STREETLIGHT DATA

VMT Monitor – Methodology and Validation

Introduction

This white paper outlines our motivation, methodology, and some validation results for StreetLight Data's daily Vehicle Miles Travelled (VMT) Monitor. This is a living document! We developed the VMT Monitor very quickly in light of the COVID-19 crisis to help the transportation industry track VMT on a daily basis. We may improve our methodology as time moves forward, or as our community suggests improvements.

StreetLight's VMT Monitor provides our estimate of total vehicle miles travelled (VMT) by residents of each county, each day since the COVID-19 crisis began, as well as a change from the "baseline." We define baseline as the average daily VMT of January 2020. It is updated every two days with a ~48hour lag to real time. To access the VMT Monitor – go to this webpage: https://www.streetlightdata.com/VMT-monitor-by-county/

Motivation – Why VMT Monitor and not a "Mobility Index"?

At StreetLight Data, our purpose is to help the transportation community put big data to work. We know that several companies have published different metrics derived from locational data to describe activity during the COVID-19 crisis, many in the form of a "Mobility Index." After reading the methodology documentation and reviewing the results of several such indices, we felt that these did not serve to answer some important questions for the transportation community. **This is not a criticism of Mobility Indices!** They are a tool developed for the epidemiological community. We are not epidemiologists and cannot comment on their usefulness. In fact, as this methodology paper describes, we built our VMT Monitor in part on a derivative of the Mobility Index published by our partner Cuebiq. We are transportation professionals – so we derived a different metric for our community of practice.

Why focus on VMT?

We decided that a daily Vehicle Miles Travelled (VMT) update would be most useful for the transportation community for several reasons:

1. It is a well-known and widely used metric in the field in general. Therefore, it will be more intuitive for many practitioners.
2. VMT is directly correlated with things that matter to transportation practitioners including: gas tax revenues, road wear-and-tear, greenhouse gas emissions, and other smog/particulate creating emissions.
3. VMT captures and measures all individuals in a region, not just one representative individual (see description of Mobility Indices below and why this matters).
4. VMT takes into account the frequency of trips, not just the length of the median trip (see description of Mobility Indices below and why this matters).

What is a Mobility Index and why does it not represent "how much transportation is happening"?

We reviewed the methodologies shared by those publishing Mobility Indices[1] (thank you to the companies who were transparent on methodology!). Most of them use a common framework that is based in the epidemiological community. Again – we emphasize that we're not criticizing any Mobility Index. We're simply pointing out their shortcomings for transportation purposes.

The main gap – **Mobility Indices are not correlated with total travel**. This is for two main reasons:

1. Mobility Indices use median distance traveled from "home" (as opposed to mean)
2. Mobility Indices focus on the average individual's trip length – they're not scaled by the total number of individuals or the number of times per day the individuals take a trip.

