Replica-rMove Comparison

1.1 INTRODUCTION

While Replica is informed by several real-world datasets, the trips produced by the synthetic population in Replica are simulated. Because of this, it is important to validate Replica estimates like we would validate other simulated data. RSG conducted a household travel survey (rMove) in the Tampa region in Fall 2019, including around 9,000 persons from over 4,500 households making over 75,000 trips. The rMove survey allowed respondents to participate via two survey modes: a smartphone application that collected 7 days of travel data for each participant ("rMove app") or a web-based survey that collected a single day of travel data ("rMove diary"). While data from rMove app participants include both weekday and weekend travel days, rMove diary participants were all assigned a single weekday (Tuesday, Wednesday, or Thursday) and therefore no weekend travel days are present in the rMove diary sample. This document compares Replica's 2019 estimates for the Tampa region to this travel survey to a measure of relative confidence in different aspects of Replica's simulation.

Generating an apples-to-apples comparison

Before comparing Replica estimates to available HTS data, a fundamental step is to define the sub-population within the larger Replica population that facilitates an apples-to-apples comparison between the two datasets. Because the Replica data used in this study includes trips taken by all persons in the D7 region, including visitors, non-residents were first removed from the Replica data—in other words, the comparisons presented here focuses on residents of the five Florida counties in District 7: Hillsborough, Pinellas, Pasco, Hernando, and Citrus counties in 2019. It's important to note that the Replica population differs slightly form the expanded rMerge population (Table 1). This may be because of differences in how seasonal and university students are defined in each dataset. Further, the Replica population excludes children under the age of 3.

1.2 TRIP RATE COMPARISON

The first set of comparisons between Replica and the Tampa HTS data focus on trip rate, both in aggregate and by various demographic characteristics of the two datasets including age, employment status, household income, household size, vehicle availability, travel purpose, and travel mode. HTS trip rates are calculated using the expanded data (i.e., incorporating survey weights) and, unless otherwise noted, use data from both rMove app and eMove diary participants. Because only weekday trips are weighted in the rMove data, the trip rate comparison described below focuses exclusively on weekday trips.

Age

Across nearly all age groups, trip rates in the Replica data are roughly 30% lower than trip rates derived from the HTS data (Table 1). Interestingly, Replica trip rates are higher for one category

(16-17), though the trip rate for this group is much lower than expected in the HTS data (Figure 1). It is also highly likely that trip rates for persons in the youngest two age groups in the Replica data are asserted: in both the lte_4 and 5-15 age groups, all Replica persons have exactly 2 trips (home-school-home_ per day). Finally, rMove trips per home-based tour rate are higher, showing more trip chaining behavior compared to Replica trips (Figure 2).

TABLE 1. TRIP RATE BY AGE GROUP

|	Replica			rMove		
	Persons	Trips	Trip rate	Persons	Trips	Trip rate
lte_4	30,230	60,460	2	140,723	396,235	2.82
5_15	330,754	661,508	2	356,857	1,005,741	2.82
16_17	72,120	175,459	2.43	66,755	115,323	1.73
18_24	233,408	990,843	4.25	180,143	782,814	4.35
25_34	392,564	1,677,731	4.27	289,024	1,417,191	4.9
35_44	351,125	1,507,404	4.29	401,818	2,279,535	5.67
45_54	398,144	1,718,455	4.32	377,495	1,944,743	5.15
55_64	386,423	1,686,603	4.36	414,612	2,003,161	4.83
65_74	304,163	1,201,212	3.95	479,391	2,426,556	5.06
75_plus	262,875	964,100	3.67	147,636	636,169	4.31
Total	2,761,806	10,643,775	3.85	2,854,455	13,007,465	4.56
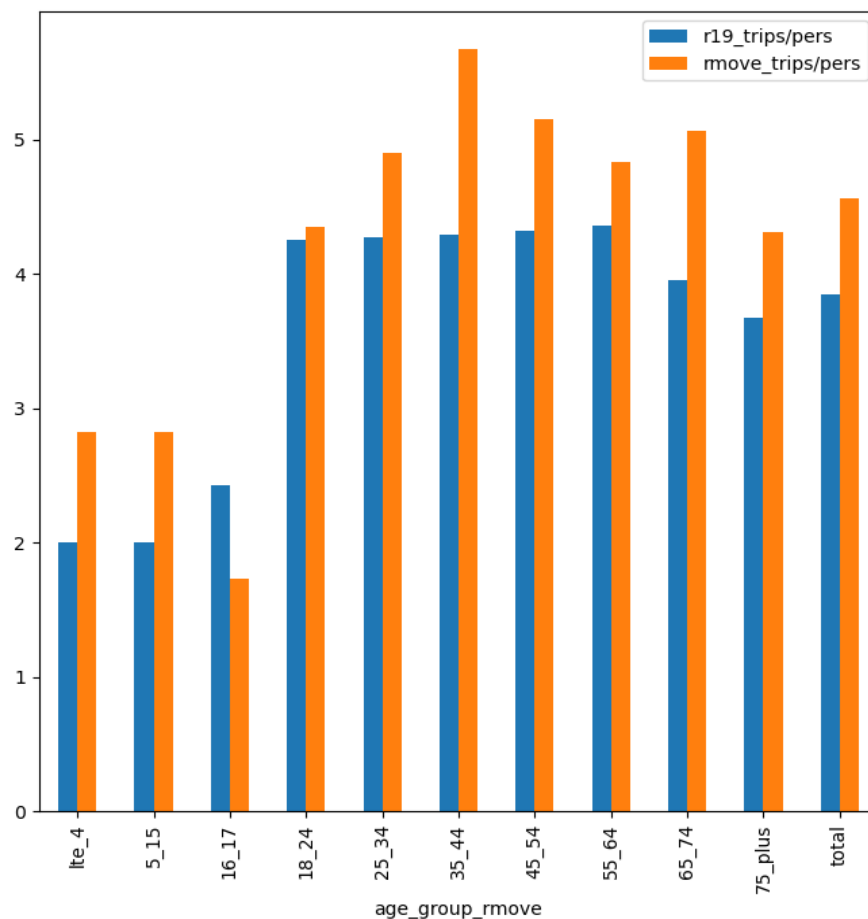
FIGURE 1. TRIP RATE BY AGE GROUP
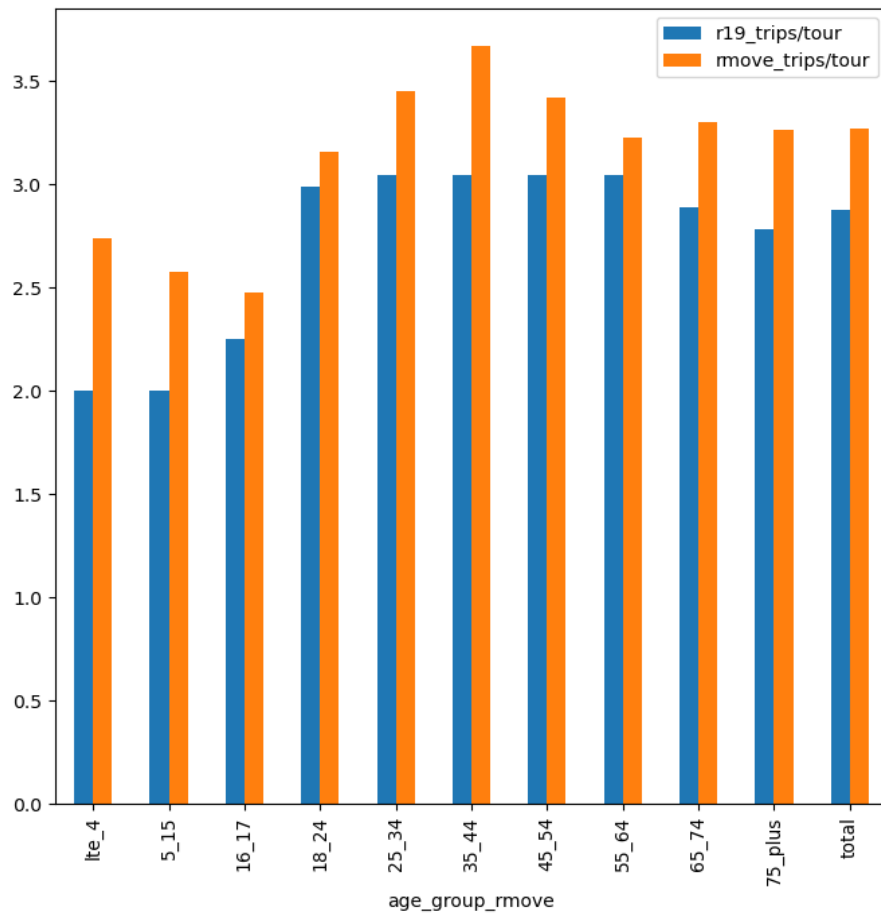
FIGURE 2. TRIP PER TOUR BY AGE GROUP

Employment Status

Employment status directly affects trip rates, as workers usually make trips to work, and have less time for non-work activities. Replica and rMove trip rates were compared across four employment statuses: employed, unemployed, not in labor force, and under 16. An important difference to note in this comparison in that Replica classifies retired persons as not in the labor force while rMove classifies these persons as unemployed; thus, the Replica data has many more persons not in the labor force and many fewer unemployed persons as compared to the rMove data. Additionally, Replica does not differentiate between full-time and part-time workers; both are included in the employed category. Overall, the trip rate for workers in Replica is 22% less than worker trip rates in rMove, which is a higher difference than the global difference of 18% (Table 2, Figure 3). Interestingly, Replica underestimates trip rates to a greater extent for those not in the labor force and has nearly identical trip rates for unemployed persons; however, it is difficult to draw strong conclusions from these comparisons given the definitional difference in the not in labor force and unemployed categories between the two datasets as described above.

