Black Lives Matter: U.S. Protesters Tracked By Secretive Phone Location Technology

Zak Doffman ⋮ 6-7 minutes ⋮ 6/26/2020

Los Angeles Times via Getty Images

Following Black Lives Matter protests in cities across the U.S., a marketing company that uses AI to categorize phone users by race, gender and even religion, has now published a report using phone location data secretly collected during those protests. In a presentation titled “George Floyd Protester Demographics: Insights Across Four Major U.S. Cities,” consumer insight firm Mobilewalla says it has published its report “out of a sense of social responsibility.”

Mobilewalla buys mobile phone data from aggregators that, in turn, buy their data from the operators behind the apps we install on our phones. The company says that it has 80-90% device coverage in the U.S., fairly evenly split between iOS and Android. It also says that at any given event, it can usually poll devices from somewhere between 30% and 60% of those present.

“Normally we’re fairly introverted,” Mobilewalla CEO Anindya Datta told me, “this is one of the rare occasions we publish a report like this.” And while claiming it was socially responsible to do so, because the data suggested “the vast majority of protesters were white and marched within their own cities,” Datta acknowledged that such a report would inevitably trigger a response, given the backlash against surveillance and seemingly secretive phone data collection.

Mobilewalla says its dataset, which Datta told me “is never real time—we can store hundreds of petabytes of historic data,” covers 1.6 billion devices and 25 billion daily data points globally. As for this report, the company says it is based on “devices observed for geographic areas around specific protest locations in New York, Los Angeles, Minneapolis and Atlanta,” both during the day and at night, across three days at the end of May.

The data itself focuses on gender, ethnicity and those it infers come from inside versus outside the city, where phones with a “common evening location within a 150km radius of city center” are considered to be “inside.” The vast majority of tracked phones, unsurprisingly, were from inside the relevant cities. Although Atlanta, the firm says, saw more than one in five phones coming from outside.

Mobilewalla

Most protesters were males, in all cities, and the clear majority were under 34 years of age, again in all cities. The report then steps into ethnicity distribution—the vast majority of tracked phones belonged to “caucasian/ other.” Despite the percentage of African Americans being much lower, the report then also presents its “gender distribution” for African Americans, but not for other ethnicities.

I asked Datta about the accuracy of the gender and ethnicity classification. “We do a lot of AI to assign attributes to individuals,” he explained. “For instance... if you use a female centric app once or twice you may or may not be a female, but if you use it 600 times over 10 months the likelihood goes up.”

Mobilewalla

As for ethnicity, Datta told me “we look at signals—there are a small number of signals that make the likelihood the person is of a certain class, ethnicity or gender or anything. We gather data where we know the class and use AI to extract patterns and project those patterns on others, it’s all probabilistic.” It is impossible to know how accurate the data is, although Datta says the firm runs controlled tests for commercial customers to prove out its models.

Mobilewalla

None of those being tracked had any idea at the time, nor do they know now. Mobilewalla says that “the

information is based on mobile device data observed and derived from [its] consumer data set. The data is aggregated based on mobile devices present at protest locations and then cleansed and de-identified in accordance with applicable privacy and consent regulations."

Datta told me that all the data, which it buys and does not collect itself, is "opt in," but when challenged on this, he accepted that this "opt in" relies on so-called permission abuse, where apps gather way more data than they need to function. This data includes location pings, access to our contacts, web activity, phone information and even hardware if they have sought the right permissions.

Last year, to make the point, one project analysed simple flashlight apps to show just how pervasive this has become. Of the 937 Android flashlight apps tested, 180 requested permission to access our contacts and 131 our precise locations. This is an issue that impacts Android much more than iOS, although there are iPhone apps that collected our data as well. But Apple has been cracking down, and thankfully Google has now announced plans in Android 11 to do the same.

By way of illustration, Mobilewalla says in its privacy policy that this is the kind of data it may receive: "Device type and model, network provider and connection type, browser type, language, operating system, cookie identifiers, mobile carrier, mobile IDs, MAC address, IMEI, unique identifiers, hardware type, device specifications, operating system, IP address, location, Internet service provider."

The usual defense for the collection of this kind of unique but anonymized data, is that it cannot be attributed to individuals. But previous media investigations have shown that by cross-referencing anonymized locations with other data, you can tie phones to people. How many other people work and live in the same building as you? How many people follow your same travel patterns?

In a week where Apple and Google have announced changes to better protect the privacy of billions of users, this news shows just how invasive the leakage of data from our phones has now become, and why those changes are so badly needed.