Validation Report for Measure Up! LOCUS 2019/2020

Cambridge Systematics, Inc. (CS) developed the Measure Up! LOCUS 2017 Flows Dashboard from trip tables prepared by employing thoughtful algorithms to transform raw location-based services (LBS) data and transit ridership data (smart farecard and APC) into carefully validated travel patterns for Los Angeles County. This user guide focuses on the description of the dashboard and provides instructions on how to utilize the dashboard.

Learn more about LOCUS on www.camsys.com/locus

CAMBRIDGE SYSTEMATICS

Contents

About the Data

1

The Measure Up! LOCUS 2019-2020 Flow Dashboard leverages three datasets that reflect granular travel and activity patterns observed in Los Angeles County (1) LOCUS, which captures overall travel as well as estimates for non-motorized travel, (2) regional transit ridership data from a combination of TAP transactions and automated passenger counts (APC) datasets, and (3) truck movement data from SCAG's regional model.

Location-based services (LBS) data collected passively from mobile devices are becoming an increasingly valuable source of information about travel patterns. LBS data are collected by GPS applications running either in the background or foreground on cellular devices, where the device user has opted to allow access to the app to import the device's geographic location. LBS data are spatially more accurate than other forms of cellular data because they collect geographic locations on a GPS platform.

These data can provide detailed information about how people are moving, where they are going, and when their travel is occurring. Compared to household surveys, LBS data can be collected for longer periods of time, at more regular intervals, and from a larger sample size. As a result, these datasets are massive in size, often containing millions of records collected over a period of months, rather than the typical 1 to 2-day travel diary often collected by travel surveys. Not only does this generate a larger overall sample, but travel patterns of individual devices can be measured over a period, while maintaining sufficient degrees of privacy since the device ID cannot be tied to any demographic or personally identifiable information.

LBS data serve as the primary data source for Measure Up! LOCUS. The development of LOCUS is grounded in a deep understanding of travel behavior and econometrics coupled with strong big data analytics and validation expertise. In the development of Measure Up! LOCUS, the following key steps have been adhered to:

- Passenger travel across a variety of market segments including travel purposes, time-of-day, and day of the week are captured. Flow patterns are developed at the census block group level, which supports granular evaluation of travel.
- Using the inferred home location of the devices, LOCUS helps identify those devices that live in equity-focused communities (e.g., low-income and/or minority population communities).
- Data from Q3 and Q4 of 2019 and 2020 are included in the development of the dashboard. LOCUS only retains the most reliable data using rigorous heuristics and data cleaning techniques.
- Data in the final product have been expanded to match Census population data and validated against national sources of travel behavior.
- LOCUS data (and Transit Ridership data) are displayed using a Tableau interface that allows for quick and easy summaries.

This validation document focuses on the checks conducted as part of the Measure Up! LOCUS data development process.

** Vinayak P., Wafa Z., Cheung C., Tu S., Komanduri A., Overman J., and Goodwin D., 2019, Using Smart Farecard Data to Support Transit Network Restructuring: Findings from Los Angeles, Transportation Research Record (TRR), Volume: 2673. 6, 202-213.

Transit Trips

2

The transit trip estimates for 2019 and 2020 are derived using a combination of the TAP (smart farecard transactions) and automated passenger counts (APC) datasets. Collectively, the cover a vast majority of transit travel in the region on both Metro bus and rail routes as well as bus routes operated by regional transit agencies. Due to the unavailability of a complete set of TAP data for 2020, the transit trip tables derived using 2017 data (base year) are used as a seed matrix and adjusted using ridership estimates derived using APC (for Metro Bus) and published ridership numbers (for Metro Rail and regional transit agencies).

2.1 Base Year (2017) Transit Trip Table

The TAP data available collected over 4 months in 2017, as a part of LA Metro's NextGen Bus Study, is used as the seed matrix to generate the transit trip tables. Key attributes in the TAP data include boarding location (spatial attributes), transaction time, route, and a unique ID for each card user. The unique ID is pivotal to the translation of unlinked trips (transactions) to linked trips – each record representing a complete transit itinerary, with the transfers accounted for.

A multi-step expansion procedure is implemented to account for cash transactions, non-mappable transactions, and other missing records. The fully expanded transit trip table represents the overall transit travel in the region. For additional information on processing and validation of the transit ridership dataset, please refer to the article referenced in the footnote.

