

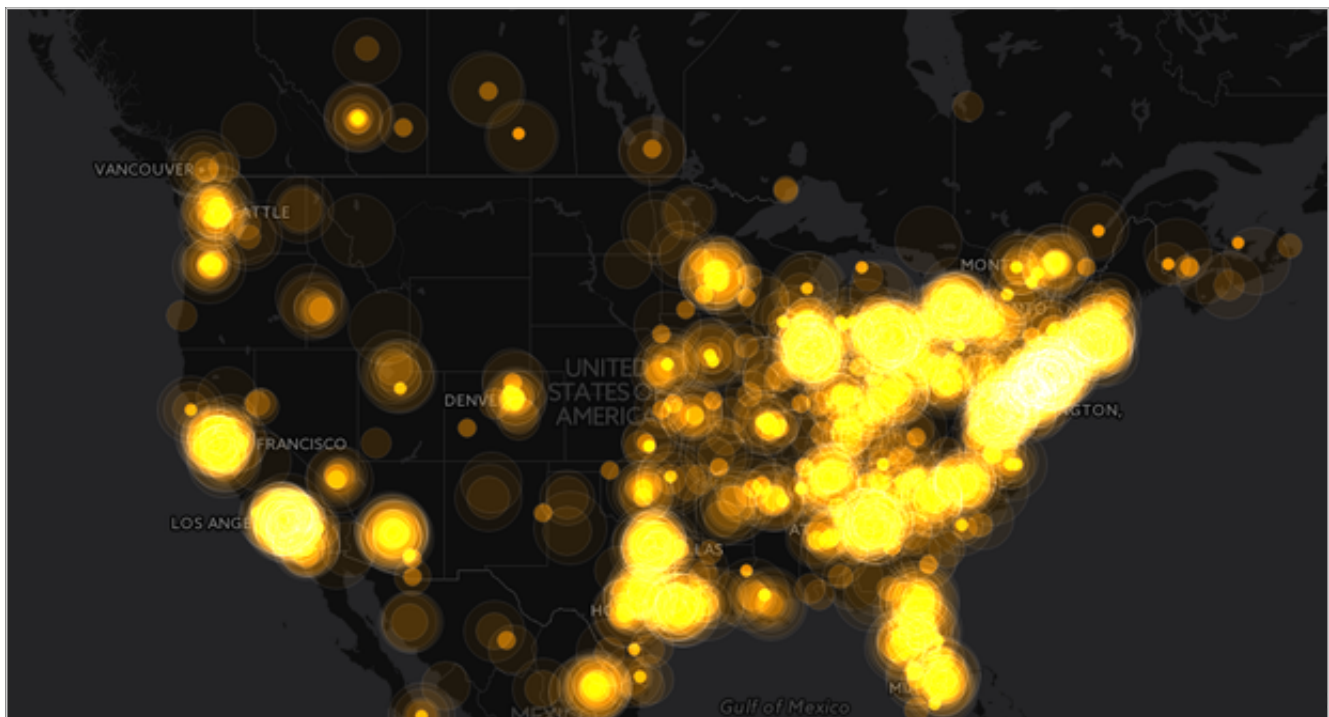


MAPS

Why Most Twitter Maps Can't Be Trusted

A sophisticated age demands a more sophisticated social media cartography.

