*How might some of the issues outlined by Zimmer arise when collecting and analyzing data from Twitter?*

The issues that Zimmer outlines in his article around the OKCupid study also appear when collecting and analyzing data from Twitter. Twitter is a very public forum where millions of users each day can post, reply, and “retweet” other users all on the platform. While users can set their profiles to private and limit interaction to just those who follow them, most remain public. Further, and perhaps unknowingly, users can have their information pulled through use of APIs and used in various data analytic work. Zimmer notes throughout that just because data is public does not mean that users have consented to their information being used in reporting or analyses. Just because data is available publicly does not remove a researcher’s ethical responsibility in ensuring that the data is handled appropriately and with the proper guardrails in place.

I agreed with the author’s argument and push back on the OKCupid researcher’s assertions throughout the article. Even if Twitter makes users aware of how their data may be used in Terms of Service or other agreements, the researchers must bear the “burden” of ensuring that the data is obtained with best practices around consent. The biggest concern in the OKCupid research in my mind was around the lack of anonymity given to the subjects in the dataset. Even if the OKCupid researcher is correct in that the data is public (which is still not an excuse for a lack of obtaining consent) harm was by no means minimized with their approach. In thinking about the Twitter API and Tweepy, user information tied to tweets is readily available, and careful considerations would need to be applied in any data analytic work that looks to use this information. Anonymization is a very clear path forward in reducing harm when this “public” information is pulled through in any reporting.

In consideration of our Week 1 programming assignment, and having reviewed the Twitter API documentation, it’s very clear how powerful the API is and the wealth of information that can be obtained. Most individuals today are on at least one, if not multiple, social media platforms. With such a wide population to pull from, the advantages of using social media in reporting are clear when leveraged for various sociological or other analyses. However, ethics should not be ignored just due to the ease of information access. In some of my prior work experiences within the healthcare space, specifically clinical trial recruitment in hospitals, consent AND an understanding of what the trial is, what it’s measuring, how the data will be used, etc. were all absolutely critical to obtain from a potential test subject before they were onboarded into the trial. These types of strict guardrails should not just exist within the healthcare space but should help govern how data science and analytic work is handled moving forward.