TABLE 2. TRIP RATE BY EMPLOYMENT STATUS

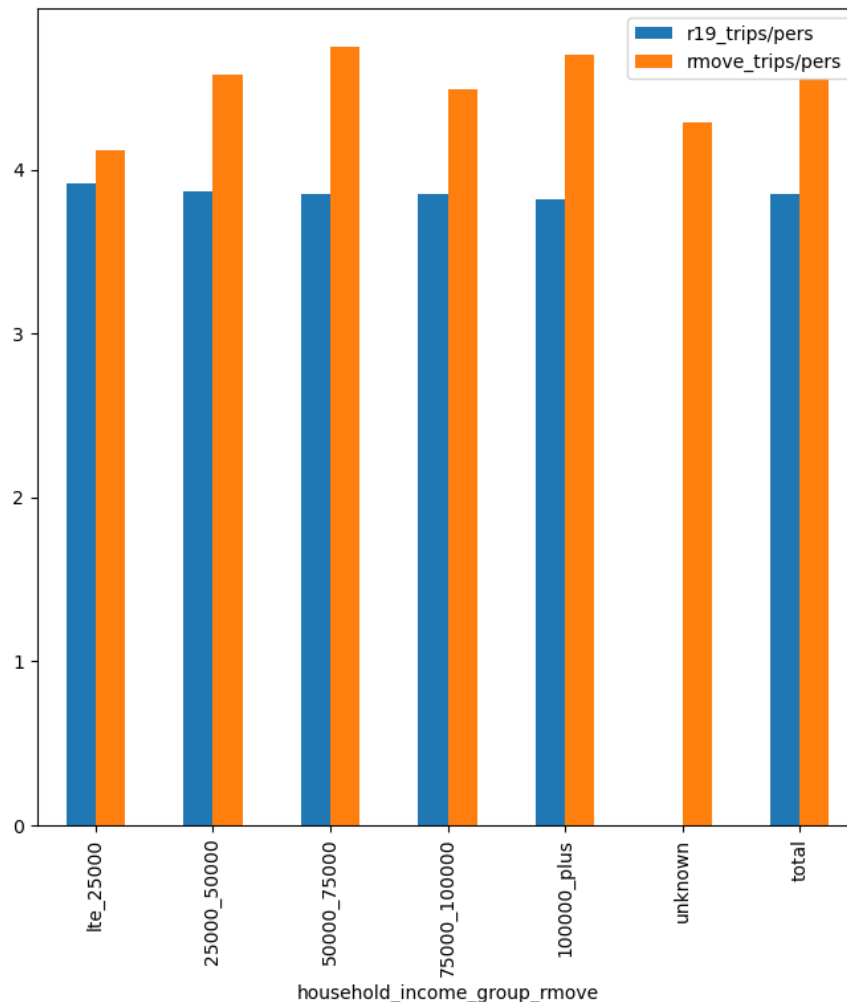|	Replica			rMove		
	Persons	Trips	Trip rate	Persons	Trips	Trip rate
Employed	1,418,433	5,929,262	4.18	1,360,466	7,294,415	5.36
Unemployed	74,932	325,218	4.34	975,151	4,191,580	4.3
Notinlf	907,457	3,667,327	4.04	21,257	119,495	5.62
Under_16	360,984	721,968	2	497,581	1,401,975	2.82
Total	2,761,806	10,643,775	3.85	2,854,455	13,007,465	4.56
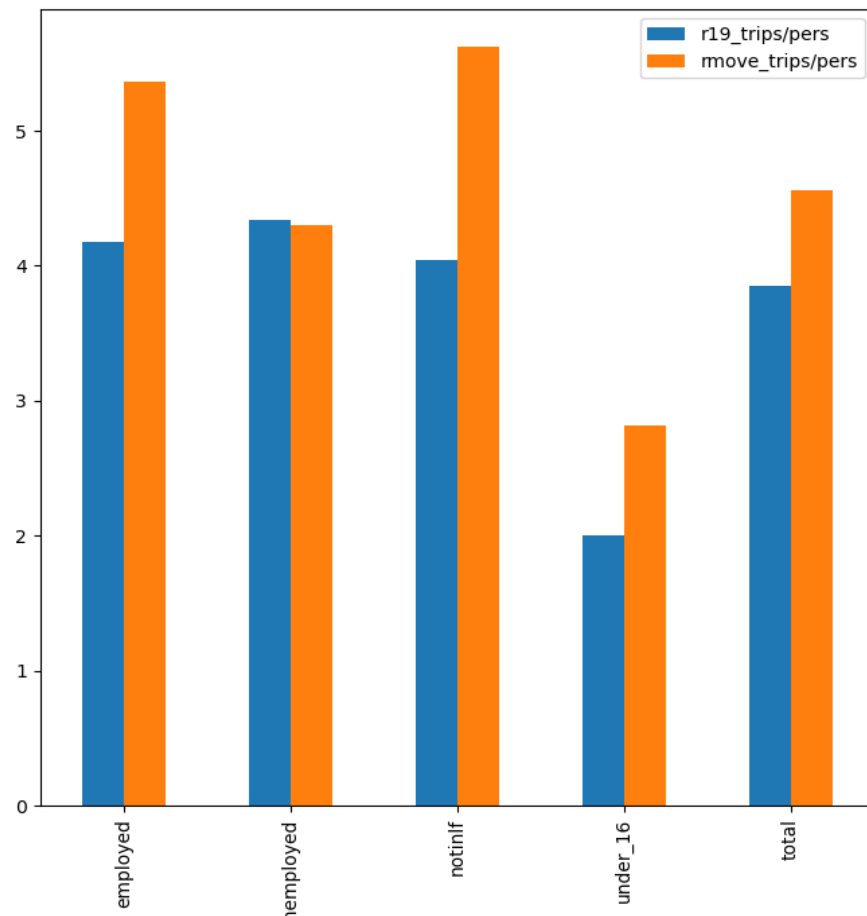
FIGURE 3. TRIP RATE BY EMPLOYMENT STATUS

Household Income

It is generally understood that household income has a positive impact on trip rate because households with higher income tend to have more budget for discretionary activities. The relationship between trip rates and household income is largely as expected in the rMove data: the lowest income group has the lowest trip rate, while the highest income group has the highest second-highest trip rate. Conversely, Replica data shows that trips per person are lower as income goes up, which is contrary to expectations (Table 3, Figure 4).

TABLE 3. TRIP RATE BY HOUSEHOLD INCOME

|	Replica			rMove		
	Persons	Trips	Trip rate	Persons	Trips	Trip rate
lte_25000	414,757	1,626,299	3.92	428,444	1,766,405	4.12
25000_50000	567,705	2,196,740	3.87	685,235	3,138,611	4.58
50000_75000	512,532	1,972,651	3.85	509,357	2,417,540	4.75
75000_100000	385,037	1,481,128	3.85	375,199	1,685,228	4.49
100000_plus	881,775	3,366,957	3.82	799,833	3,757,900	4.7
Unknown	0	0	0	56,387	241,782	4.29
Total	2,761,806	10,643,775	3.85	2,854,455	13,007,465	4.56

FIGURE 4. TRIP RATE BY HOUSEHOLD INCOME

Household Size

In both rMove and Replica datasets, trip rates decrease as household size increases for both, as household maintenance activities can be shared with larger households. Comparing the two datasets, rMove trip rate per person is 28% higher than Replica trip rate for 1- and 2-person households, significantly higher than the global difference of 18% between the two (Table 4, Figure 5).