AARIAN MARSHALL |  @AarianMarshall | Mar 26, 2015 |  4 Comments

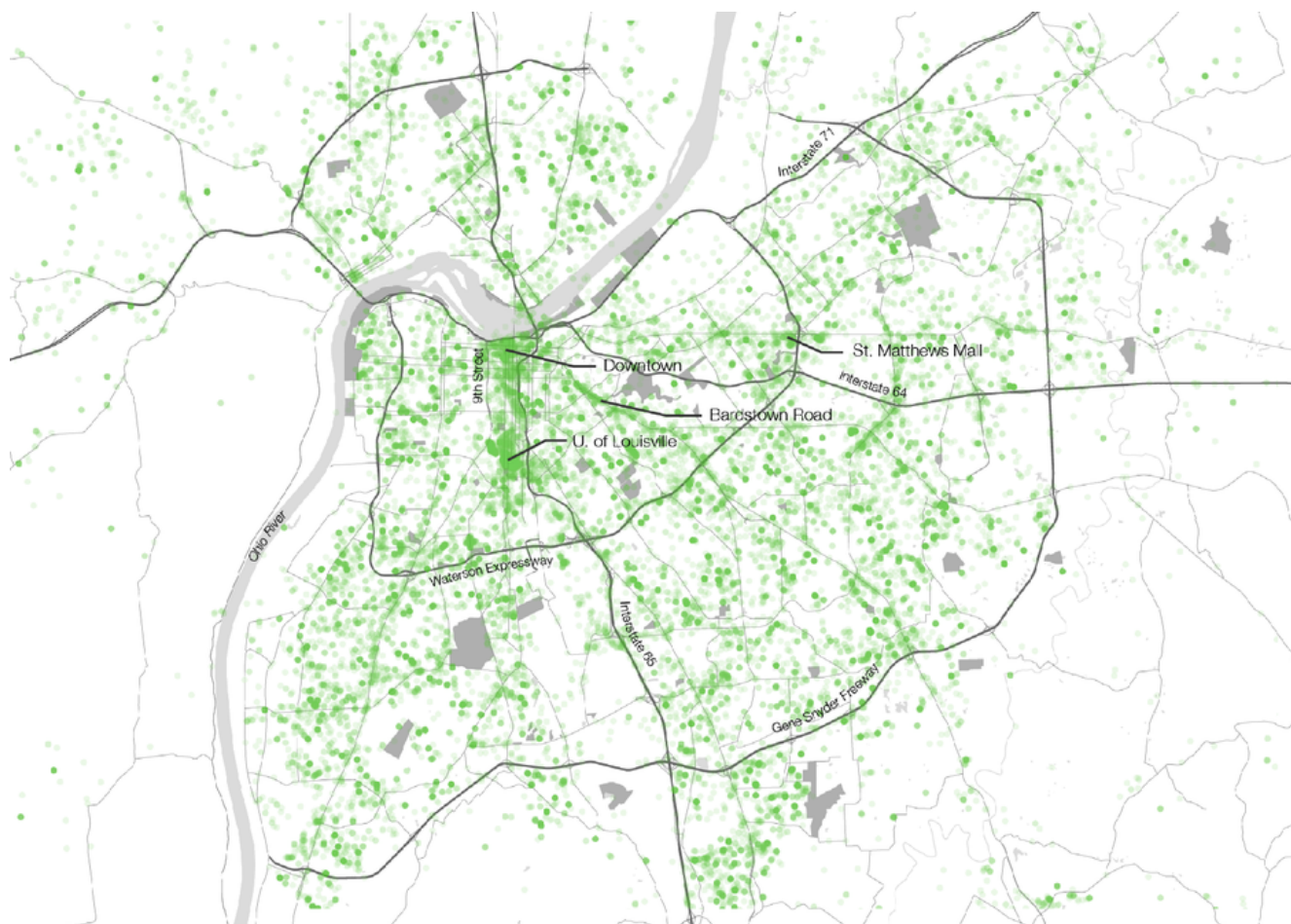


[CartoB](#), via [Twitter](#)

When geographers [Taylor Shelton](#), [Ate Poorthuis](#)* and [Matthew Zook](#) decided to study the tricky phenomenon of race-based spatial segregation in Louisville, Kentucky, they had a few options. They could have, as Poorthuis put it in an interview, "set up an entire study [and] given [Louisville residents] GPS trackers and diaries." But "that would require 18 months of planning and lot of money," he says. "Too much money."

So the geographers went the cheaper, faster route: Twitter data. Specifically, they pulled data from 5.7 million tweets geotagged to Louisville, Kentucky, between June 2012 and July 2014.