2.2 Future Year (2019 and 2020) Transit Trip Tables

Since a complete set of TAP data was unavailable for future years, the following procedure is deployed to generate transit trip estimates using the base year transit data.

Step 1: Calculation of average daily boardings in future years

- For Metro Bus, stop level APC data for comparable months in 2019 and 2020 are processed to generate average daily boardings by stop ID, day type (Weekday, Saturday, and Sunday), time of day periods, and year. The stop ID is tagged to neighborhoods in LA county, and the data is aggregated to neighborhood spatial resolution.
- For Metro Rail, average daily boardings are available by day type and year.
- For Regional Transit Agencies, ridership data are available at different levels of aggregations. For agencies with data available by day type and year, average daily boardings are calculated using the total ridership numbers and number of days of each day type. For agencies with no day type splits available, the totals are distributed between the day types using the Metro bus splits and then converted to average daily boardings.
- See Table 2.1 for agencies included in the dataset.

Vinayak P., Wafa Z., Cheung C., Tu S., Komanduri A., Overman J., and Goodwin D., 2019, Using Smart Farecard Data to Support Transit Network Restructuring: Findings from Los Angeles, Transportation Research Record (TRR), Volume: 2673. 6, 202-213.

Table 2.1 Agencies with ridership data available in 2019/2020

Agency Name
Metro (Bus & Rail)
ACT
AVTA
BCT
Big Blue Bus
Commuter Express
DASH
El Monte Transit
Foothill Transit
Gtrans
Long Beach
LA County Shuttles
Montebello
Pasadena Transit
Santa Clarita

Step 2: Calculation of Transfer Ratios

- Base year TAP data includes both transactions and linked trips, extracted from those transactions by collapsing transfers. The transfer ratio is calculated as the ratio of transactions (unlinked trips) to linked trips. This ratio captures, on average, how many of the boardings are transfers vs origins of trips.

- This ratio is generated at the aggregation levels at which the daily boardings are available for Metro Bus, Metro Rail, and Regional Transit Agencies.

Step 3: Conversion of boardings to linked trips and adjustments of base year transit table

- Average daily boardings are converted to estimates of linked trips using the transfer ratios. This gives us estimates of the total number of linked trips in the region for 2019 and 2020.

- Base year tables are scaled to match the new control totals along the adjustment dimensions. Table 2.2 shows the average daily boardings in the future years and average daily trips in the future years (after translation using transfer ratios) for all agencies combined.

Table 2.2 2019/2020 Average Daily Boardings and Average Daily Trips

Day of Week	2019 Average Daily Boardings	2019 Average Daily Trips	2020 Average Daily Boardings	2020 Average Daily Trips
Weekday	1,482,334	1,001,806	684,238	460,718
Saturday	609,088	609,091	499,168	345,125
Sunday	488,743	502,785	422,059	299,559

Table 2.3 shows the relative drops in average daily boardings (from Metro's ridership dashboard) and estimated average daily linked trips between 2019 and 2020. Note that the Metro ridership numbers reflect only Metro operated routes, while the linked trips cover regional agencies as well - but we can compare the two metrics since Metro routes carry over 80% of the total ridership in the region. As seen below, the drops are consistent with Metro's reporting on each day type.

Table 2.3 Changes in Metro Avg Daily Boardings and Avg Daily Linked Trips (2020 – 2019)

Day of Week	Avg Daily Boardings (Metro Ridership Dashboard)		Estimated Avg Daily Linked Trips		% Change (2020 - 2019)	
	2019	2020	2019	2020	Avg Daily Boardings	Avg Daily Linked Trips
Weekday	1,178,825	563,788	1,001,808	460,719	-52%	-54%
Saturday	745,130	443,932	609,091	345,202	-40%	-43%
Sunday	593,605	380,698	502,793	299,663	-36%	-40%

