



# Opportunities and Potential Bias in New Transportation Data

By Lia Cattaneo March 21, 2018

Technology companies that provide mobility services, such as ride-sharing or navigation, are awash in widespread, real-time transportation data. Uber, Lyft, Waze, Apple, TomTom, and others have gathered massive amounts of data about the behavior of people and vehicles across the country. They know how long it takes people to drive from home to work in the morning, which roads are most congested, where and when accidents occur, and more.

Some of these companies are making data available for free to the public sector: Uber recently launched Uber Movement and Google-affiliate Waze provides data to users through its Connected Citizens Program.<sup>1</sup> Uber Movement aggregates and anonymizes billions of trip records from various locations across the world, including four U.S. cities, providing travel times between census tracts or traffic analysis zones for specific dates or average days. Waze provides anonymized user-generated alerts, including information on weather, accidents, and traffic jams.

These companies, and others, have significantly more data than they have offered to provide—including travel speed on specific roadway segments, police activity, occupancy rates, and specific ride-sharing information—but the data that are available still offer unparalleled insights into the transportation system. Importantly, transportation planners now have access to tools that may help them reduce congestion and the pollution it causes.

However, relying on these new data sets may result in unforeseen costs. This issue brief first considers congestion's effects on the environment and the potential role of new data sources in solving congestion issues in the United States. It then considers the limitations of these data: how their vehicle-based focus could lead to a reliance on driving as the primary mode of transportation and how data sets are very likely imbued with demographic biases that may exacerbate challenges facing vulnerable populations.

---

## Congestion has significant effects on the environment

The Texas A&M Transportation Institute estimates that the average commuter spends 42 hours per year delayed, costing them \$1000 in wasted time and fuel.<sup>2</sup> In some large

cities, delays can cost drivers nearly \$3000.<sup>3</sup> And as if stop-and-go traffic is not frustrating enough, it is also a major contributor to local air pollution and climate change.

Fuel combustion in vehicle engines releases unburned hydrocarbons, nitrogen oxides (NO<sub>x</sub>), carbon monoxide (CO), sulfur oxides (SO<sub>x</sub>), volatile organic compounds (VOCs), particulate matter (PM), and carbon dioxide (CO<sub>2</sub>). These air pollutants have a slew of negative health impacts, from increased frequency of asthma attacks, to infants born pre-term and with low birthweight, to premature death.<sup>4</sup> Indeed, living, working, or attending school near major roads is often associated with health problems such as these.<sup>5</sup> Like many other pollutants, nonwhite communities are often at greater risk of exposure. Transportation contributes 55 percent of NO<sub>x</sub> nationwide<sup>6</sup> and nonwhite communities are exposed to this pollution at significantly higher concentrations.<sup>7</sup>

Furthermore, pollution impacts from cars extend beyond the human environment. For example, hydrocarbons, NO<sub>x</sub>, and sunlight react to form smog, leading to acid rain and harming sensitive plants. PM also contributes to acid rain and can upend delicate nutrient balances in coastal and terrestrial ecosystems, damage forests and crops, and reduce biodiversity.<sup>8</sup> The transportation sector is now the largest contributor to U.S. greenhouse gas emissions that cause climate change,<sup>9</sup> the impacts of which are well-documented.<sup>10</sup>

Nearly all of the millions of vehicles on the road burn fossil fuels, and they consume more when stuck in traffic. Large traffic volumes and high vehicle miles traveled in a region typically contribute to congestion emissions, as does poor traffic flow.<sup>11</sup> More vehicle travel intuitively leads to more emissions, while traffic flow is more complicated. Below about 20 miles per hour, emissions rates of CO<sub>2</sub> and other air pollutants—including PM, CO, NO<sub>x</sub>—are significantly higher.<sup>12</sup> Getting vehicles out of stop-and-go traffic can reduce emissions, provided that the reduced congestion does not encourage more driving by people who currently forgo inconvenient vehicle trips.