TABLE 4. TRIP RATE BY HOUSEHOLD SIZE

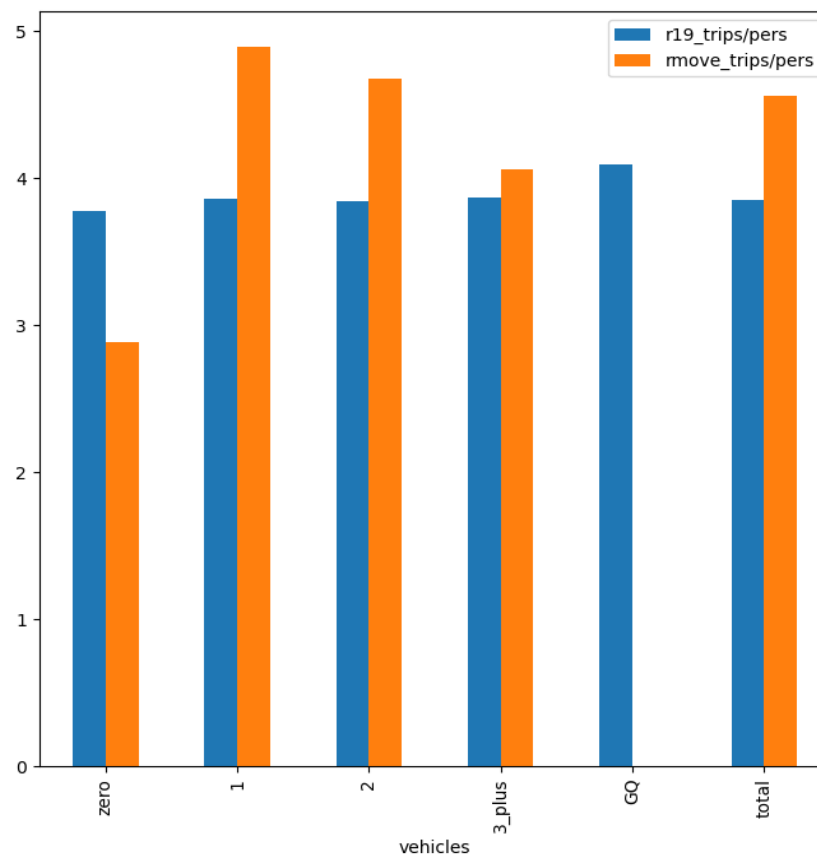|	Replica			rMove		
	Persons	Trips	Trip Rate	Persons	Trips	Trip Rate
1_person	332,516	1,355,553	4.08	406,904	2,245,208	5.52
1_person_GQ	41,451	169,382	4.09	20,732	119,325	5.76
2_person	838,677	3,403,596	4.06	937,057	4,714,912	5.03
3_person	483,803	1,886,639	3.9	562,145	2,198,583	3.91
4_person	476,879	1,748,137	3.67	505,117	1,966,508	3.89
5_person	259,886	928,132	3.57	310,184	1,402,859	4.52
6_person	140,208	501,447	3.58	91,993	343,335	3.73
7_plus_person	188,386	650,889	3.46	20,322	16,734	0.82
Total	2,761,806	10,643,775	3.85	2,854,455	13,007,465	4.56
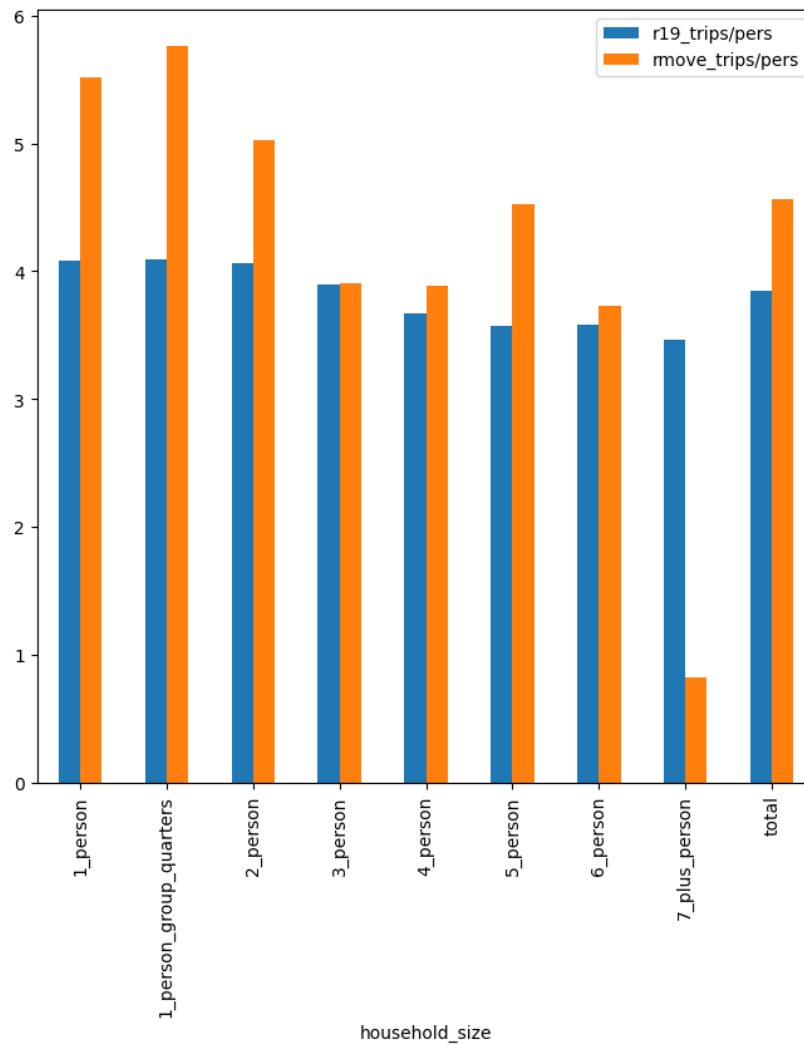
FIGURE 5. TRIP RATE BY HOUSEHOLD SIZE

Household Vehicles

The number of household vehicles available affects the trip rate as zero-auto households have less mobility and accessibility compared to households with vehicles. The group quarters (GQ) category in Replica vehicles is not defined in the documentation, so rMove was not mapped to this category. Interesting, the GQ category trip rate is higher than all other categories in Replica. Replica data shows that people in zero-auto households have similar rates as people with household vehicles, despite having less accessibility (Table 5, Figure 6). This is a dramatic difference compared to the rMove data.

TABLE 5. TRIP RATE BY HOUSEHOLD VEHICLES

	Replica			rMove		
	Persons	Trips	Trip rate	Persons	Trips	Trip rate
Zero	93,242	352,128	3.78	112,002	322,153	2.88
1	802,885	3,099,330	3.86	899,776	4,400,912	4.89
2	1,145,225	4,394,368	3.84	1,289,704	6,039,855	4.68
3_plus	679,003	2,628,567	3.87	552,974	2,244,545	4.06
GQ	41,451	169,382	4.09	-	-	
Total	2,761,806	10,643,775	3.85	2,854,455	13,007,465	4.56

FIGURE 6. TRIP RATE BY HOUSEHOLD VEHICLES

Trip Rates by Travel Purpose

In addition to the aggregate comparison described above, work, school, and maintenance purposes by age group were compared between Replica and rMove.