Here's what happened when the geographers initially mapped Louisville's tweets. The map below, from their [upcoming paper](#) in *Landscape and Urban Planning*, shows a 1 percent random sample of all geotagged Louisville tweets from 2013. To the west of Ninth Street, labeled on the map, is the majority-black West End. In the minds of many white Louisvillians, the researchers write, the West End is a world "fundamentally separate and apart." [The area has a](#) median household income of \$21,700—63 percent lower than the city's mostly white East End. Thirteen percent of West End residents are unemployed, and just 7 percent hold Bachelor's degrees. Compare that to the East End, where only 4 percent are unemployed and 55 percent have four-year college degrees.



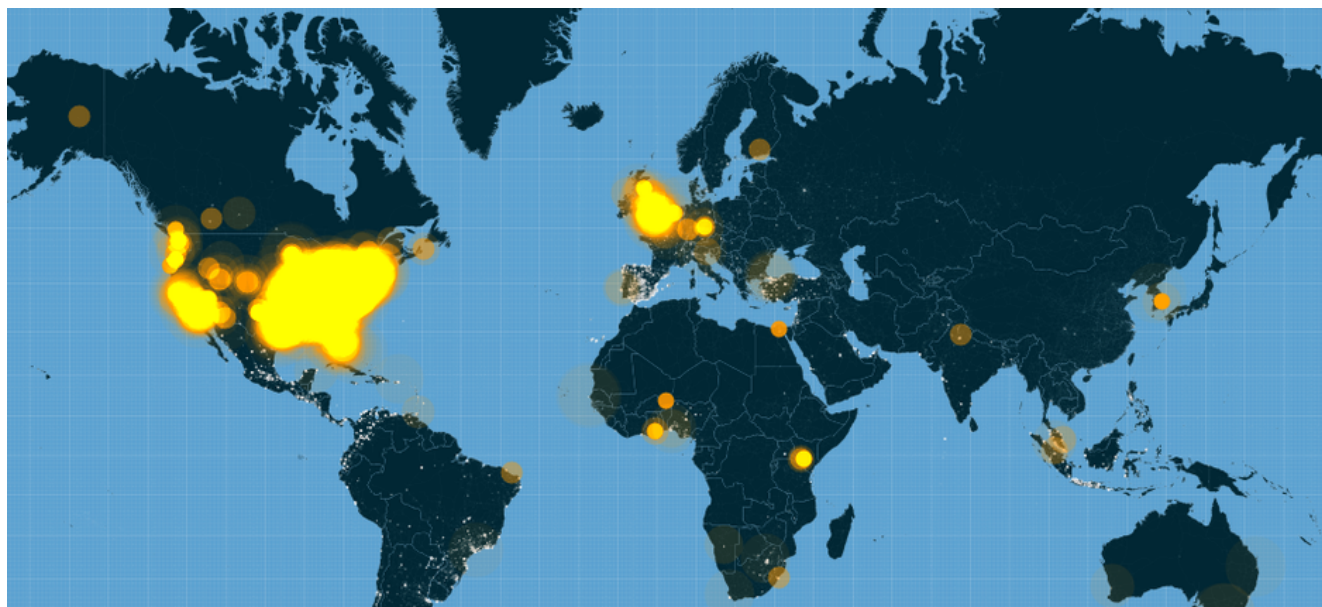
(Shelton, Poorthuis, Zook)

What sorts of conclusions can one draw from this map? To start, it looks like residents west of Ninth Street aren't tweeting nearly as much as those to the east. So maybe, a social scientist might posit, there are simply fewer Twitter users in the West End. And maybe this is yet more evidence of the "[digital divide](#)," where those who live in the nation's poorest and least advantaged neighborhoods are locked out of the fast-paced digital age without access to

high-speed Internet. Maybe some of the West End's problems could be solved if Louisville just gets the West End online.