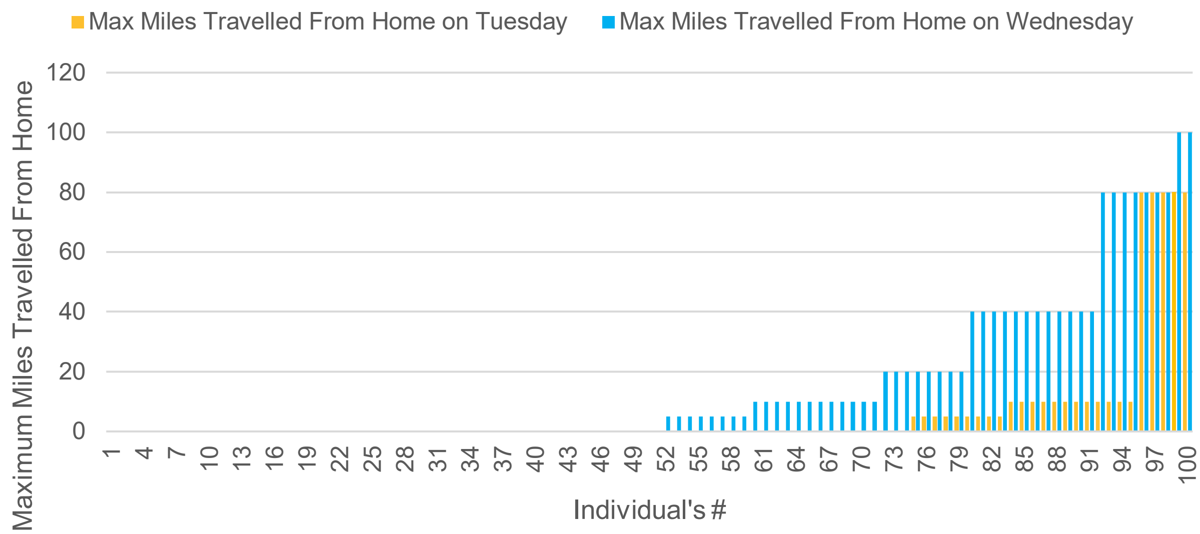
To illustrate the importance of the median vs. mean question, let's imagine a county that has 100 individuals. Figure 1 shows the maximum distance travelled from home on Tuesday and Wednesday. On both days, 50 of the individuals (50% of the population) fully stay at home. But the other 50 individuals behave quite differently on Tuesday (orange) and Wednesday (blue). On Tuesday, 25 more individuals also fully stay at home and the remaining individuals travel 5-20 miles (though a few travel 80 miles) from home. But on Wednesday, all of the remaining 50 individuals travel, with several going 120 miles away from home. The fundamental problem with using "median" is that both Tuesday and Wednesday get the exact same score. From a transportation perspective, these days are in fact quite different. Using the "mean" distance from home, as opposed to the median, captures a lot of this difference, as shown in Table 1.

Table 1 – Comparison of Mean and Median miles for Tuesday and Wednesday

	Tuesday	**Wednesday**
Median miles travelled from home	0	0
Mean miles travelled from home	5.65	15.6

[1] https://help.cuebiq.com/hc/en-us/articles/360041285051-Reading-Cuebiq-s-COVID-19-Mobility-Insights, https://github.com/descarteslabs/DL-COVID-19, https://www.unacast.com/post/the-unacast-social-distancing-scoreboard

Figure 1: Distribution of Maximum Miles Travelled from Home by 100 Individuals- Tuesday and Wednesday



Next, to understand the full scope of the travel for the region we need to understand how many trips the population took. Two trips of 10 miles is the same as four trips of five miles in terms of total travel. But if we just looked at the mean they are different.

Table 2 – Comparison of two individuals' trip patterns, mean trip length, and total VMT.