Work trips

In both datasets, work trips are predominantly made by adults, with some by young adults aged 16-17 and retirees (Figure 7). rMove trip rates seem to be higher than expected in the 45-54 age category, likely due to small samples sizes when splitting the rMove data into small subcategories. While Replica trip rates tend to be lower than rMove across nearly all other dimensions, it is interesting to note that the Replica work trip in the 65+ age category is much higher than the rMove work trip rate for this same group. This seems to indicate that many retirement-age individuals still make work trips in the Replica data. Finally, rMove work trip per person aged 18-24 is much higher than Replica work trip rate. In addition to sampling bias, this can be affected by the breakdown of full-time and part-time workers, and how part-time workers working more than one jobs are represented.
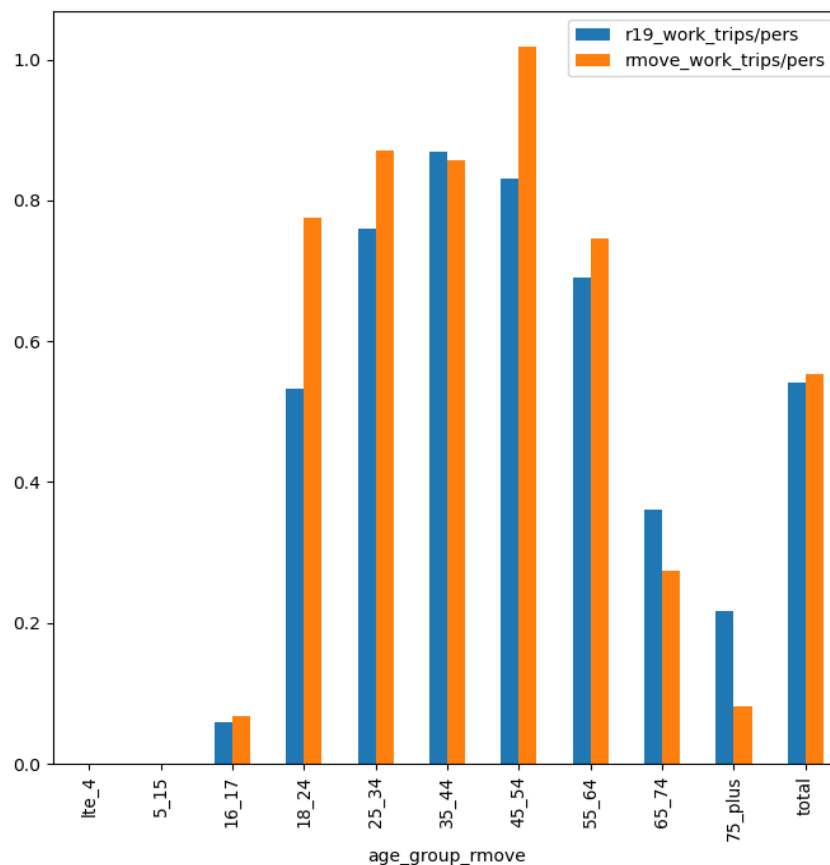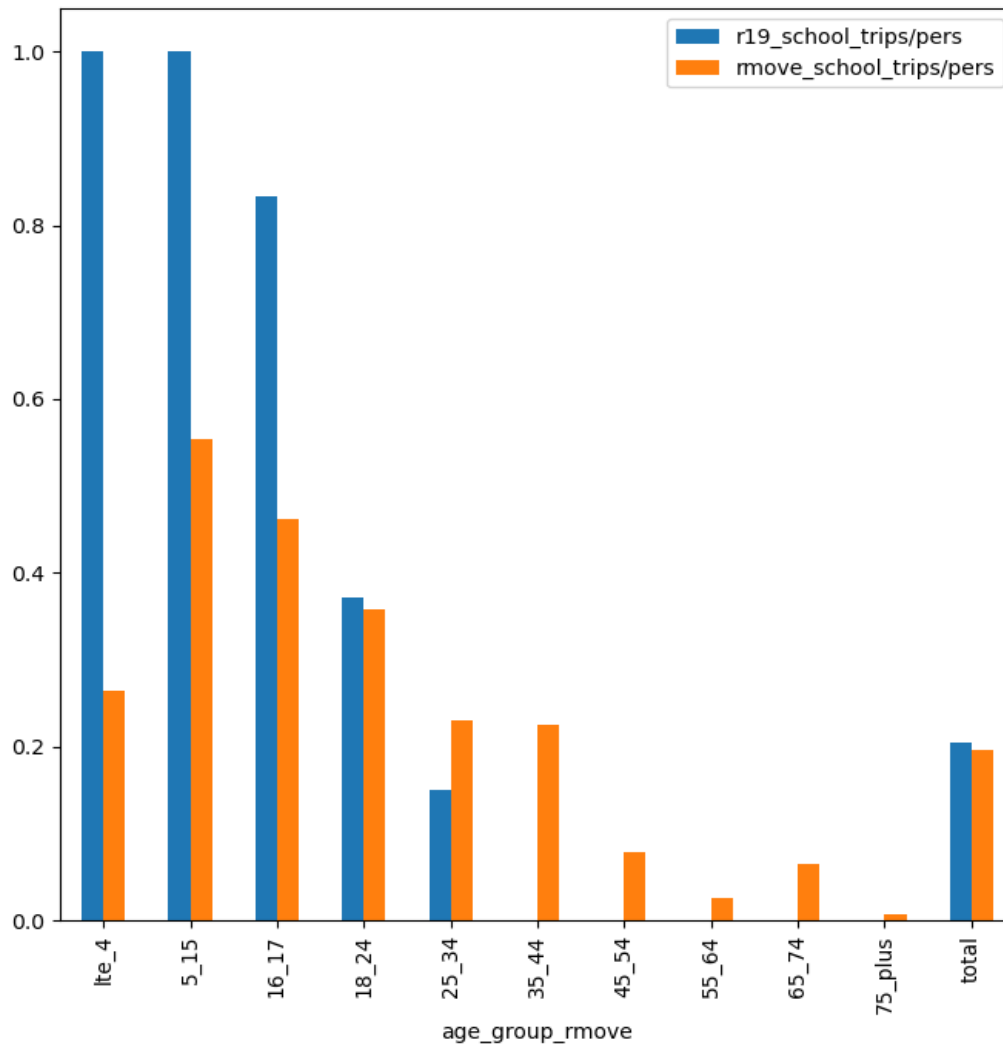

FIGURE 7. TRIP RATE BY AGE GROUP AND WORK PURPOSE

School trips

In both datasets, school trips are predominantly made by aged 24 or under (Figure 8). As mentioned previously, Replica data assumes that every person aged 15 and under travels to school every day and seems to assert trips for these age groups, resulting in a trip rate exactly equal to 1 (representing the home -> school trip). Replica data also assumes that every person aged 35 and over does not make school trips.

FIGURE 8. TRIP RATE BY AGE GROUP FOR SCHOOL PURPOSE

Maintenance trips

Maintenance trips include household maintenance activities such as errands, shopping, and picking up or dropping off children. Across nearly all age categories, the rMove maintenance trip rate is much higher than the Replica maintenance trip rate (Figure 9). Further, Replica seems to assume that children aged 15 and under do not make maintenance trips or accompany their parents for errands and shopping trips.
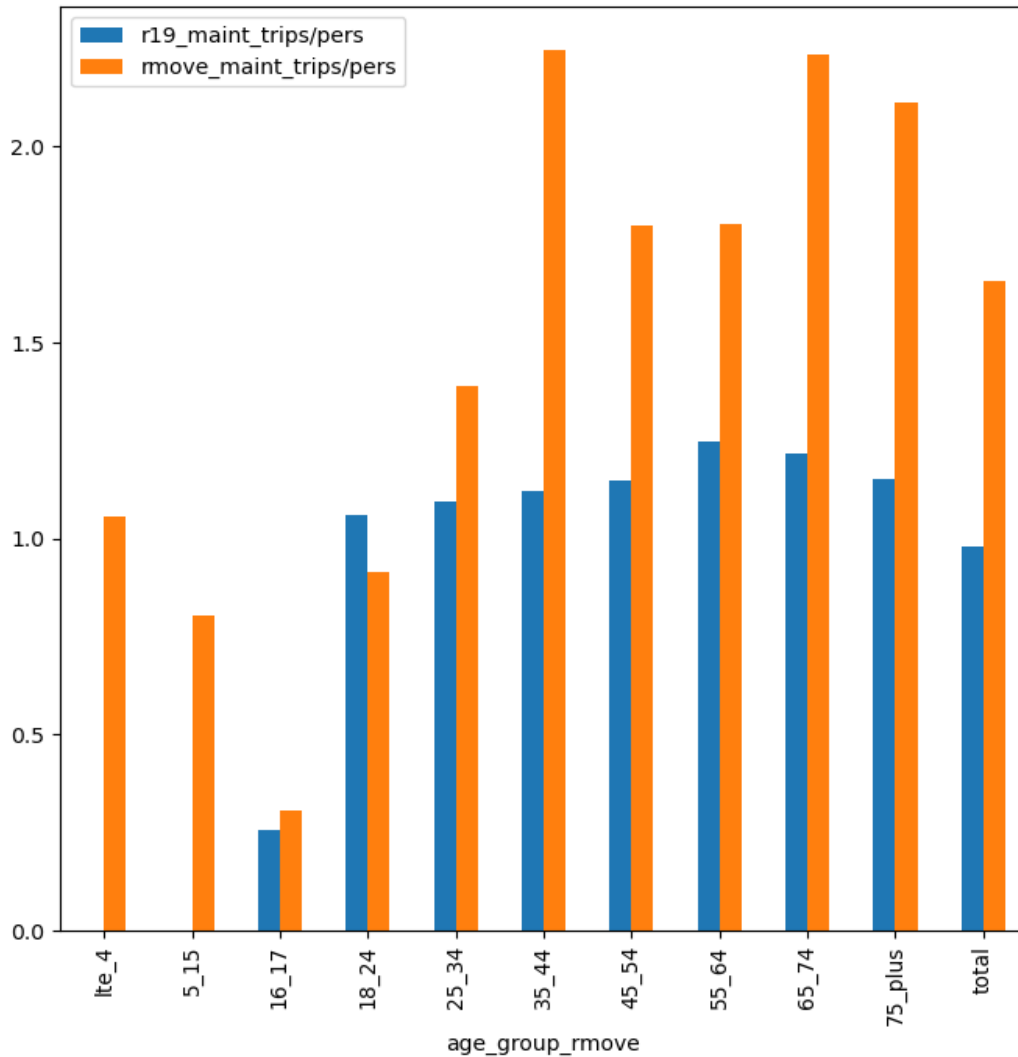
FIGURE 9. TRIP RATE BY AGE GROUP FOR MAINTENANCE PURPOSE

Trip Rates by Travel Mode

Trip rates for auto driver (the person drives a vehicle, with or without accompanying passengers), auto passenger (the person is a passenger in a private vehicle), public transit (a trip including one public transit ride, possibly mixed with other modes), and walking (walking only) modes by age group were also compared between Replica and rMove.

Auto driver trips

Overall, the rMove auto driver trip rate is 30% higher than the Replica auto driver trip rate (Figure 10). This difference is nearly double the global difference in trip rates between the two datasets. Interestingly, the Replica auto driver trip rate is higher than rMove trip rate for 16- and 17-year-olds; however, rMove has an unexpectedly low trip rate for persons in this age group (Table 1).

