You can’t find ‘super-spreader’ businesses with old GPS data

Today, as America faces a difficult debate about “reopening” state economies during the coronavirus pandemic, *The New York Times* seemingly offered a helpful piece of guidance. The paper’s opinion section published a visualization of how people engage with different businesses like bars and gyms, estimating their risk of infection at each. Alarmingly, it dubbed some of these places dangerous “super-spreader businesses” — warning that “through the lens of contagion, a yoga class, a busy corner store, or a crowded neighborhood bar may look a lot like a wet market in China.”

But at least in this story, the real worry isn’t contagion. It’s the sweeping assumptions being trained on some very limited data. And it doesn’t say much about whether your local bar can avoid serving coronavirus with its Coronas. That’s because the story doesn’t include any infection data.

Location data can track patterns, but not whether they’re still relevant

The *Times* story is written by four professors at the University of Chicago, who aggregated anonymized phone location data from April of 2019. Using that data, they plotted how people flowed through locations that corresponded to businesses — learning which kinds of stores saw customers linger the longest, which drew the biggest crowds, and which squeezed the most people into the smallest spaces. They bolstered that data with surveys asking how much participants interacted with other people or touched shared surfaces in these places.

The tracking project produced some interesting factoids. Denny's and the Original Pancake House served similar numbers of people for similar amounts of time, but the latter saw a breakfast rush that produced more crowding. People stayed twice as long at electronics stores as lawn and garden stores — the authors speculate that "you don't linger over fertilizer."

But data from a year ago might tell us very little about the present or future. Many of these businesses have either been closed for weeks or drastically altered their operations. They'll probably reopen (if they reopen at all) under restrictions that render a lot of these observations moot, like spacing rules that will stop those large brunch crowds or arrangements that decrease close-range interactions and shared surfaces.

Even assuming GPS records perfectly track who's inside or outside a building — not a given — we don't know if they reflect how post-pandemic consumers will behave. People might maintain those old browsing styles, but they might also consciously avoid lingering anywhere, or the coronavirus might produce unpredictable new patterns. (For all we know, if everybody's planting a COVID-19 victory garden, maybe they *will* linger over fertilizer.) People's safety doesn't hinge on how they behaved in stores and restaurants a year ago, but about how well a given business can adapt to promote new habits and minimize their risk.

Without infection data, naming 'super-spreading' businesses is a huge stretch

The authors acknowledge that limitations exist, and they note that "these data alone cannot tell us which businesses to open first, and we can't simplify all these different metrics into a 'yes' or 'no' decision on any single business." But this uncertainty doesn't make it into the overall framing of the story, particularly the claim that certain businesses are "super-spreaders" — which is not just an oversimplification,

but an absurd rhetorical stretch.

“Super-spreader” is normally reserved for a person or place that’s linked to dozens or hundreds of COVID-19 cases, including a funeral in Georgia, a party in Connecticut, and numerous meatpacking plants, nursing homes, and prisons. The *Times* editorial co-opts it for businesses that *might* be risky compared to other stores and restaurants. And it does this without using hard data to establish how much each of its risk factors practically matters, not just whether they hypothetically make infection more likely. We have no idea how good the assumptions of this model are.

In fact, it’s extrapolating danger based on a painfully incomplete knowledge of the coronavirus. Epidemiologists know the mechanics of how the virus can spread, but they’re not sure what’s specifically driving many infections and how they’re affected by a multitude of environmental factors. Countries’ rapidly-expanding contact tracing programs can help solve that mystery by finding links between cases. But beyond facts like extreme crowding and poor ventilation being dangerous, there’s not a precise risk assessment right now for different densities and exposure times.

There are huge gaps in our understanding of coronavirus
It’s normal for research and reporting to include simplifications or caveats. And mis-evaluating Red Lobster’s pandemic fitness is less pernicious than, say, promoting drugs that probably don’t work. But it’s bad to pitch simple conclusions (like “bookstores are high-risk” or “florists are safe,” to name two of the model’s assertions) as policy recommendations when they’re backed up by so little evidence and so much speculation, especially when people are grasping for advice about a deadly and often baffling disease.

Warnings — or reassurances — about coronavirus can become conventional wisdom even if they’re mostly speculative. One pre-print research paper with flashy graphics, for instance,

seemed to show joggers and cyclists barreling down streets in a vast miasmatic breath cloud. The paper was an aerodynamic simulation that didn’t establish whether the air could actually infect bystanders. But it was still widely shared with warnings against exercising outside. And even that paper didn’t go as far as calling runners “super-spreaders” — which, as *Guardian* reporter Julia Carrie Wong noted on Twitter, is an incredibly loaded term for businesses that aren’t accused of spreading anything.

Amid a lot of bad-faith attacks on media during the pandemic’s first several weeks, there’s a fair critique that news outlets have promoted research and recommendations far too authoritatively, instead of reflecting the incredible confusion among experts. Meanwhile, Silicon Valley critics have cautioned against overemphasizing high-tech solutions to the disease — including the *Times* itself, which warned that putting too much stock in tracking data might simply “provide an opening for technologists to oversell what they do.”

Overselling is exactly the problem here. This kind of location data could provide some limited guidance for businesses. But if, as the *Times’s* headline puts it, you’re curious if it’s safer to visit a coffee shop or a gym? These clean charts don’t hold the answer.

https://www.theverge.com/2020/5/6/21249404/covid-19-coronavirus-super-spreader-business-gps-data-model-flawed