Bike & Pedestrian Trips

3

3.1 Active Transport Approach

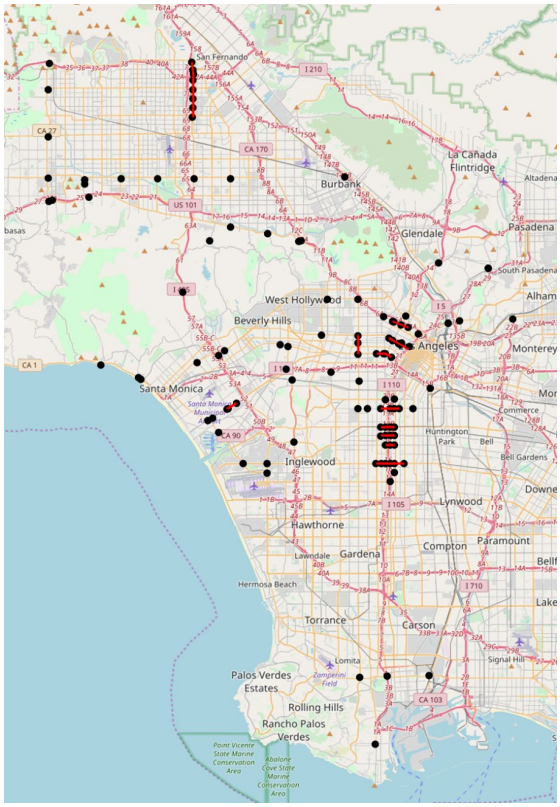
The active transportation methodology framework is designed to identify travel made by these modes in the LOCUS data. The approach was developed using bike and pedestrian trip records from surveys collected around the country. By using survey data with information about the location and path of active transport trips, key factors that determine modal choice including travel distance, network characteristics, elevation change, urban form, weather, and time of day were identified.

This approach was built using a nationwide panel and served as the starting point for the Measure Up! Dashboard. During the validation step, it was observed that the nationwide approach underestimated walk trips and overestimated bike trips when compared to local count data. We believe that this is consistent with the fact that bicycling in LA is localized and limited to areas with bike-lanes and traffic calming measures. So, the nationwide model was calibrated for LA County to better reflect ground conditions.

3.2 Bike & Pedestrian Trip Validation

For the Measure Up! LOCUS dashboard development, bike and pedestrian counts collected by local jurisdictions were made available to the project team. These counts were taken from 116 locations across the county. Figure 3.1 shows the count locations (shown as black dots on the map).

Figure 3.1 Bike and Pedestrian Count Locations and Screenlines



A total of 10 screenlines were developed (as shown as red lines on the map in Figure 3.1), each of which connects a series of count locations with one another. By comparing the aggregate counts for a given screenline against LOCUS estimates, we can decipher how well LOCUS estimates match counts.

Originally 11 screenlines were developed for this effort. However, it was determined that one of the screenlines did not have count information for the movements that cross the screenline. Instead, counts were available only for volumes running parallel along the screenline itself. Therefore, this screenline was ultimately dropped from the analysis. The screenline that was dropped the screenline just to the north and west of LAX, shown in Figure 3.1. Table 3.1 provides additional detail about the 10 screenlines used in the analysis.

Table 3.1 Screenline Descriptions

Screenline	Description	CrossStreet1	CrossStreet2
1	Hwy 42, S of Downtown (E-W)	Vermont	Avalon
2	Pico Blvd, W of Downtown (E-W)	Vermont	Union
3	Florence Ave, S of Downtown (E-W)	Figeroa	Main
4	Gage Ave, S of Downtown (E-W)	Hoover	Main
5	Slauson Ave, S of Downtown (E-W)	Hoover	Main
6	Vernon Ave, S of Downtown (E-W)	Hoover	San Pedro
7	Wilshire Blvd, W of Downtown (E-W)	Alvarado	110 Frwy
8	Temple St, NW of Downtown (E-W)	Benton	Glendale
9	Western Ave, Koreatown (N-S)	Wilshire	Pico
11	Sepulveda Blvd, SW of San Fernando (N-S)	Rinaldi	Nordhoff

Each of the count stations collected counts during primarily the peak periods. Therefore, the peak periods were the emphasis of the validation process. Two different approaches were conceived for walk and bike validation. Those two methods are documented below.

3.2.1 Pedestrian Validation Approach

Pedestrian trips are typically very short-distance trips, with trips greater than 1 mile being quite rare. As a result, the approach devised to develop estimates of LOCUS screenline volumes was based on creating two bounding boxes: one on either side of each screenline. Then, LOCUS trips with one trip end each in the two bounding boxes were identified and the weighted pedestrian flow of those trips was taken as the LOCUS screenline pedestrian volume. Screenline pedestrian counts were taken as the sum of counts for each count location associated with the given screenline.