This sort of Twitter mapping technique is popular, and probably reached its zenith last summer, when a geotagged map of users tweeting the hashtag "Ferguson" [itself went viral](#). Maps like these seem to illuminate broad socio-cultural responses to big-time national events. "Look at how many people care about Ferguson!" is the implicit "geez, whiz" message of that particular Twitter-generated map. But as Shelton, who worked on the Louisville study, points out, this technique has serious flaws.

For one, keep in mind that relatively few of us are actually on Twitter to begin with—just 23 percent of Americans, according to the [Pew Research Center](#). But even more importantly, Shelton says, Twitter mappers often fail to normalize their data, meaning that many Twitter maps are less representations of deep, social phenomena and more depictions of population patterns. The Ferguson map, for example, doesn't meaningfully diverge from "typical tweeting," Shelton says. Anything that goes viral—be it #Ferguson, #Obamacare, or #BachelorNation—will look similar as it "trends" and lights up a Twitter map.



#Ferguson lights up Twitter on August 12, 2014. ([CartoDB](#))

Furthermore, this "points on a map" approach doesn't teach the map viewer anything about tweet density. What looks like one point on a map may actually be 14 on top of each other. Where does one tweet begin and another end? And how do researchers adjust for spambots, which oftenglom onto trending

Twitter topics to auto-promote their wares? What about "power users," people who obsessively tweet with the same hashtag over and over and over again?

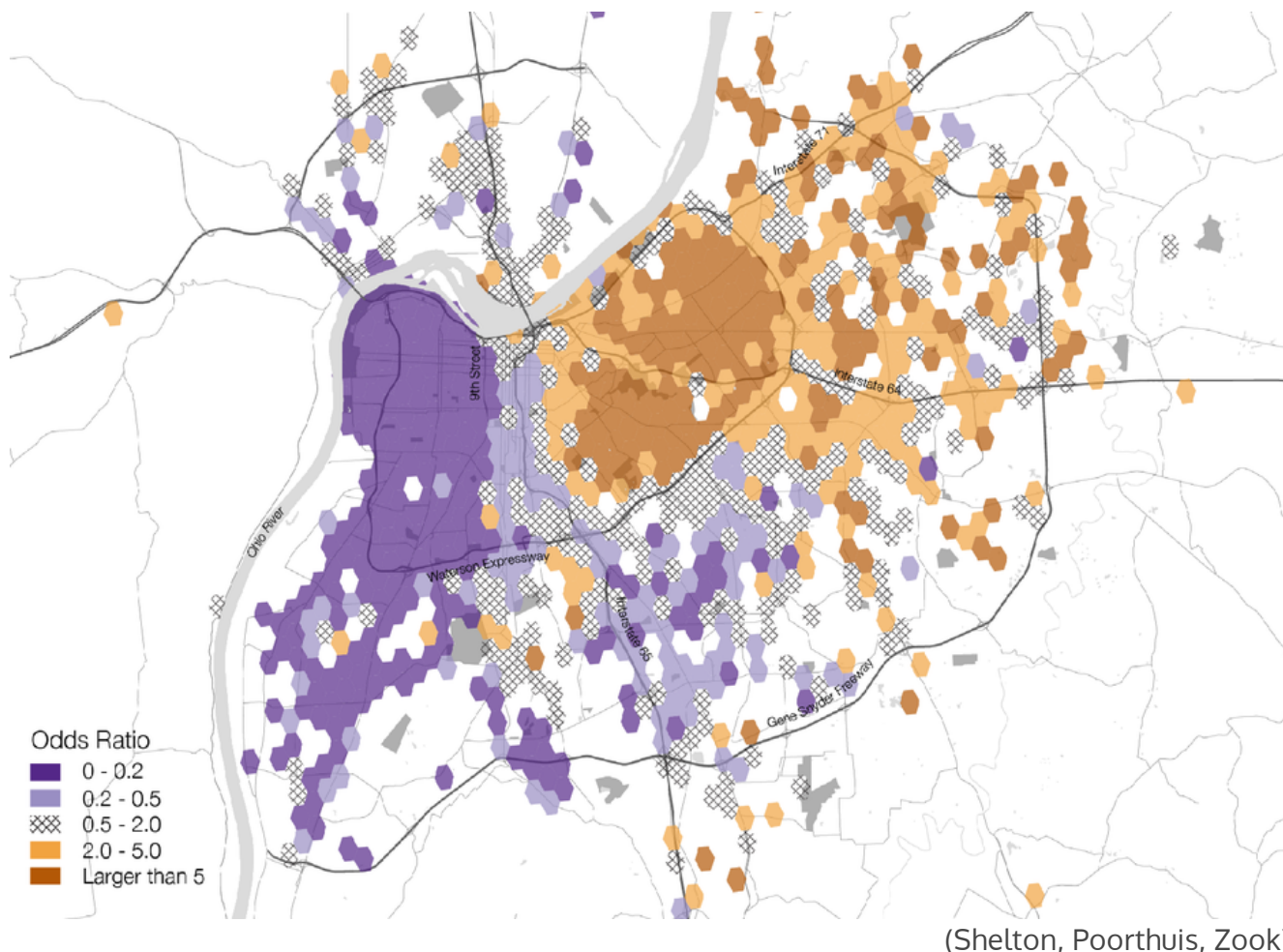
The problem with Twitter maps isn't that social media data is inherently flawed—it's that the people who make them get lazy.