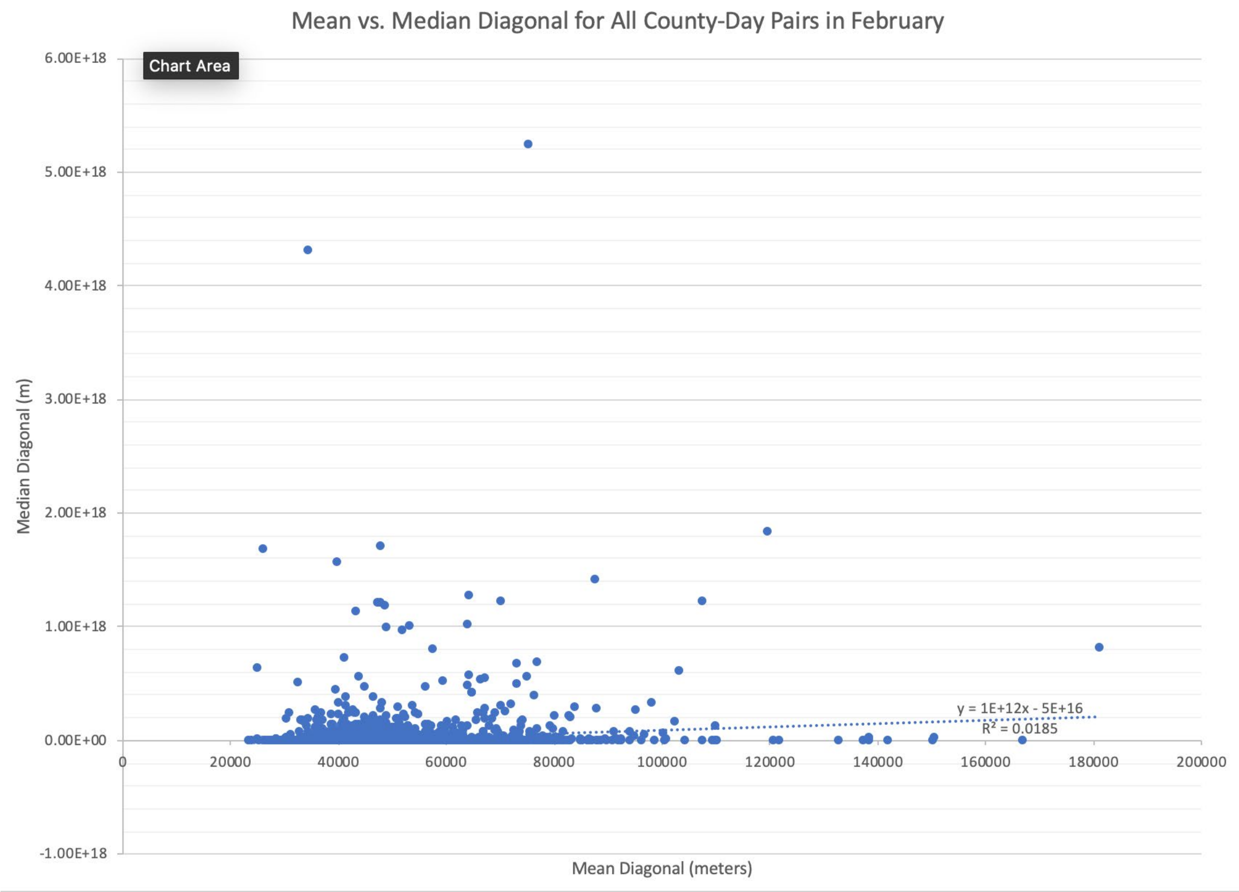
	Tuesday Trip Pattern	**Mean Trip Length**	**Total VMT**
Individual 1	2 trips of 10 miles	10	20
Individual 2	4 trips of 5 miles	5	20

Thus – to capture total travel we need a metric that measures both mean trip length and the total number of trips taken by the full population as shown in this framework equation:

$$VMT = \overline{Trip\ Length} * Total\ \#\ of\ Trips\ by\ All\ Individuals$$

(NB – one cannot multiply the median trip lengh or distance from home by the population to get total travel, even if one knows the frequency of trips).

It is even possible that the total travel in a county could go up (if the 51^{st} through 100^{th} person started travelling much more) but the median/Mobility Index could go down (if the 50^{th} person started travelling less), which would lead to very erroneous overall conclusions. The following figure shows the scatter plot of the mean diagonal of all the places travelled from home vs. the median, and shows the real lack of relation. Note – we dropped the 1000 most extreme medians to make this chart, as they were so far off the axes it made the data hart to read.



Methodology

We had to find a way to follow the framework equation above (mean trip length * total number of trips by all individuals) but also was able

1. To be scaled and normalized to represent all the VMT in any county, even though we only capture a subset of the population
2. Be computable nationwide, quickly
3. Adjust for the fact that many individuals are not sheltering in their normal “residence” county – college students have gone home, city-dwellers have often gone to stay with relatives or vacation homes, etc.

We developed the following methodology:

Step 1: Calculate Baseline VMT per County

StreetLight has procession billions of locational data points each month (see this link for our standard methodology document). Therefore, we used data already residing in our *StreetLight InSight* app to capture the average daily VMT for every county in the continental US for January 2020, which we call the closest baseline travel.