Bounding boxes were set to capture the area starting at one extreme of the screenline and extending to the other extreme of the screenline (that is, one edge of the bounding box was set to be identical to the screenline). The depth of each bounding box was then set to 1 mile.

3.2.1 Bike Validation Approach

Bike trips tend to be longer than pedestrian trips making the pedestrian approach unsuitable for analyzing bike trips. Instead, bike trips were routed on an OpenStreetMaps network for the county. The routing assignment was based on simply minimizing travel distance. Links with prohibitions against bikes (like on freeways) were excluded in the assignment process.

While this approach offers more opportunity for error because it introduces a secondary assignment model which is a generalization of reality, it offers the benefit of being able to consider all LOCUS bike trips and makes no pre-determined assessment about what path is used for any given trip. It is important in this case to aggregate count stations to minimize the possibility of over-assigning trips to certain paths and under-assigning trips to other parallel paths. By aggregating count stations for each screenline, many of these assignment errors will offset one another for comparing to counts.

We found after assigning bike trips to the network that some links along the screenline were assigned bike volumes that did not have corresponding counts. That is, no count station existed for some links along each screenline. To avoid an apples to oranges comparison of LOCUS to counts, the links along each screenline that did not have counts were filtered from the analysis. Each location considered in the comparisons for each screenline, therefore, had LOCUS volume estimates and also had count information.

3.3 Validation Results

While we do not see a perfect match between bike volumes and counts (Table 3.2), the numbers are similar to each other. Furthermore, we see very high correlations between counts and volume estimates—0.87 in the AM Peak and 0.75 in the PM Peak. This suggests that LOCUS reasonably captures where bike travel is highest. Regionwide, LOCUS predicted bike share is about 0.7 percent, consistent with survey and other data sources for a large region.

Table 3.2 Bike Screenline Validation Results

Screenline	AM Peak			PM Peak		
	Cnt-Bike	LOC-Bike	%Dif	Cnt-Bike	LOC-Bike	%Dif
1	150	137	-9%	302	225	-26%
2	151	223	48%	246	313	27%
3	81	78	-4%	162	122	-25%
4	104	123	18%	179	172	-4%
5	92	120	30%	166	166	0%
6	173	237	37%	229	308	34%
7	157	169	8%	228	248	9%
8	67	95	42%	116	143	23%
9	114	164	44%	215	259	20%
11	57	91	61%	131	133	1%
Total	1,146	1,437	25%	1,974	2,088	6%

Table 3.3 Pedestrian Screenline Validation Results

Screenline	AM Peak			PM Peak		
	Cnt-Wlk	LOC-Wlk	%Dif	Cnt-Wlk	LOC-Wlk	%Dif
1	1,969	369	-81%	2,898	527	-82%
2	1,676	658	-61%	3,899	1,211	-69%
3	1,025	77	-93%	1,502	115	-92%
4	576	141	-76%	948	163	-83%
5	918	102	-89%	1,216	160	-87%
6	1,583	321	-80%	2,288	499	-78%
7	2,999	2,562	-15%	4,705	4,188	-11%
8	873	428	-51%	1,398	488	-65%
9	1,855	1,351	-27%	4,019	2,655	-34%
11	1,700	429	-75%	2,645	630	-76%
Total	15,174	6,438	-58%	25,518	10,634	-58%

Table 3.3 shows the pedestrian validation results for the 10 screenlines. Overall, the LOCUS estimates of pedestrian volumes are low by more than 50 percent in both the AM Peak period and the PM Peak period. While on its face, this seems like a wide margin, there are a couple of reasons for these results.

- First, capturing very short trips is challenging using LBS data. The reason is that a short movement of a device does not always indicate that a trip occurred, it could be signal inconsistencies, or it could be movement within a building where an activity is occurring. So, we err on the side of caution and flag short trips which we consider false positives. Because of this, the LOCUS estimates of trips less than 0.25 miles using network facilities is likely lower than reality.
- Second and perhaps most importantly, loop trips are not captured by LOCUS. Loop trips are ones typically engaged in as recreational trips. For instance, going for a walk or a jog and ending at the same place as the start would typically be considered a loop trip.

We believe the first issue tends to be isolated in highly dense urban areas such as DTLA and is not a region-wide challenge. The second issue doesn't fundamentally affect mode share baselining because walk/bike loop trips are unlikely to be captured in any future baselining/trend analysis efforts.

