Breelyn Betts

Dr. Johnson

CMSI 401

7 October 2020

Assignment #1

Stacey Gray’s article “A Closer Look at Location Data: Privacy and Pandemics” explores the potential consequences of allowing government access to mobile location data collected by major technology companies. The emergence of a global pandemic has spiked interest in this data as a means to inform individuals when they have come in contact with the virus as well as assess the effectiveness of social distancing. While these insights might be beneficial to reducing the spread of COVID-19, the article introduces the potential side effects of relinquishing this data. Researchers requesting this data would require precise “information such as ‘works in the same building’ or ‘attended the same restaurant at the same time as a diagnosed person’ … precise enough to single out an individual with reasonable specificity.”[1] Because device movement mirrors its owner, the data could reveal information about a person’s identity, behavior, associations and activities.

Location data is commonly collected using the following methods: GPS, Cell Towers, Wi-Fi Networks, and Bluetooth Beacons. Modern cell phones use a combination of these methods to better infer where devices are located, increasing location accuracy. This accuracy of device location (and thus the device owner’s location) raises concerns due to the difficulty to fully anonymize data. Some common methods of “anonymizing” data involve replacing names

[1] Gray, Stacey. “A Closer Look at Location Data: Privacy and Pandemics.” Future of Privacy Forum iCal, March 25, 2020. https://fpf.org/2020/03/25/a-closer-look-at-location-data-privacy-and-pandemics/?mkt_tok=eyJpIjoiWXpCbFpHUTJNVGRsTVdSaiIsInQiOiJLSjVjSFwvMnI0TVI4bExcLzJ5aGJ0VHd3VytEMTErbStacUdRdTZRZitQTXVuT2lkUTVRSmZEa2xZZEJLN3FhNHlncjFxTjFFWWJhdEs1alpjUkdnazFaRmRLVkFYU2JPVForWm1mbnFUcENJWmRWaUEzZ1EyK1wvb01MSXo4cno0WCJ9.

with unique identifiers, however behaviors such as where a device remains at night, is enough to easily identify and distinguish people. Additionally, the location data in question is not fully representative of the population, leaving out the individuals who do not own cell phones. Another concerning consequence of sharing this data is that once shared with a public health agency, it will be difficult to limit the future uses of it. A similar article, "Mobile Location Data and Covid-19: Q&A", highlights some of the potential human rights abuses that may occur if this data is not properly protected. In the past, well intended data-driven technologies "such as surveillance measures put in place to counter terrorism, shows that they often go too far, fail to have their desired effect, and, once approved, often outlast their justification."[2] Once granted access to this data, it would be difficult for major technology companies to relinquish access to it. There is also no indication of how effective the research on the data would be in preventing and lowering the spread of the virus. The question remains, do the risks outweigh the potential rewards?

This article relates to Software Engineering in many aspects. In particular, the level of specificity the article outlines subtly indicates subsequent data (in addition to mobile location data) that would be necessary to make such insightful inquiries. In order to accurately provide predictions to who was exposed to the virus, the researchers would need to know who actually has the virus. In order to know who has the virus, the researchers would need access to medical records or COVID-19 test results. This raises the question: Does a person's positive infection status entitle the researchers to access all their personal data on their whereabouts and life patterns? How does this comply with HIPAA laws and overall privacy? It appears that granting

[2] "Mobile Location Data and Covid-19: Q&A." *Human Rights Watch*, May 13, 2020. https://www.hrw.org/news/2020/05/13/mobile-location-data-and-covid-19-qa.

access to mobile location data is simply the first step in a slippery slope of the data that would be deemed “necessary” to make an accurate prediction of who was exposed to the Coronavirus.

Additionally, marginalized and vulnerable communities who cannot afford cell phones are not included in the mobile location data. As a result, any conclusions drawn from the data will not be equitable. This is especially problematic in this current pandemic; the same group disproportionately vulnerable to infection is also the group that is not accurately represented in the dataset. With affluent individuals more represented in the dataset, they are more likely to receive resources and suggestions constructed from the data. It can be argued that any reduction in the spread of COVID-19 is good, but neglecting marginalized and vulnerable communities is an oversight that can have devastating consequences. As programmers, it is important to recognize the lack of representation that datasets may contain. Taking it a step further, it is then even more important to attempt to compensate for these disparities. There will always be some sort of bias in the data we work with and it is our responsibility to identify it.

Even as more questions arise in terms of data security, big reputable companies still are not questioned enough by everyday users. This is seen mainly in the lack of consideration given to the information they are choosing to share with applications. I too am guilty of agreeing to the terms and conditions of an application without checking what that means for data collection. While not all data is sensitive, people may be sharing more than intended or desired. It is important to note that users should not have to be vigilant about keeping their data private and anonymous, this responsibility should fall on the companies collecting the data. It is our job as programmers to make sure that users are aware of the data they are providing to an application. Developers and companies appear to be moving in this direction with well-known applications like Instagram allowing users to limit the number of photos they share, instead of requiring full access to the phone's gallery.

While data analytics are valuable to companies and people in general, it is important to keep in mind that there are biases present in every dataset. We have an obligation as programmers to identify and mitigate these biases in order to reduce any potential harm or discrimination. As the article implicates, it is important to explore the ethical repercussions of selling and using data. There are few federal laws in place to regulate the exchange of data. Within the realm of software development, it is important to consider the type of data that is collected in the systems we develop and how we use the data. Users have an expectation that the sensitive information they provide is protected, so we have the obligation to ensure that it is.

Article: A Closer Look at Location Data: Privacy and Pandemics