Step 2: Calculate Mean Trip Length and Total Trips Per County

To do this with daily updates, we partnered with Cuebiq, one of our data suppliers. We proxied the mean trip length as “mean diagonal of the bounding box capturing all trips for all devices whose most frequent location is in this county.” We proxied total trips in the county as “total number of trip starts and ends for all devices whose most frequent location is in this county.” Thus our VMT-proxy is expressed as:

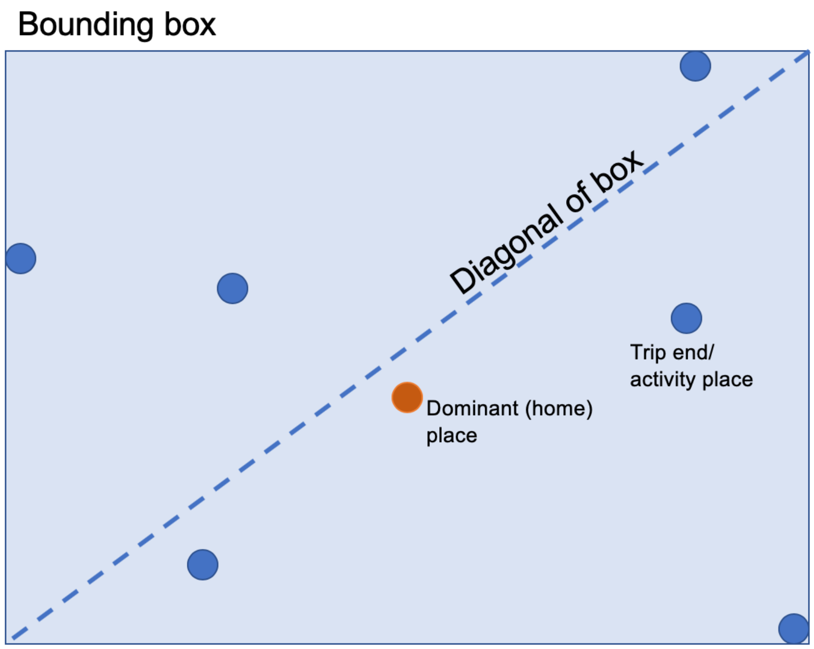
$$VMTPROXY = \overline{diagonal\ of\ trip\ bounding\ box} * Total\ \#\ of\ Trips$$

For all devices in the sample whose most frequent location is in the county.

After we determined (see validation, below) that the product of these two metrics yielded the best results, Cuebiq began calculating them on a daily basis in conjunction with their Mobility Index**. It is important to note – these methodology choices mean our VMT represents VMT of people residing in the county on a particular day. It will not county the VMT of, say, trucks cutting through the county.**

We eliminated counties on days where they had fewer than 500 trips in our sample for the given day for lack of statistical validity due to low sample.

Figure 2: Illustration of Bounding Box diagonal concept.



Step 3: Scale and Normalize VMT-proxy to total VMT

We calculated the VMT-proxy for every day in January 2020 (our baseline month) for all counties. Then we excluded counties with fewer than 5,000 VMT on a baseline day. Then, we expressed the VMT-proxy for each day in March, and now April as a percent of the January VMT-proxy. We used this factor to adjust the January VMT down and create a normalized VMT estimate for each of the more recent days. A few examples are shown in Table 3 for three imaginary counties that all had the same January Average Daily VMT of 100,000 miles.