Despite the differences in volumes, the correlation between counts and LOCUS volume estimates is very good. The correlation between counts and volumes in the AM Peak was 0.84 and in the PM Peak was 0.86. These correlations indicate that the LOCUS volume estimates are highest in the same screenlines where actual counts are highest. Overall walk shares countywide are in line with observations from other datasets of about 11 percent on weekdays.

Freight Trips

4

Freight flows are developed based on the SCAG Heavy Duty Truck (HDT) model**, model output year 2019.

4.1 Data Structure in the Model

The SCAG model generates summaries at a TAZ (Traffic Analysis Zone) level for the entire 6-County SCAG region. The model has a total of 2300 TAZs in Los Angeles County. The model outputs flows for five different times of day:

- AM Peak: 6:00 AM – 9:00 AM
- Midday: 9:00 AM - 3:00 PM
- PM Peak: 3:00 PM - 7:00 PM
- Evening: 7:00 PM – 9:00 PM
- Night: 9:00 PM – 6:00 AM

The SCAG HDT model has trips for three different truck types. These truck Types are defined based on the vehicle's gross vehicle weight (GVW). The truck type definitions are shown in the table below.

Table 4.1 Truck Type Definitions

Truck Type	Gross vehicle weight (GVW)
Light Truck	8,500 to 14,000 lbs.
Medium Truck	14,001 to 33,000 lbs.
Heavy Truck	More than 33,000 lbs.

4.2 Data Structure in LOCUS

LOCUS data for the passenger uses the geography at the Census Block Group (BG) and Neighborhood levels. TAZ-level truck trips were redistributed to the same spatial resolution (BG) of passenger data to help better understand truck movement by combining the passenger trips.

The time-of-day periods of the LOCUS passenger data are defined to better meet LA Metro's modeling and planning usages. The LOCUS TOD periods are as follows:

- Early AM: 4:00 AM- 6:00 AM
- AM Peak: 6:00 AM - 9:30 AM
- Midday: 9:30 AM – 2:00 PM
- PM Peak: 2:00 PM – 6:30 PM
- Late Evening: 6:30 PM – Midnight
- Owl Period: Midnight – 4:00 AM

** SCAG 2008 Regional Mode. Chapter 7 – Heavy Duty Truck Model. https://scag.ca.gov/sites/main/files/file-attachments/heavy_duty_truck_model_documentation.pdf?1605572862

4.3 Redistributing Truck Trips

The truck data from the SCAG Model were reallocated to Census BGs and to the time-of-day bins already used in the LOCUS passenger dataset using a series of industry-approved reallocation frameworks. Further, since SCAG has validated their model thoroughly, only data fidelity checks were performed as part of this effort.

4.3.1 Geographic Redistribution

In performing the geographic redistribution, we adopted truck trip rates from establishment surveys conducted by CS across the US. This helps create a relative share of truck trips generated by different Census BGs and is used to implement the reallocation methodology

Table 4.2 Truck Trip Rates by Industry

NAICS Industry Type	Truck Trips per Employment
Agriculture, Mining, Utility, Construction	0.38
Manufacturing	0.57
Retail	0.49
Wholesale, Transportation & Warehousing	2.01

4.3.1 Temporal Redistribution

For the reassignment of the TOD, we assumed that there is a uniform trip rate within a SCAG TOD for 30-minutes intervals. Thus, we can assign the trips for every 30-minutes to a different TOD based on the half-hourly trips for each TOD. The overlap and reassignment factors are estimated in the table below.

Table 4.3 Time of Day Reassignment Factors

SCAG TOD	LOCUS TOD	Overlapped Time (a)	Hours in SCAG TOD (b)	Reassignment Factor (a ÷ b)
AM	AM	3.0	3.0	1.00
Evening	Late Evening	2.0	2.0	1.00
Midday	AM	0.5	6.0	0.08
Midday	Midday	4.5	6.0	0.75
Midday	PM	1.0	6.0	0.17
Night	Early AM	2	9	0.22
Night	Late Evening	3.0	9.0	0.33
Night	Owl	4.0	9.0	0.44
PM	Late Evening	0.5	4.0	0.13
PM	PM	3.5	4.0	0.88

4.4 Validation and Reasonableness Checks

Reasonableness checks were performed with the redistributed truck trips to ensure that no trip leakages occurred. Table 4.4 lays compares the truck trips by TAZ and Census BG to ensure totals remain constant.