Recognizing these flaws, the geographers tried to add a little more context to their study of Louisville. First, they traced the geotagged tweets to discover where specific users spent most of their time, West or East Louisville. And instead of mapping with dots, they chose gradated hexagonal areas. As they explain in the study:

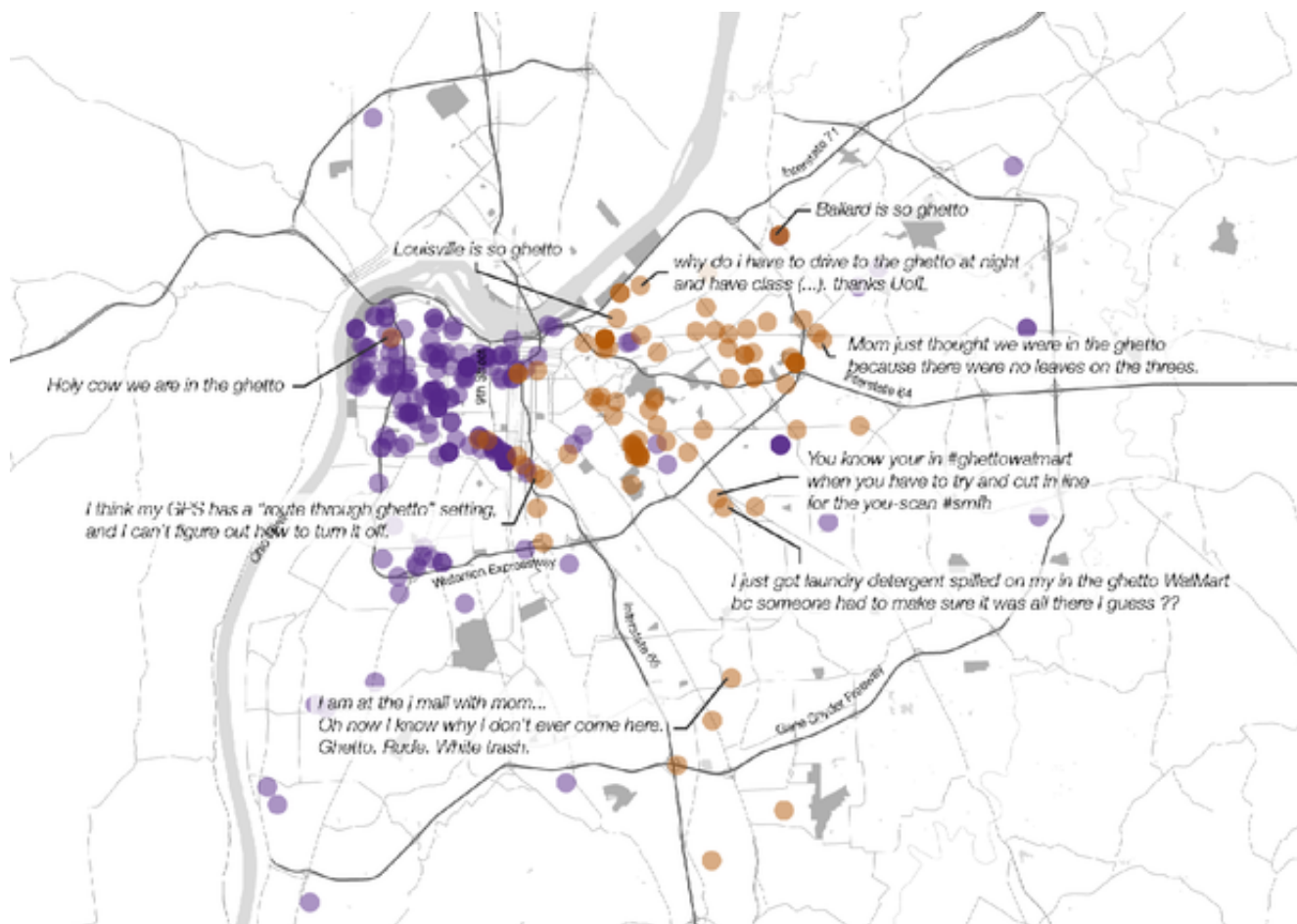
[W]ithin the larger dataset, one user created 65 tweets from the area around 2nd and Market Streets in Louisville in one six hour period but never again tweeted from this area. Unadjusted, this activity would give equal weighting to each of these 65 tweets as to the tweets of individuals who travel regularly to this place, or individuals who only visit once but produce a much smaller amount of content.