10
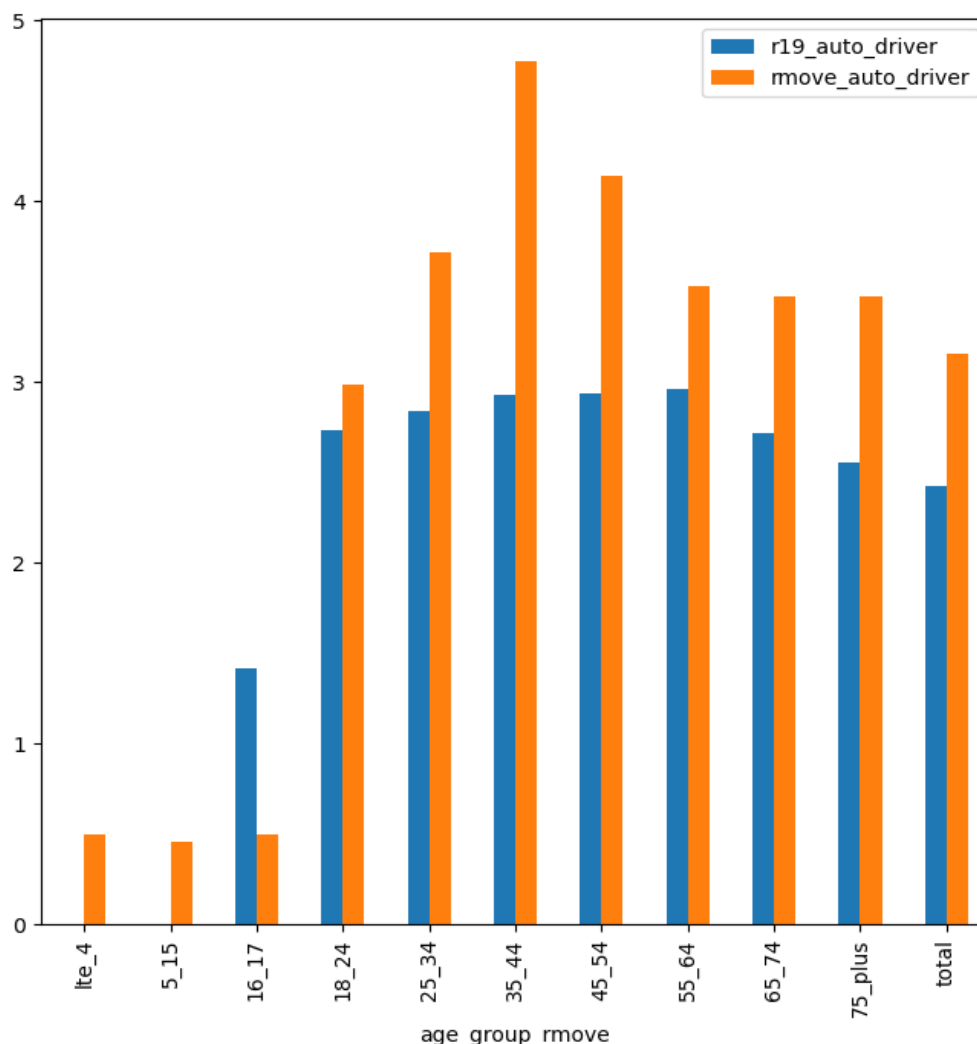
FIGURE 10. TRIP RATE BY AGE GROUP FOR AUTO DRIVER MODE

Auto passenger ("carpool") trips

Replica classifies all trips taken as a passenger in a vehicle as a carpool trip, even if that trip is with someone from the same household. Replica has much higher auto passenger trip rates across all adults above age 25 and much lower auto passenger trip rates for individuals under 25 (Figure 11). For adults, this seems to indicate that Replica dramatically over-estimates carpooling. In fact, Replica indicates an average auto occupancy rate of 1.4 persons per vehicle compared to 1.24 persons per vehicle in rMove. Conversely, Replica dramatically underestimates carpooling for children, which is consistent with the earlier finding that Replica children do not make non-school trips. Interestingly, if Replica included carpooling trips for children at rates closer to those observed in the survey, auto occupancy would increase further above 1.4 persons per vehicle.
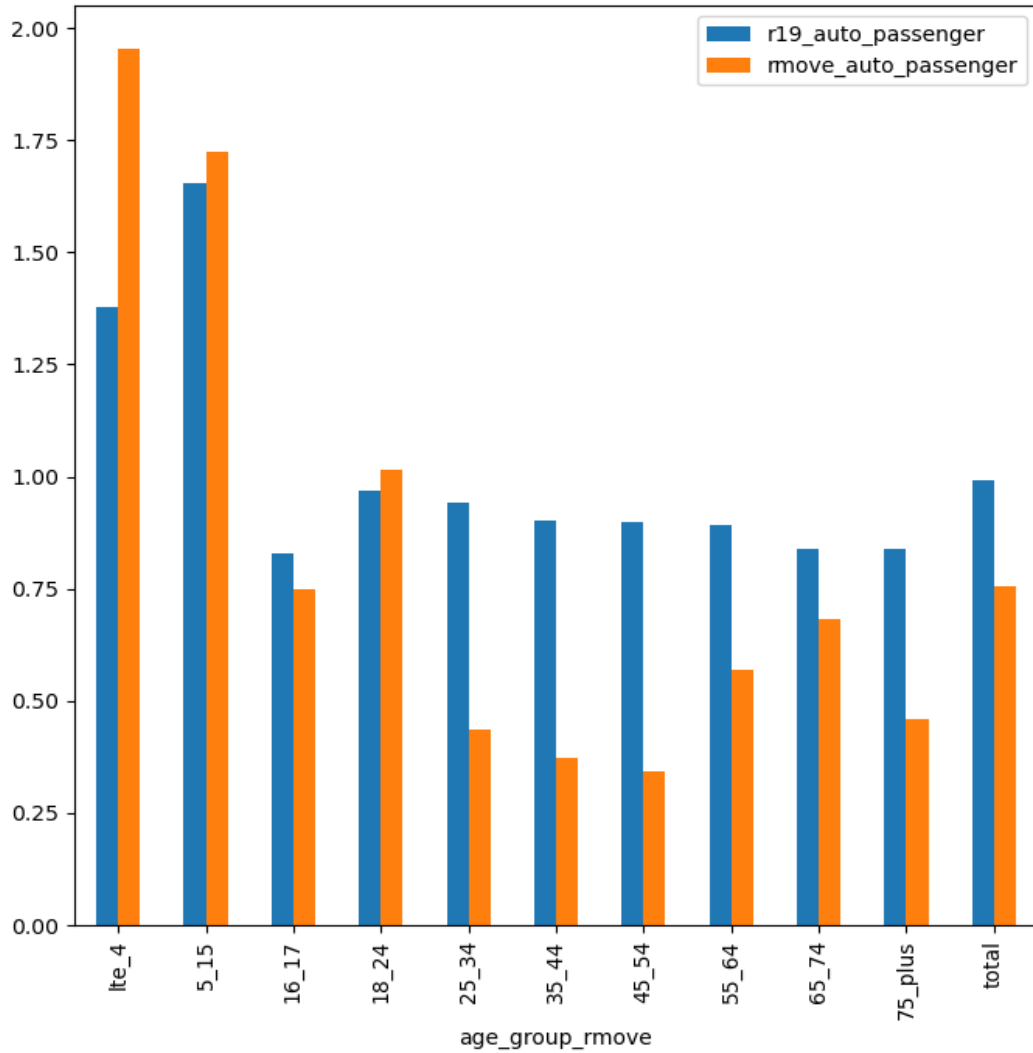
FIGURE 11. TRIP RATE BY AGE GROUP FOR AUTO PASSENGER MODE