Yet comprehensive solutions to congestion must focus on all modes of transportation, not just vehicles. Only then can emissions across a city or transportation system be reduced. Bicycle and pedestrian travel, public transit, parking management, and land-use planning are all proven strategies for managing congestion and its related emissions.<sup>13</sup>

---

## The potential uses of new data in managing congestion

To plan for and improve the systems they manage, transportation planners build models that forecast travel patterns, then craft detailed 15- to 30-year plans.<sup>14</sup> Modeling involves collecting data describing where people go, when they go, and which modes of transportation they use in order to project how travel patterns and volumes may change over time.

To make these assumptions, planners often rely on a variety of sources: traffic data, transit ridership statistics, census information, and household surveys of travel patterns. Household surveys are particularly important, but they are expensive to administer and thus only carried out about once every five to 10 years.<sup>15</sup> And while these public sector data sets are improving, they are often limited and lead to poorly-informed decision-making.<sup>16</sup> Low cost, immediately-accessible, and widespread data from private companies such as Uber and Waze could revolutionize transportation planning and management, particularly for congestion relief. Planners could use new data sets to identify traffic bottlenecks, for example, then test different operational or policy changes to fix issues.

Transportation planners will only be successful in reducing congestion-based pollution if they fully understand the severity and causes of congestion in their area. Private sector data offer benefits over traditional sources—such as surveys—because models and traffic management systems can be updated more quickly and at lower cost. Instead of waiting five to 10 years for new survey results, planners can capitalize on free data from a source such as Waze, which provides information in real time. According to researchers at the University of California, Berkeley, as of September 2015, no cities routinely measure traffic volume, occupancy, and speed on main streets.<sup>17</sup> Other research finds a severe lack of evaluation of congestion-reduction strategies.<sup>18</sup> Planning in a big data environment will be far more informed, responsive to exogenous factors, and better at predicting and measuring the impacts of policy or infrastructure changes.

In the future, departments of transportation might use new data to address poor management of incidents, poorly-designed infrastructure, and large volumes of vehicles—all contributors to traffic congestion.<sup>19</sup> For example, Boston used Waze data to identify problematic intersections. The city reduced congestion 18 percent by experimenting with signal timing adjustments.<sup>20</sup> And researchers at Stanford University used Uber Movement's data to identify mobility patterns and pinpoint traffic bottlenecks during rush hours in several cities.<sup>21</sup>

There are myriad possible applications for this type of data collection. Data could be input into traditional transportation planning models, helping determine the size and location of future streets and bridges; the infrastructure needs around ports; or new locations for bus stops. On the traffic management and policy sides, new data could help explain where ride-share demand occurs and inform curbspace policies to reduce backups or vehicle idling. Data could help increase the efficiency of emergency response to traffic crashes and expedite traffic flow after accidents. It could also help determine when, where, and at what level to charge tolls for highway use to reduce vehicle travel during peak hours. These new private sector data sets could be key to fighting traffic pollution in the 21st century.

Yet while these data are exciting, planners need to consider potential pitfalls.

---

## Why planners should be wary of new data

New data poses serious risks if not managed properly. First, the vehicle-based nature of these data sets could encourage driving as the primary mode of transportation in the United States. Second, demographic biases could skew models in a way that disadvantages vulnerable U.S. communities.

### A focus on vehicle-based travel

Currently, the majority of these new data sources, including all of the freely-available ones, are vehicle-based. If transportation models become extremely dependable for vehicle-based travel but advances are not made in collecting, procuring, and integrating data about bicycles, pedestrians, and public transit, then planning solutions will likely skew toward vehicles. Companies such as Strava are starting to collect and market data for biking and walking,<sup>22</sup> but few other companies are making such efforts, and there are currently significant paywalls. Congestion is a systemwide issue, so planners must use new data sources to enhance planning for all modes, or risk furthering reliance on vehicles and their associated emissions.

Uber and other ride-hailing companies provide interesting and important data but could contribute more to congestion than to its solutions. According to a University of California, Davis study, if the service was not available, about half of ride-hailing trips would not have been made or would have been made by walking, biking, or public transit.<sup>23</sup> Planners and policymakers need to be aware of the double-edged aspect of ride-hailing in transportation emissions and find ways to encourage lower-emission transportation options.