Table 4.4 Total Truck Trip Before and After Redistributions

	Light Trucks	Medium Trucks	Heavy Trucks	Total Trucks
Total at TAZ Level	375,914	319,509	569,974	1,265,396
Total at BG Level	375,914	319,509	569,974	1,265,396

Trip distributions at both the TAZ and the Block group level were compared visually to make sure that trip distributions were not affected by the geographic disaggregation. Figures 4.1-4.3 show that no discernable changes were observed in high frequency locations.

Figure 4.1 Light Truck Trips Before and After Redistribution

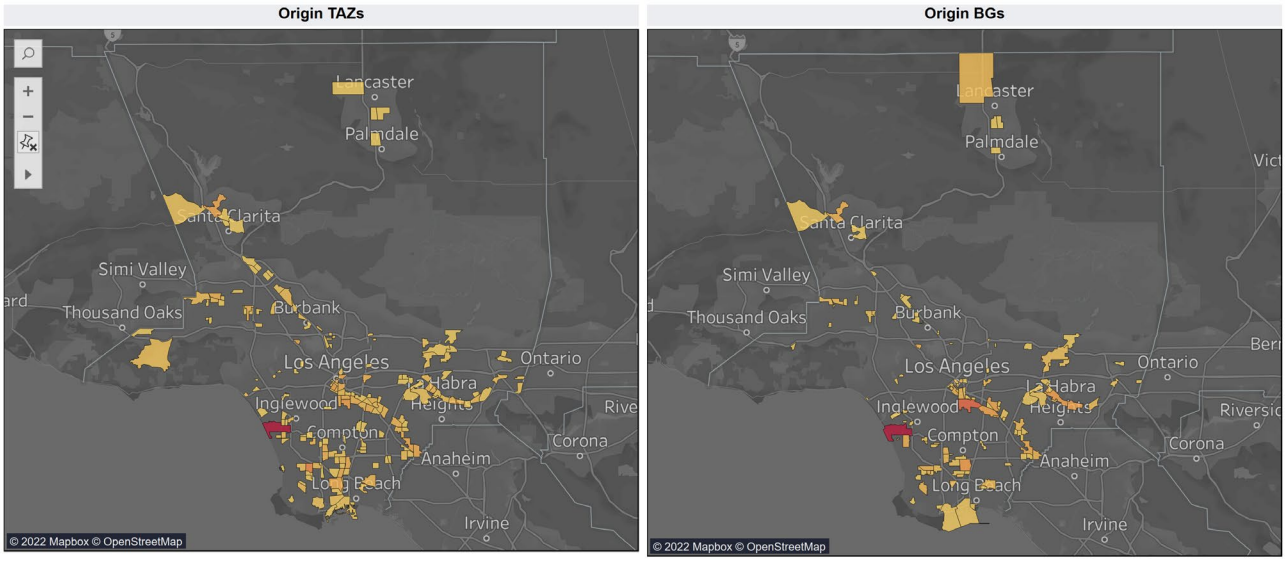


Figure 4.2 Medium Truck Trips Before and After Redistribution

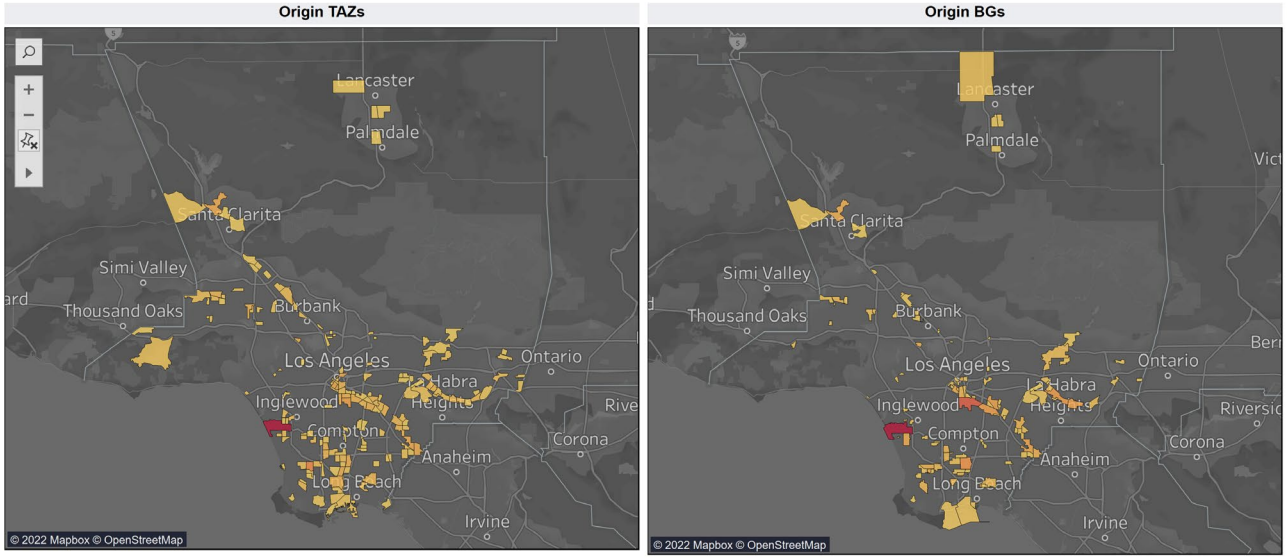
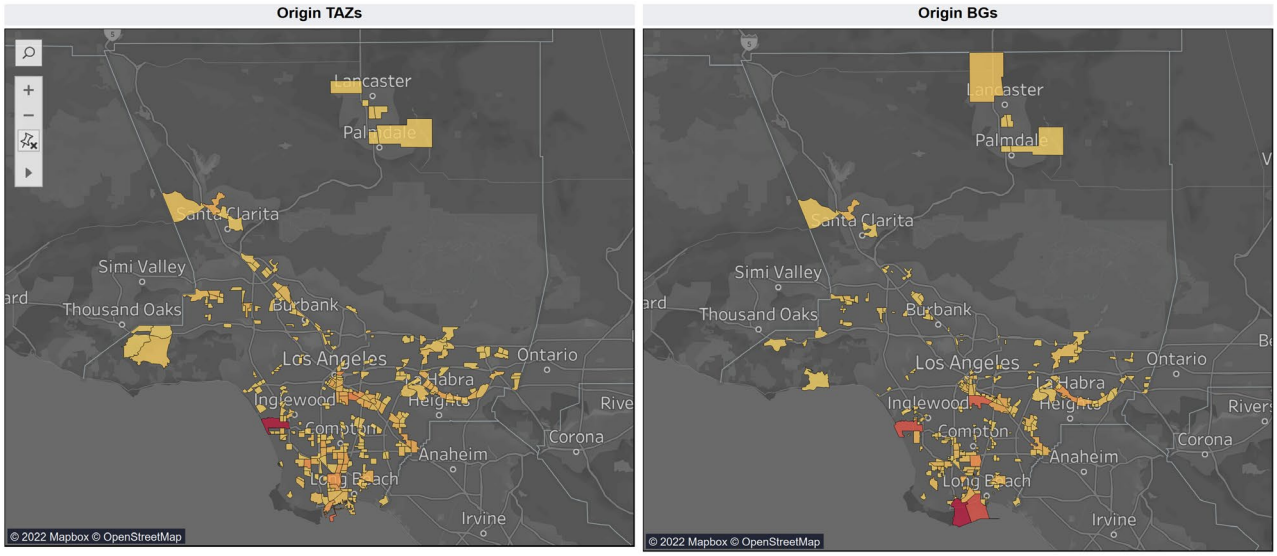


Figure 4.3 Heavy Truck Trips Before and After Redistribution



The redistribution of trips was carried out for the entire SCAG region to ensure that external trips to LA County were captured accurately. As an additional reasonableness check, truck movements at special generators (i.e., ports, intermodal facilities, and external count locations) were assessed both at the TAZ and at the BG level.

Table 4.5 Special Generator Truck Trips Before and After Redistribution

Truck Type	TAZ with Special Generator Trips	BG
Light Truck	198,213	198,213
Medium Truck	166,752	166,752
Heavy Truck	318,922	318,922

Commuting Patterns

5

Commuting patterns have changed tremendously over the past two years, primarily due to the COVID-19 pandemic and because of advances in technology. LOCUS provides a detailed picture of commuting activity to regular activity locations like workplaces or school locations both as aggregate measures of number of commuters and also as detailed metrics of number of commuters commuting on any given day of the week.