Transit trips

Overall, the rMove transit trip rate is 53% higher than Replica transit trip rate (Figure 12). The largest difference is observed in the 16-17 age category; however, rMove data for this group may be unreliable due to sampling issues. Given the large differences between these two datasets and the relatively low number of transit trips in the rMove dataset, external sources such as transit on-board surveys may be a more reliable source to help calibrate and validate transit usage.
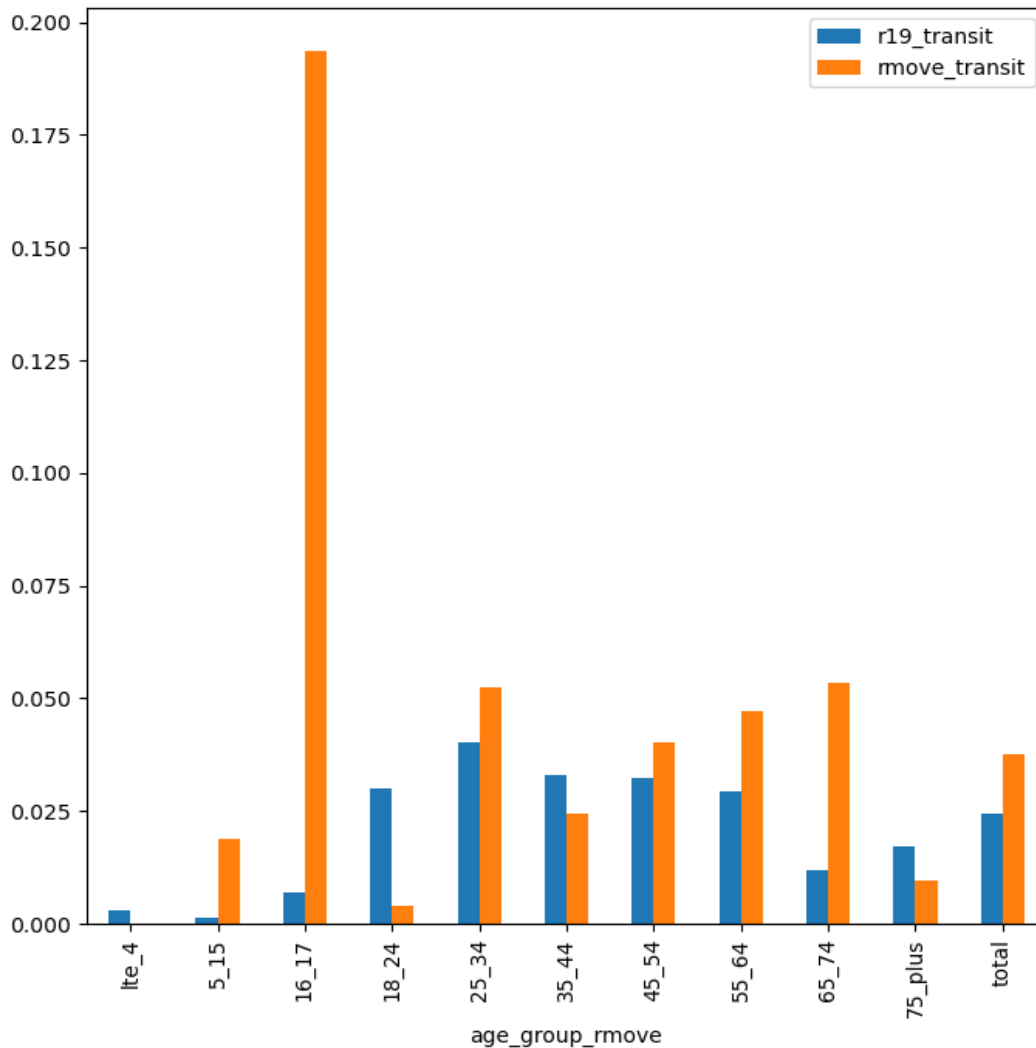
FIGURE 12. TRIP RATE BY AGE GROUP FOR TRANSIT MODE

Walk trips

Overall, the rMove walk trip rate is 38% higher than Replica. However, there are large differences by age group: for adults aged 25+, the rMove walk trip rate is nearly 60% higher than Replica while the walk trip rate for children is lower in rMove than in the Replica data (Figure 13). Curiously, the Replica walk trip rate for persons aged less than 4 is higher than all other age groups, suggesting that they walk without being accompanied by an adult.
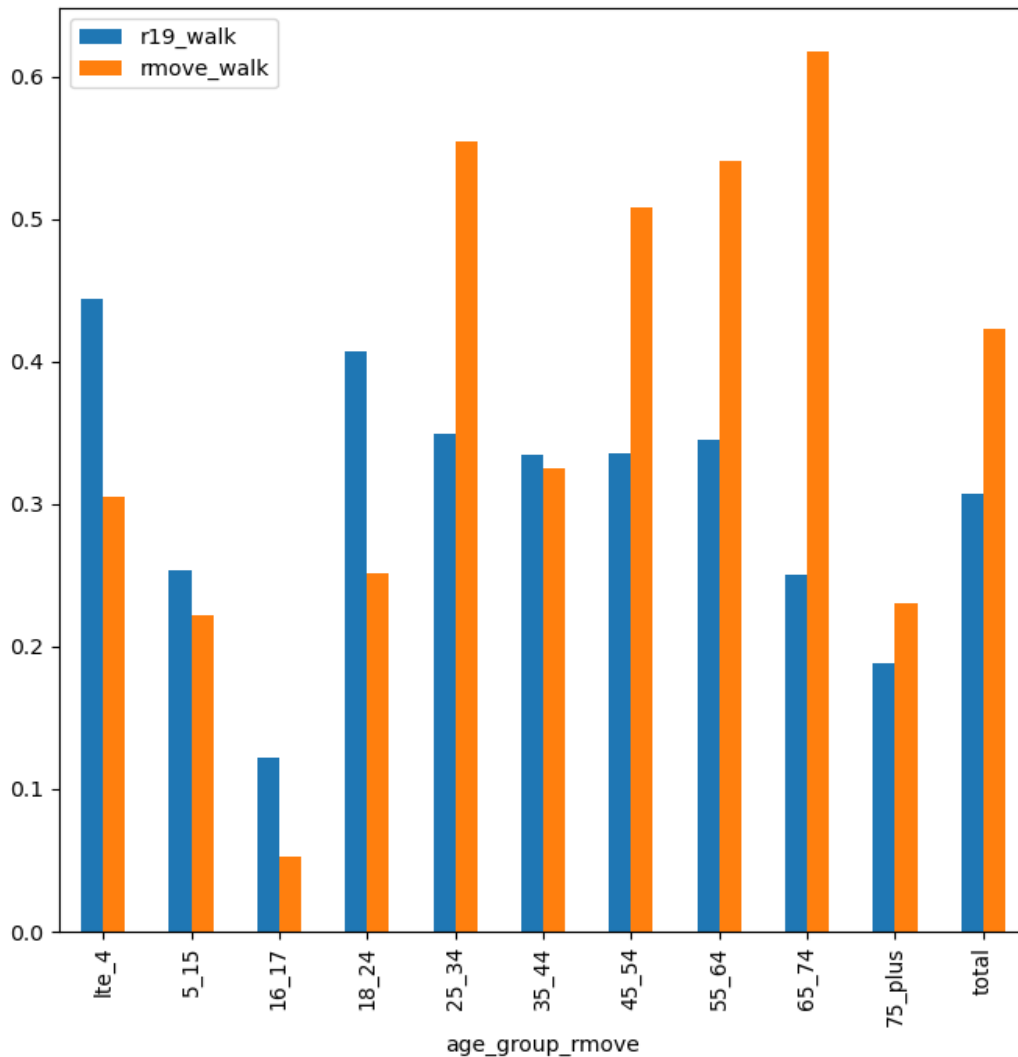
FIGURE 13. TRIP RATE BY AGE GROUP FOR WALK MODE

Trip Rates by Survey Mode

As noted previously, the rMove household travel survey was collected via mobile apps tracking locations, call center staff recording respondent's information, and web survey filled out by the respondent. Trip under-reporting is very common due to shorter trips being forgotten, or a proxy reporting being unaware of trips made by other members of the household. By using the rMove mobile app to track and record trips, trip rates that are frequently forgotten in call center or web responses can be adjusted for trip under-reporting. As such, rMove survey modes generally have higher trip rates than Replica. rMove web version has the fewest number of responses within rMove, so it has higher fluctuations in trip rates by age group due to small samples (Figure 14).
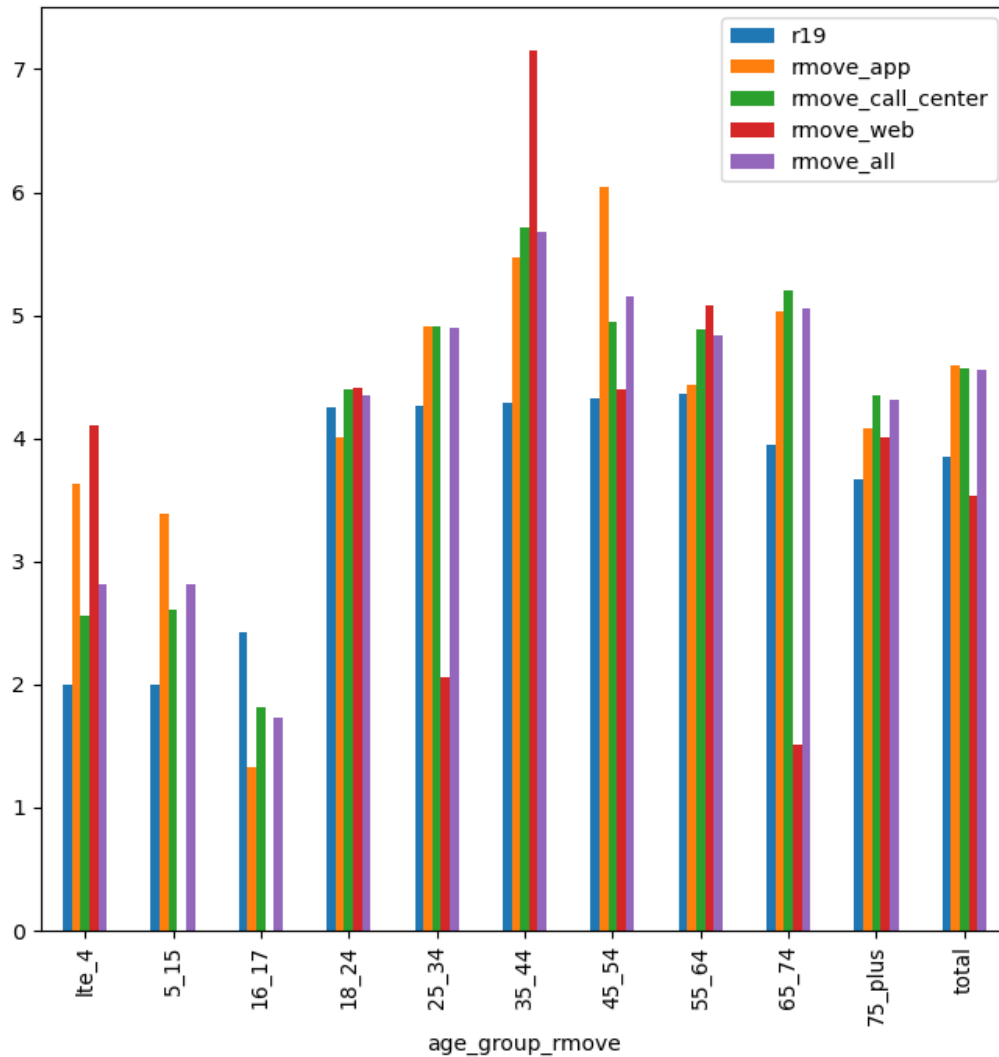
FIGURE 14. TRIP RATE BY AGE GROUP AND SURVEY MODE