### A potentially biased look at transportation demand

Perhaps most importantly, new data may reflect biased accounts of where transportation demand and congestion occur. First, there is a danger of explicit and intentional bias when ride-hailing drivers choose not to drive in certain areas. One study of three ride-hailing apps in Seattle and Boston found that “the cancellation rate for African American sounding names was more than twice as frequent compared to white sounding names.”<sup>24</sup> And anecdotal evidence has found that drivers strategically redline communities perceived as dangerous.<sup>25</sup> Not all studies have found better service in predominantly white and upper-income areas,<sup>26</sup> but planners must be mindful of potential biases when incorporating new data into their models.

More commonly, there is a danger of unintentional bias. For example, Uber demand is inherently lower in areas that cannot afford Uber services or that do not have as many smartphone users, so data availability and reliability are limited in these areas. Several

years ago, the city of Boston planned to partner with a mobile application company called Street Bump that collected data about road condition by measuring “bumps” passively sensed through the accelerometer in users’ phones. The data could then automatically inform the Boston Public Works Department maintenance team. Before deploying the app, however, Street Bump recognized that some populations—such as low-income people and the elderly—may be less likely to carry smartphones or to use the app, so the company was at risk for directing city services to certain neighborhoods over others.

To account for this, Boston first gave the app to its road inspectors, who service all parts of the city equally. Then, the public was able to provide additional, supporting data.<sup>27</sup> The success of Street Bump—which helped Boston identify utility covers, not potholes, as the biggest road condition issue—was predicated on accounting for inequity before incorporating the app into the city’s maintenance program. The same critical thinking must be applied to use of new transportation data for congestion mitigation. To reap the public health benefits of reduced emissions, all communities need transportation data and solutions that reflect their needs.

---

## Conclusion

The transportation community needs all the information it can get to help drive climate- and health-smart transportation policies. New data sources such as those provided by Uber and Waze could help reduce congestion-related emissions. But transportation planners must critically review these data sets and their applications; address issues of demographic representation; and look to nonvehicle modes for congestion solutions in order to ensure that the transportation system of the future is environmentally friendly, equitable, and multimodal.

*Lia Cattaneo is a research associate for Energy and Environment Policy at the Center for American Progress.*