For data disclosure concerns, the analysis uses the share of commuters, instead of commuter estimates for both home and work locations to provide an overview of how frequently workers commute to work by day of week. Commuter share measurement is available for all 7 days of the week, from Monday to Sunday.

California Household Travel Survey (CHTS) indicates that 55 percent of teleworkers (workers who have a regular workplace location that is not their home, but they have a teleworking option and work from home at least once a week, on average) commute to work on an average weekday; 75 percent non-teleworkers make a commute trip to work on a weekday. National Household Travel Survey (NHTS) shows that 59 percent of telecommuters, and 64 percent of non-telecommuters commute to work on a weekday.

LOCUS commuting patterns show a similar share (55 percent to 72 percent) of commuters making a commuting trip to their workplace from Monday to Friday before the pandemic in 2019 (Figure 5.1).

In the absence of robust information about telecommuting post-pandemic, greater number of validation checks are not possible. However, the team focused on performing reasonableness checks such as telecommuting shares dropping in 2020 compared to 2019, e.g., from 72 percent to 60 percent on a typical Tuesday, shown in Figure 5.2; and commuters who still went to work dropping higher in high-income communities like Beverly Hills (47%) compared to lower-income communities like Historic-South-Central (which have a higher share of essential workers).

Table 5.1 Share of Workers/Students Commuting on Weekdays

Trip-making Characteristic	CHTS		NHTS	
	Non-telecommuter	Telecommuter	Non-telecommuter	Telecommuter
% of workers making a work trip on any given weekday	75%	55%	64%	59%

Source: 2012-2013 CHTS. 2017 NHTS.
Baseline Estimates of Teleworking, SCAG Future Communities: Future of the Workplace, 2019.

Figure 5.1 LOCUS Commuting Pattern

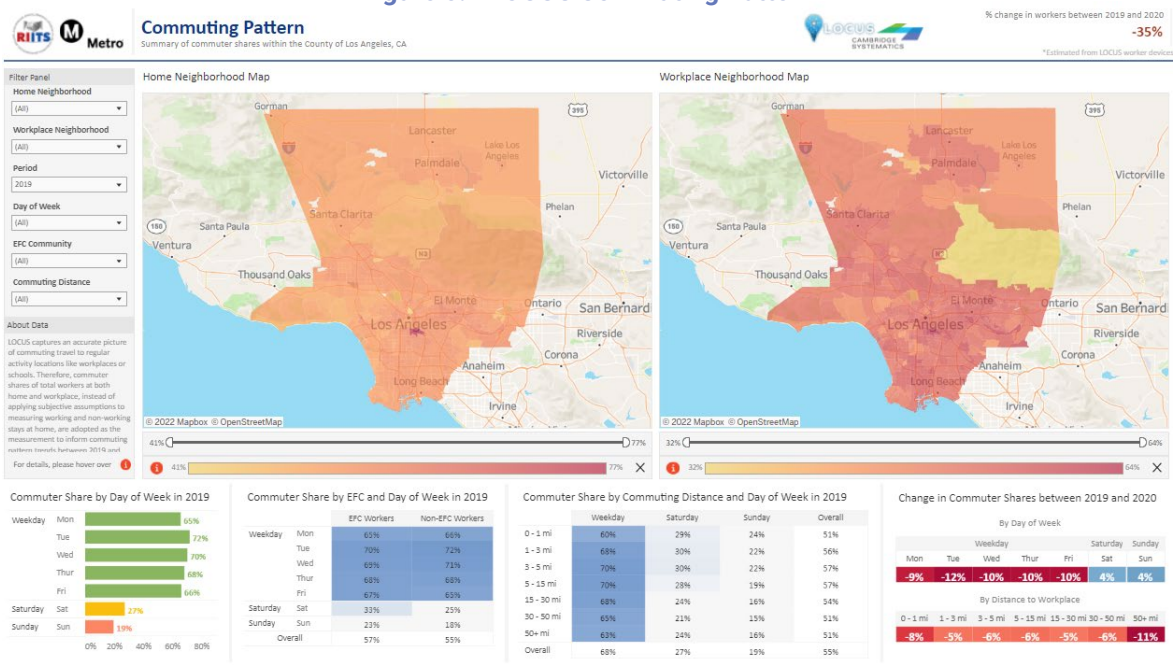


Figure 5.3 Weekday Commuter Share 2019 vs. 2020

Figure 5.3 Commuter Share Beverly Hills & Compton