To correct for this, the researchers chose to map a maximum of five randomly selected tweets per user in any given hexagon.

Below is their corrected map, with the patterns of predominantly West End residents in purple and East End residents in orange. The areas in grey are those where users from both areas tweet almost equally.



Compared to the raw data map above, this technique leads to nearly the opposite conclusions. Louisville's West End residents are actually moving around and tweeting *a lot*. In fact, West End residents are much more likely to enter the East End than East End residents are to go West. That all-important "Ninth Street divide," which seemingly cuts the city into two separate and unequal worlds? It now looks more like a very porous border. And looking beyond the geotags themselves to examine the *content* of East End users' tweets reveals even more.



(Shelton, Poorthuis and Zook)

"Holy cow we are in the ghetto," one East End user writes after they cross over the Ninth Street line. Even more interestingly, many East End users tweeted they were "in the ghetto" while well within the city's predominantly white and affluent areas. "Ultimately," the researchers conclude, "these kinds of incongruencies demonstrate the more complex relationship between urban spatial imaginaries and the everyday activity spaces of individuals and collectives as demonstrated through geotagged social media data." Put more simply: Things are not always as unsophisticated Twitter maps make them seem.

The problem with Twitter maps, then, isn't that social media data is inherently flawed—it's that the people who make them get lazy. "[When] you have these giant Twitter datasets ... it's very, very easy to get that view from above and let the data speak for itself and just sort of stop there," says Poorthuis. "That's not the right stopping point. You need to contextualize by looking at the data in more detail—the variables and dimensions combined with the local knowledge."

"It's 2015 now," Poorthuis says. "It was cool and an engineering challenge to get these points on a map. But now it's time to ask deeper and more meaningful questions."

**CORRECTION: A previous version of this article misspelled the name of geographer Ate Poorthuis.*

About the Author



Aarian Marshall is an editorial fellow with *CityLab*.

ALL POSTS | [@AarianMarshall](#)



American Express® Presale
Now - April 3 at 10pm

GET TICKETS