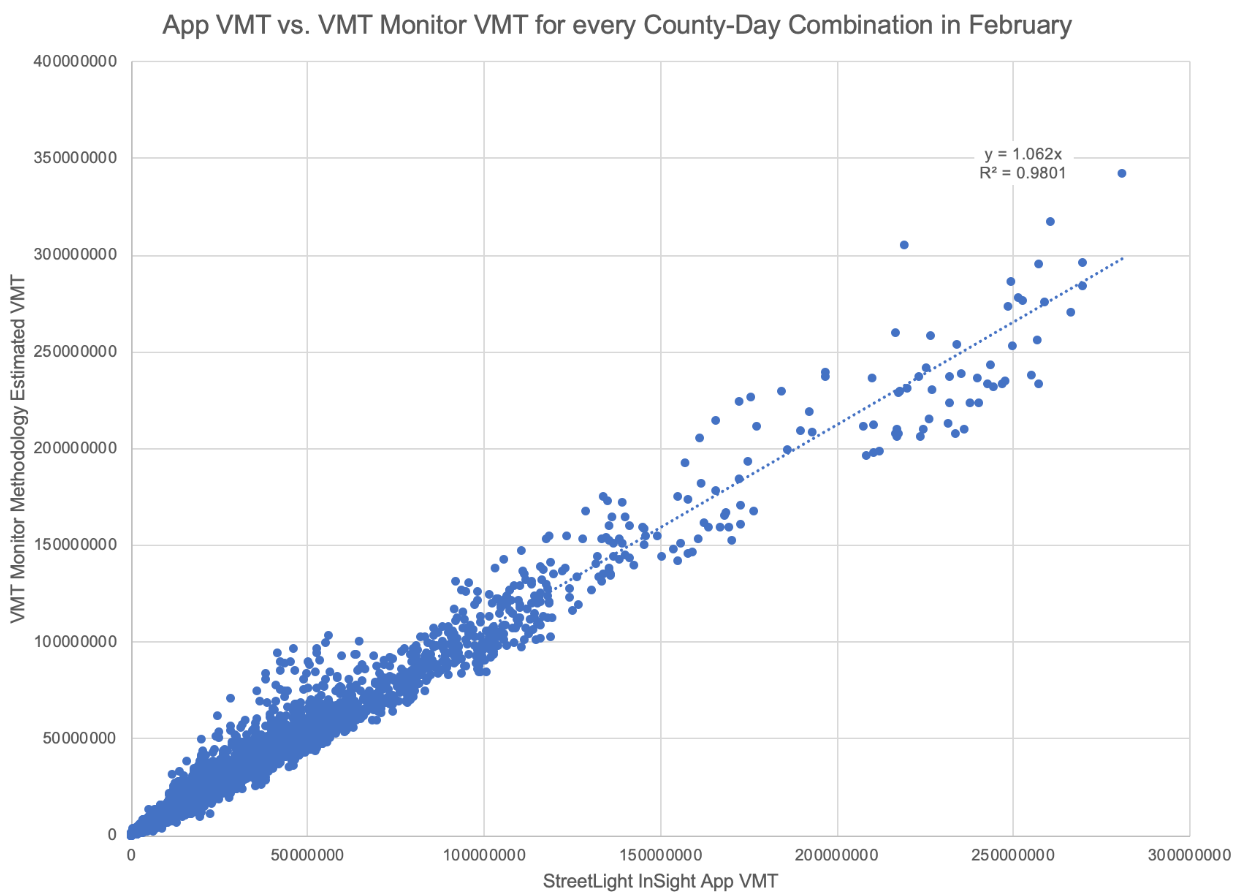
Table 3: Example Normalization for Three Imaginary Counties

County	January Avg Day VMT	Jan – Avg Day VMT-proxy	March 23rd – VMT-Proxy	March 23rd proxy as % Jan proxy	Estimated March 23rd VMT
A	100,000	49.6	23.1	47%	47,000
B	100,000	12.5	2.1	17%	17,000
C	100,000	56.7	16.4	29%	29,000

Validation

We also wanted to build a metric that our community could trust, so we designed a validation study. As mentioned above, our baseline month in January 2020. StreetLight also has already processed all data for February 2020. We have scaled, QA-ed, normalized, and validated all our February 2020 data, as we do every month, several ways including detailed comparison to an array of 10,000+ permanent loop counters. Thus, we if our VMT Monitor method could estimate each day's VMT in February for each county, we could feel confident in its on-going accuracy.

As shown in the figure below, the results correlated extremely well, with an R^2 of 0.98. No outliers were dropped.