## Endnotes

- 1 Uber, "Uber Movement," available at <https://movement.uber.com/cities?lang=en-US> (last accessed March 2018); Waze, "Connected Citizens Program," available at <https://www.waze.com/ccp> (last accessed March 2018).
- 2 Texas A&M Transportation Institute and INRIX, "2015 Urban Mobility Scorecard" (2015), available at <https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015.pdf> (last accessed March 2018).
- 3 INRIX, "Los Angeles Tops INRIX Global Congestion Ranking," Press release, February 5, 2018, available at <http://inrix.com/press-releases/scorecard-2017/>.
- 4 Environmental Protection Agency, "How Mobile Source Pollution Affects Your Health," available at <https://www.epa.gov/mobile-source-pollution/how-mobile-source-pollution-affects-your-health> (last accessed March 2018).
- 5 Ibid.
- 6 Environmental Protection Agency, "Smog, Soot, and Other Air Pollution from Transportation," available at <https://www.epa.gov/air-pollution-transportation/smog-soot-and-local-air-pollution> (last accessed March 2018).
- 7 Lara P. Clark, Dylan B. Millet, and Julian D. Marshall, "National patterns in environmental injustice and inequality: Outdoor NO<sub>2</sub> air pollution in the United States," *Plos one* 9 (4) (2014): e94431; Jasmine Bell, "5 Things to Know About Communities of Color and Environmental Justice," Center for American Progress, April 25, 2016, available at <https://www.americanprogress.org/issues/race/news/2016/04/25/136361/5-things-to-know-about-communities-of-color-and-environmental-justice/>.
- 8 Environmental Protection Agency, "Health and Environmental Effects of Particulate Matter (PM)," available at <https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm> (last accessed March 2018).
- 9 U.S. Energy Information Administration, "Monthly Energy Review," available at <https://www.eia.gov/totalenergy/data/monthly/#environment> (last accessed March 2018).
- 10 U.S. Global Change Research Program, *Climate Science Special Report: Fourth National Climate Assessment, Volume I* (National Science and Technology Council, 2017).
- 11 ICF International, "Multi-Pollutant Emissions Benefits of Transportation Strategies" (2006), available at [https://www.fhwa.dot.gov/environment/air\\_quality/conformity/research/mpbeb.pdf](https://www.fhwa.dot.gov/environment/air_quality/conformity/research/mpbeb.pdf); Alexander Bigazzi and Miguel Figliozzi, "Traffic Congestion Mitigation as an Emissions Reduction Strategy" (Portland, OR: Intelligent Transportation Systems Lab, Portland State University, 2011), available at <https://www.pdx.edu/cecs/sites/www.pdx.edu/cecs/files/Bigazzi%20MS%20Summary.pdf>.
- 12 Matthew Barth and Kanok Boriboonsomsin, "Traffic congestion and greenhouse gases," *Access Magazine*, 2009, p. 2-9; ICF International, "Multi-Pollutant Emissions Benefits of Transportation Strategies."
- 13 Ibid.
- 14 Edward Beiborn, "A Transportation Modeling Primer" (Milwaukee, WI: Center for Urban Transportation Studies, University of Wisconsin-Milwaukee, 2006), available at <https://www4.uwm.edu/cuts/utp/models.pdf> (last accessed March 2018).
- 15 Jameson L. Toole and others, "The path most traveled: Travel demand estimation using big data resources," *Transportation Research Part C: Emerging Technologies* 58 (2015): 162–177.
- 16 Bent Flyvbjerg, "Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp Up and Sampling," *Transportation Research A: Policy and Practice* 39 (6) (2005): 522–530; David T. Hartgen, "Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling," *Transportation* 40 (6) (2013): 1133–1157.
- 17 Alex A. Kurzhanskiy and Pravin Varaiya, "Traffic management: An outlook," *Economics of transportation* 4 (3) (2015): 135–146.
- 18 Alexander York Bigazzi and Mathieu Rouleau, "Can traffic management strategies improve urban air quality? A review of the evidence," *Journal of Transport & Health* (2017).
- 19 Federal Highway Administration, "Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation," available at [https://ops.fhwa.dot.gov/congestion\\_report/chapter2.htm](https://ops.fhwa.dot.gov/congestion_report/chapter2.htm) (last accessed March 2018).
- 20 Waze, "Reducing congestion through data-driven experiments," available at [https://www.waze.com/ccp/casestudies/reducing\\_congestion\\_through\\_data\\_driven\\_experiments](https://www.waze.com/ccp/casestudies/reducing_congestion_through_data_driven_experiments) (last accessed March 2018).
- 21 Mackenzie Pearson, Javier Sagastuy, and Sofia Samaniego, "Traffic Flow Analysis Using Uber Movement Data" (Palo Alto, CA: Stanford University), available at <http://web.stanford.edu/class/cs224w/projects/cs224w-11-final.pdf> (last accessed March 2018).
- 22 Strava, "Strava Metro," available at [https://metro.strava.com/?\\_branch\\_match\\_id=472479228730943795](https://metro.strava.com/?_branch_match_id=472479228730943795) (last accessed March 2018).
- 23 Regina R. Clewlow and Gouri Shankar Mishra, "Disruptive transportation: the adoption, utilization, and impacts of ride-hailing in the United States" (Davis, CA: Institute of Transportation Studies, University of California, Davis, 2017), available at [http://usa.streetsblog.org/wp-content/uploads/sites/5/2017/10/2017\\_UCD-ITS-RR-17-07.pdf](http://usa.streetsblog.org/wp-content/uploads/sites/5/2017/10/2017_UCD-ITS-RR-17-07.pdf).
- 24 Yanbo Ge and others, "Racial and gender discrimination in transportation network companies," Working Paper 22776 (National Bureau of Economic Research, 2016).
- 25 Constantine Singer, "Ridesharing and Redlining: Uber, Lyft, Race and Class," *Daily Kos*, May 27, 2014, available at <https://www.dailykos.com/stories/2014/05/27/1302417/-Ridesharing-and-Redlining-Uber-Lyft-Race-and-Class#>.
- 26 Ryan Hughes and Don MacKenzie, "Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators," *Journal of Transport Geography* 56 (2016): 36–44; Mingshu Wang and Lan Mu, "Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA," *Computers, Environment and Urban Systems* 67 (2018): 169–175.
- 27 Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (2014), p. 51–52.