1.3 ACTIVITY PATTERN

An additional method of validating the trip rates and travel patterns is by looking at the activity pattern of each person. For a typical weekday, three activity patterns can be defined:

- **Mandatory:** a person makes at least one work or school trips
- **Non-mandatory**: a person does not make any work or school trips, but makes at least one trip
- **Home:** a person stays at home all day.

These definitions were used to categorize person-days in both the Replica and rMove datasets, and the distribution of these activity patterns were compared between the two datasets for each age group. The distribution of activity patterns varies significantly by age group based on the ability and needs for travel (Figure 15):

- Replica assumes that everyone goes out on a typical weekday, and nobody stays at home.
- Replica assumes that everyone aged 15 and under travels to school every day, so there are no home patterns. rMove shows that pre-school children might go to school, go out for other activities, or stay at home.
- Replica shows higher mandatory rate than rMove for all age groups.


FIGURE 15. ACTIVITY PATTERN BY AGE GROUP

1.4 TRIP CHARACTERISTICS

The second set of comparisons between Replica and the rMove data focuses on trip characteristics, including trip length distribution, trip mode, and trip time-of-day profiles. In general, Replica trips tend to be slightly longer than observed trips in the rMove dataset. This difference is especially pronounced for walking and biking modes (Table 6). While weekday auto driver trips are roughly 40% longer in Replica than in the rMove app sample, walk trips are nearly 60% longer, carpool trips over 75% longer, and bike trips are roughly 3 times longer. These differences are generally more pronounced on weekends. Box plots of these distributions further highlight these differences: for bike trips, the median distance in Replica (line in the middle of the box) is above the rMove interquartile range (Figure 16).

TABLE 6. MEDIAN TRIP DISTANCES (MILES), BY DATA SOURCE, TRIP MODE, AND DAY TYPE.

|	Replica		rMove app		rMove diary	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Auto driver	5.46	4.92	3.81	3.39	4.52	-
Carpool	5.74	6.64	3.25	3.76	4.30	-
Walk	0.50	0.47	0.32	0.28	0.32	-
Bike	3.85	4.18	1.03	1.44	1.49	-
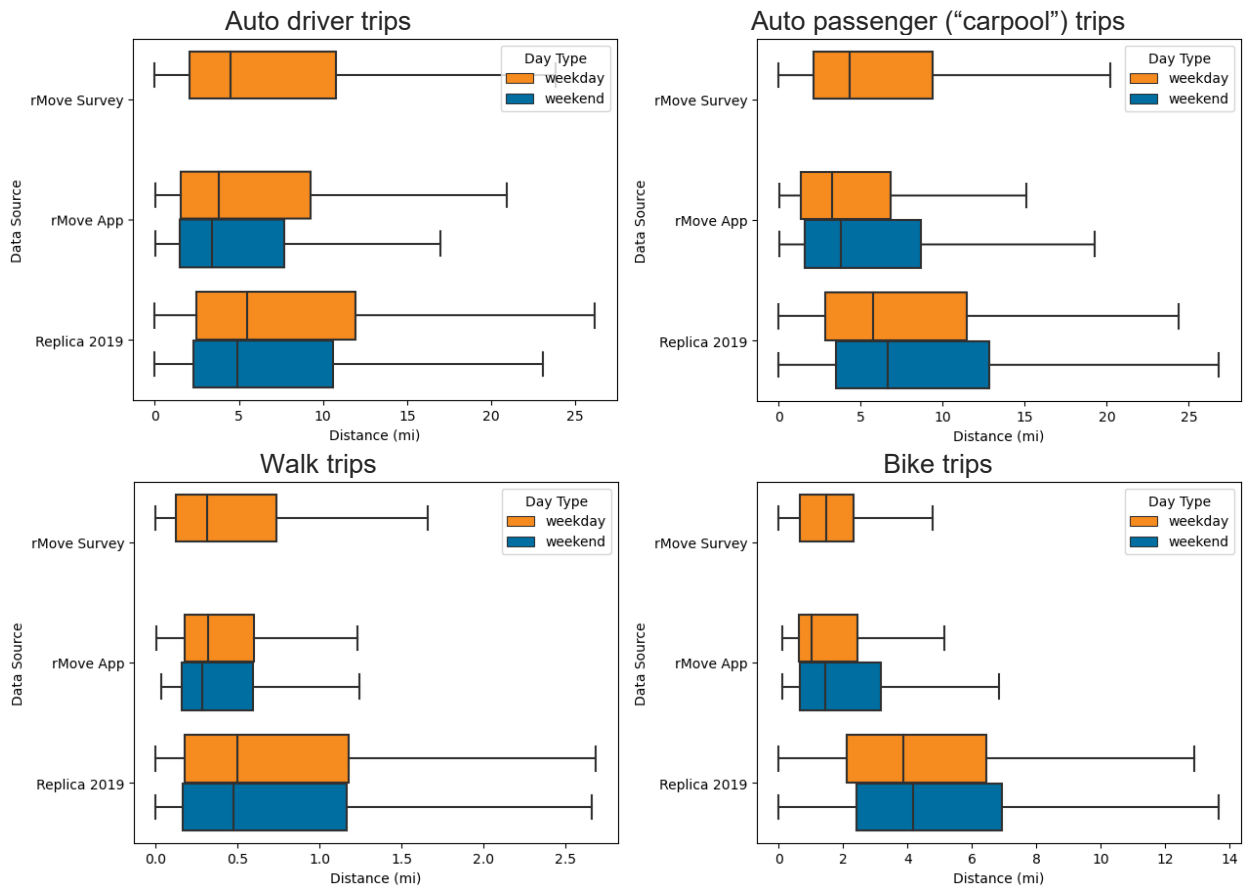
FIGURE 16. BOXPLOT OF TRIP DISTANCE DISTRIBUTIONS, BY DATA SOURCE, TRIP MODE, AND DAY TYPE.

While Replica trip distances are generally longer than rMove trip distances, there is substantial variation across trip purposes. Interesting, Replica work trips are slightly shorter than rMove work trips on both weekdays and weekends (Table 7). Across all other trip purposes, however, Replica trip distances are substantially longer than those observed in the rMove survey: Replica non-home-based work trips are nearly double the length of rMove non-home-based work trips, home-based other trips are roughly 50% longer in Replica, and non-home-based other trips are more than twice as long in Replica. As before, box plots of these distributions further reinforce these differences: for both non-home-based work and non-home-based other trips, the median trip length is nearly outside of the interquartile range in the Replica data (Figure 17).