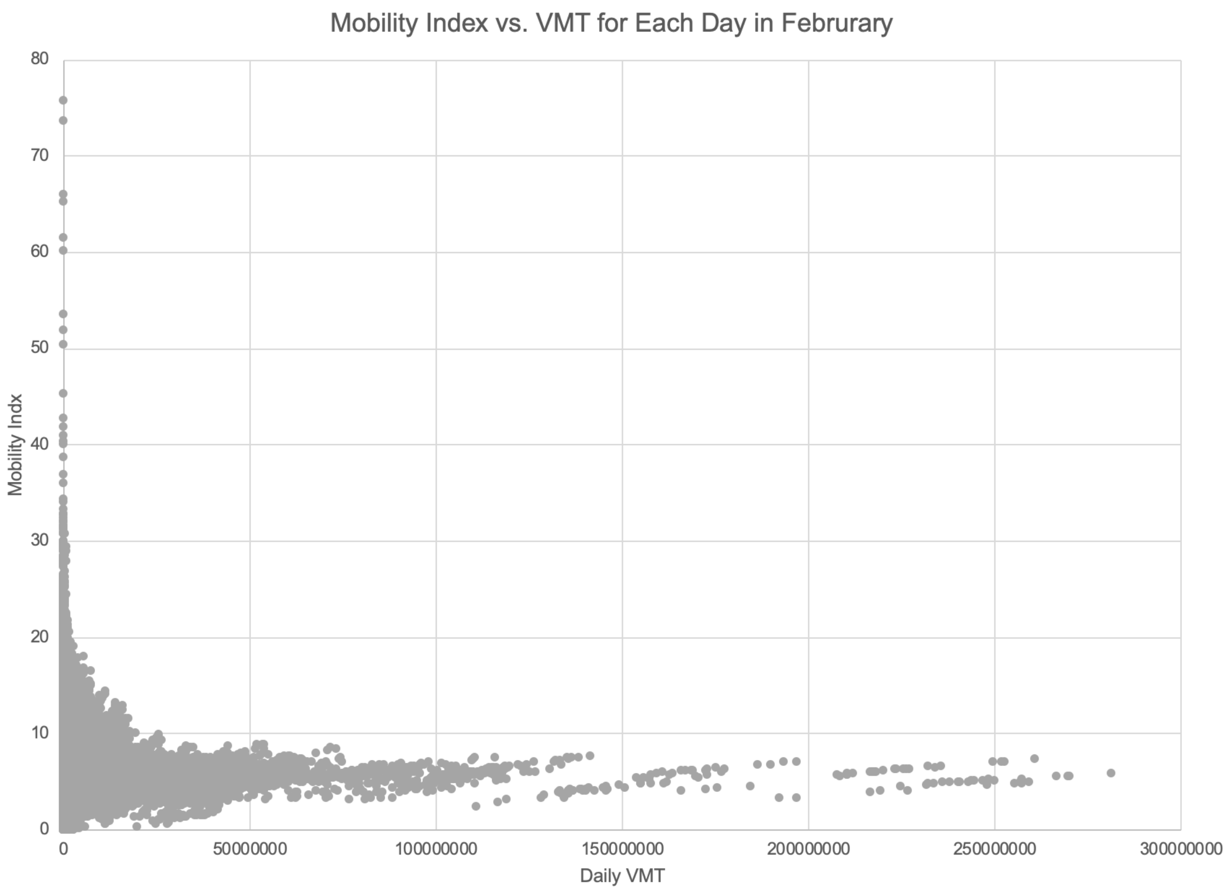


Attempting a similar methodology using the median diagonal (instead of mean) in conjunction with the number of trips yielded far worse results, as expected, shown below.

