberg JW (2009) 'Freshly generated for you, and Barack Obama': how social media represent your life. *Eur J Commun* 24(4):451–466. <https://doi.org/10.1177/0267323109345715>
- Retberg JW (2014) Seeing ourselves through technology: how we use selfies, blogs and wearable devices to see and shape ourselves. Palgrave, Basingstoke
- Retberg JW (2018a) Apps as companions: how quantified self apps become our audience and our companions. In: Ajana B (ed.) *Self-tracking: empirical and philosophical investigations*. Palgrave Macmillan, pp. 27–42
- Retberg JW (2018b) Sublime visualisations of immensely big data [Blogpost]. From Numbers to Graphics. <https://sciencenorway.no/blog-blog-from-numbers-to-graphics-statistics/sublime-visualisations-of-immensely-big-data/1618855>
- Retberg JW (2020a) Book review: automated media. *Convergence*. <https://doi.org/10.1177/1354856520906610>
- Retberg JW (2020b) Ways of knowing with data visualizations. In: Engebretsen M, Kennedy H (eds) *Data visualization in society*. Amsterdam University Press
- Risam R (2019a) New digital worlds: postcolonial digital humanities in theory, praxis, and pedagogy. Northwestern University Press, Evanston
- Risam R (2019b) Beyond the migrant "Problem": visualizing global migration. *Telev News Media* 20(6):566–580. <https://doi.org/10.1177/1527476419857679>
- Romanillos G, Austwick MZ, Ettema D, Kruijff JD (2016) Big data and cycling. *Transp Rev* 36(1):114–133. <https://doi.org/10.1080/01441647.2015.1084067>
- Ruser N (2018) Strava released their global heatmap. 13 trillion GPS points from their users (turning off data sharing is an option). [https://medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de... It looks very pretty, but not amazing for Op-Sec. US Bases are clearly identifiable and mappable. <https://twitter.com/Nrg8000/status/957326421684207616>](https://medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de...)
- Sanders R (2017) Self-tracking in the digital era: biopower, patriarchy, and the new biometric body projects. *Body Soc* 23(1):36–63. <https://doi.org/10.1177/1357034X16660366>
- Sareen S, Saltelli A, Rommetveit K (2020) Ethics of quantification: illumination, obfuscation and performative legitimation. *Palgrave Commun* 6(1). <https://doi.org/10.1057/s41599-020-0396-5>
- Selala MK, Musakwa W (2016) The potential of strava data to contribute in non-motorised transport (nmt) planning in Johannesburg. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*; Gottingen, XLI-B2, 587–594. <https://doi.org/10.5194/isprs-archives-XLI-B2-587-2016>
- Smith WR, Treem J (2017) Striving to be king of mobile mountains: communication and organizing through digital fitness technology. *Commun Stud* 68(2):135–151. <https://doi.org/10.1080/10510974.2016.1269818>
- Stacey J, Suchman L (2012) Animation and automation—the liveliness and labours of bodies and machines. *Body Soc* 18(1):1–46. <https://doi.org/10.1177/1357034X111431845>
- Strava (2019) *Strava|Metro—Web Product Feature Walkthrough*. YouTube. <https://youtu.be/9wHi9S-M1HY>
- Suchman L (1987) *Plans and situated actions: the problem of human-machine communication*. Cambridge UP, Cambridge
- Suchman L (2007) *Human-machine reconfigurations*. Cambridge UP, Cambridge
- Taylor A (2016) Data, (bio)sensing and (other-)worldly stories from the cycle routes of London. In: *Quantified: biosensing in everyday life*. MIT Press, Cambridge, pp. 189–209
- Tufte ER (1990) *Envisioning information*. Graphics Press, New York
- Valhammar R (2020) *Direkteanskaffelse—Lokasjonsbasert varslingsverktøy [Byrådsak, Bergen kommune]*. Bergen Municipality. <https://www.bergen.kommune.no/politikere-utvalg/api/2870932/Framstilling-Direkteanskaffelse-Lokasjonsbasert-varslingsverktoy>
- Weber W, Engebretsen M, Kennedy H (2018) Data stories. Rethinking journalistic storytelling in the context of data journalism. *Stud Commun Sci* 18(1). <https://doi.org/10.24434/j.scoms.2018.01.013>
- Zheng Y (2015) Trajectory data mining: an overview. *ACM Trans Intell Syst Technol* 6(3). <https://doi.org/10.1145/2743025>
- Zuboff S (2019) *The age of surveillance capitalism*. Profile Books, London

Acknowledgements

This research was funded by the research project *Innovative Data Visualisation and Visual-Numeric Literacy*, which has received funding from The Research Council of Norway (Grant no. 259536) and by the *Machine Vision in Everyday Life* project, which has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement no.

771800). I would also like to thank the many colleagues who have responded to various versions of this paper.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.W.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020