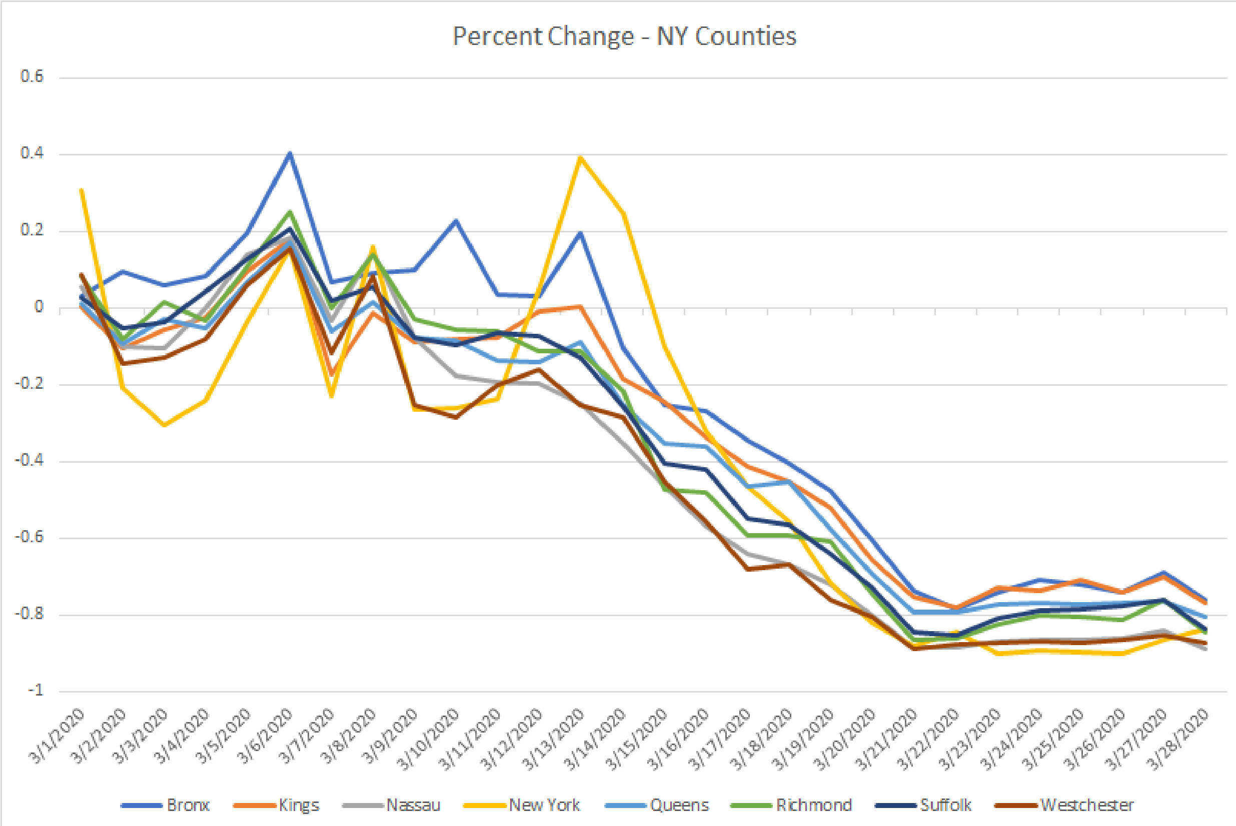
Table 4: R2 comparison for various methods of estimating total travel

Approach	**R2 to Normalized, Calibrated, Validated VMT/day for each County-Day combination in February**
VMT Monitor Scaling (mean length * # trips)	0.98
Median-based (median length * # trips)	.00001

Mobility Index	.01



The March VMT Monitor estimated VMTs show clear declines in counties in New York that have advisories or requirements to stay at home (as well as a spike in weekend travel just before the order took place), as shown in the figure below.



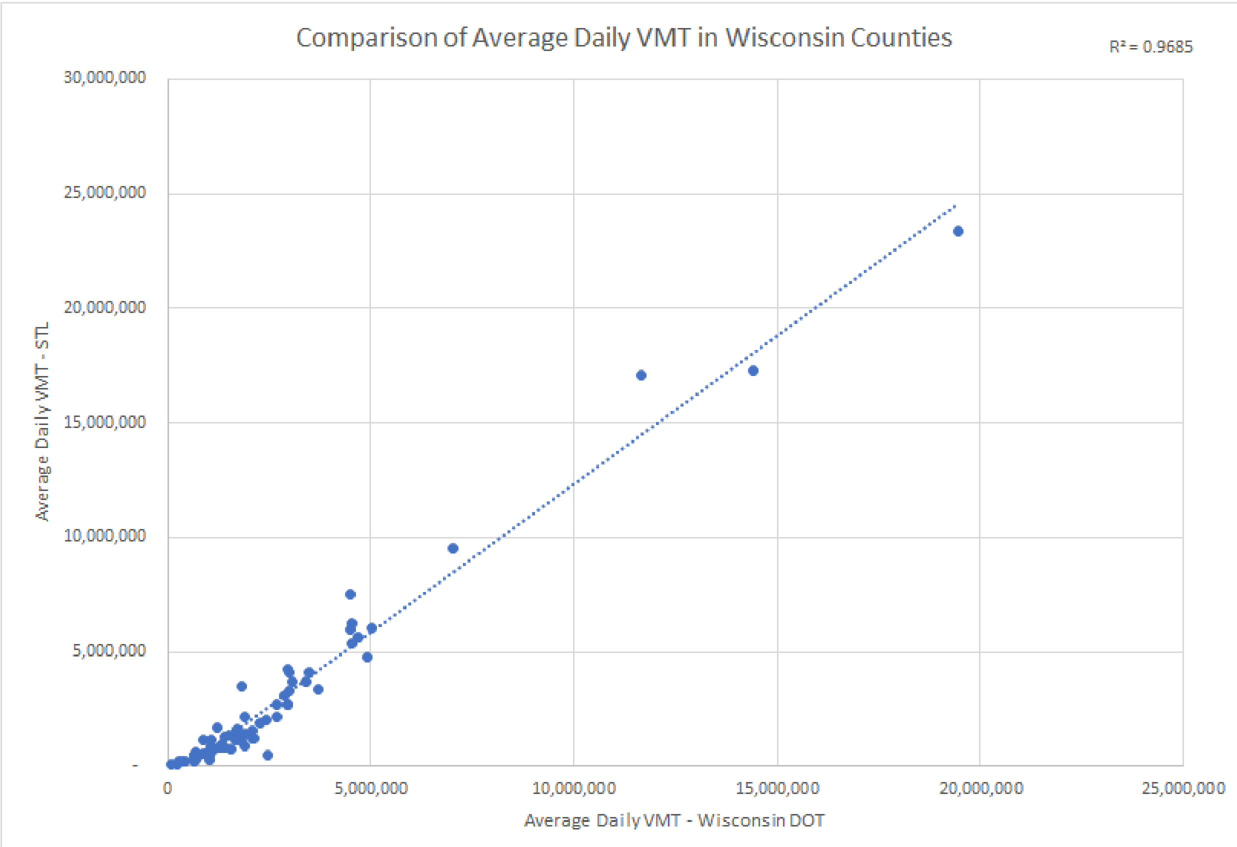
Validation Data for Baseline VMT

We validated our January Baseline VMT against several spot sources, as shown in the Figure below. The hyperlinks

Comparison to a few single counties/regions:

Source	Year	Value	StL Estimate
LA County	2017	222 million/day	237.6 million/day (Jan Avg)
Bay Area	2015	171.9 million/day	180.0 million/day (Jan Avg) for 9 counties
Houston	2015	133.5 million/day	208.2 million/day (Jan Avg) for Harris County
Chicago	2015	180.8 million/day	135.4 million/day (Jan Avg) for Cook County
Wisconsin	2018	180.5 million/day	191.7 million/day for all counties

We also did an correlation between our January 2020 baseline data for all counties in Wisconsin, finding readily available data[2] for the counties in 2018, as shown in the Figure below. R^2 was 0.96.



[2] https://wisconsindot.gov/Documents/projects/data-plan/veh-miles/vmt2018-c.pdf