TABLE 7. MEDIAN TRIP DISTANCES (MILES), BY DATA SOURCE, TRIP PURPOSE, AND DAY TYPE.

|	Replica		rMove app		rMove diary	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
HBW	8.34	6.69	9.34	6.91	9.97	-
NHBW	4.71	4.84	2.46	2.40	5.12	-
HBO	4.00	4.47	2.75	3.18	3.00	-
NHBO	5.32	5.25	2.65	2.42	3.26	-
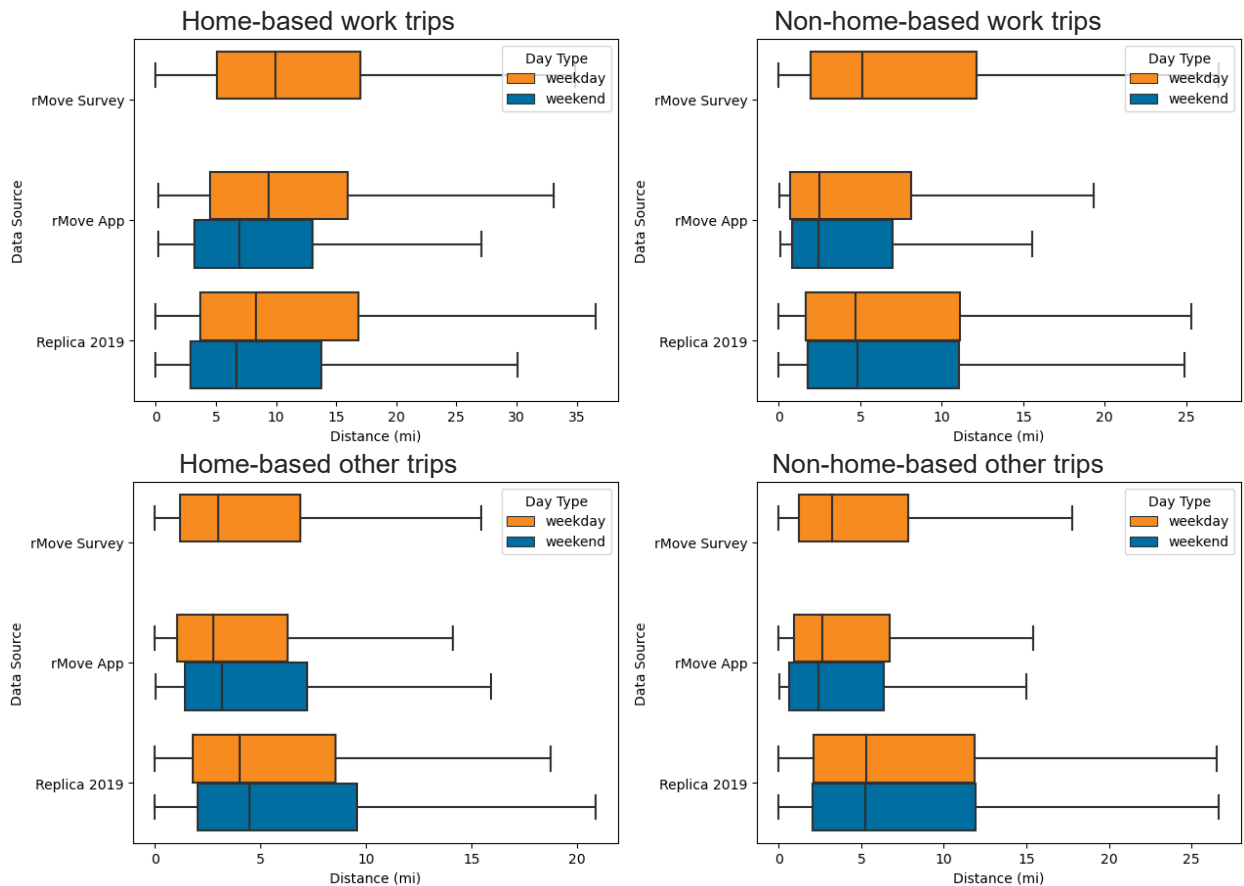
FIGURE 17. BOXPLOT OF TRIP DISTANCE DISTRIBUTIONS, BY DATA SOURCE, TRIP PURPOSE, AND DAY TYPE.

In contrast to trip distance, which varies substantially across most dimensions between Replica and rMove data, trip departure time-of-day profiles in the Replica data are quite consistent with those in the rMove data for both weekdays and weekends. Replica displays expected AM and PM peaking for both auto driver and auto passenger trips on weekdays, with a single, less pronounced midday peak on the weekends (Figure 18). Replica does tend to have slightly fewer early morning weekend trips, which are captured in the rMove data, and looks to have a slight bias towards an earlier AM peak than observed in the survey. Overall, however, the time-of-day profiles generated by the Replica simulation are remarkably consistent with those observed in the rMove survey.
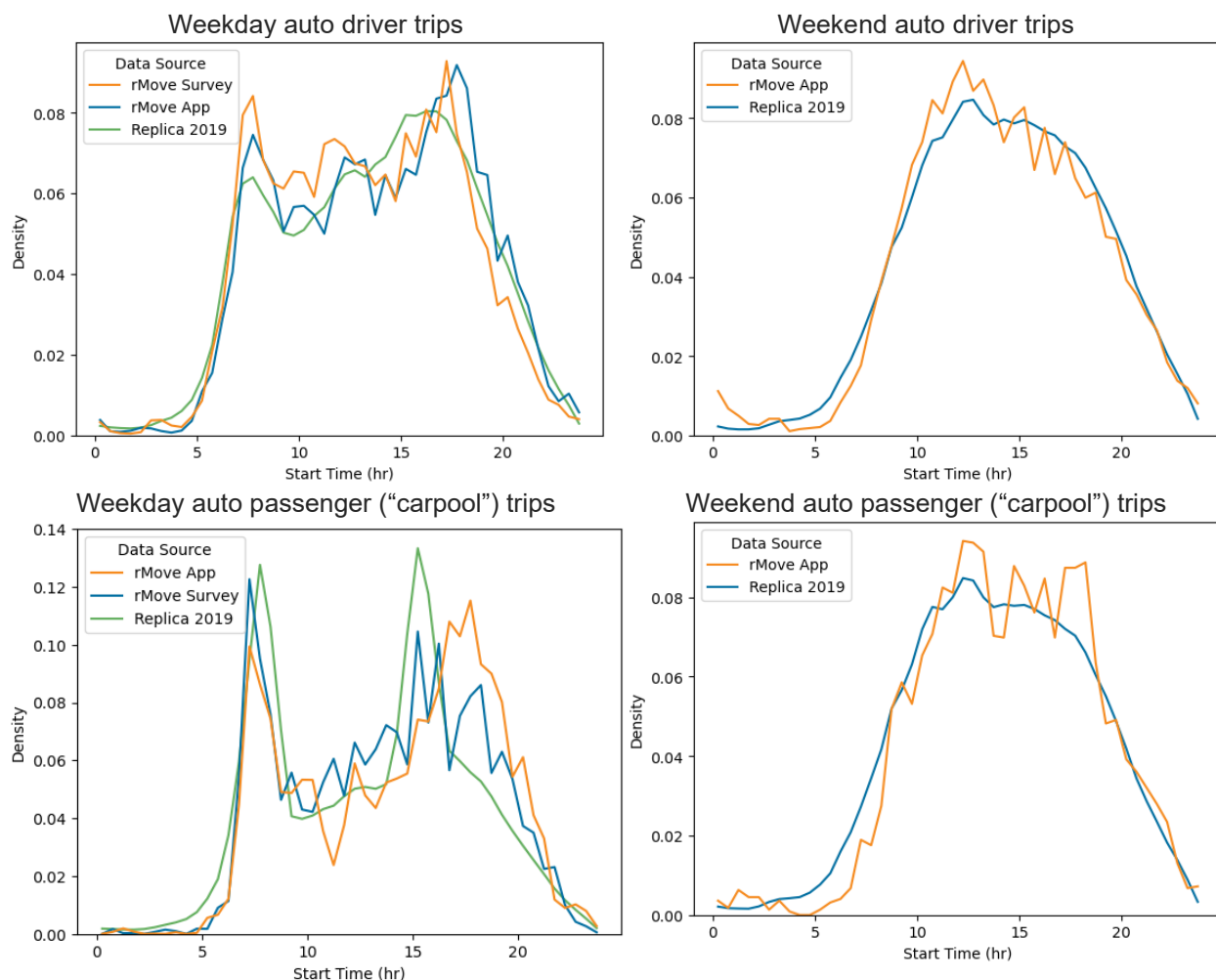
FIGURE 18. TRIP TIME-OF-DAY DISTRIBUTIONS, BY DATA SOURCE AND DAY TYPE FOR DRIVE ALONE AND CARPOOL TRIPS.

Trip departure time-of-day profiles for walk and bike trips are more difficult to compare between the two datasets, in large part due to the relatively small sample of such trips in the rMove data. However, focusing on the Replica data alone (green lines in Figure 19), there seems to be strong evidence of peaking in the Replica data during weekdays, suggesting that many walk and bike trips in the Replica data are associated with the work commute. Further, Replica documentation states that it does not model purely recreational "loop" trips, like recreational cycling or dog walking, which would be present in the rMove data. Given the limited sample sizes of these trips in the rMove data, it may be advisable to identify other ground truth sources, such as pedestrian and bicycle count data, to perform further validation. However, there are indications that Replica walk and bike trips are biased towards longer trips associated with the work commute and under-represent shorter distance, shorter duration trips (e.g., dog walking) as well as longer-distance recreational trips such as walking or biking for exercise.
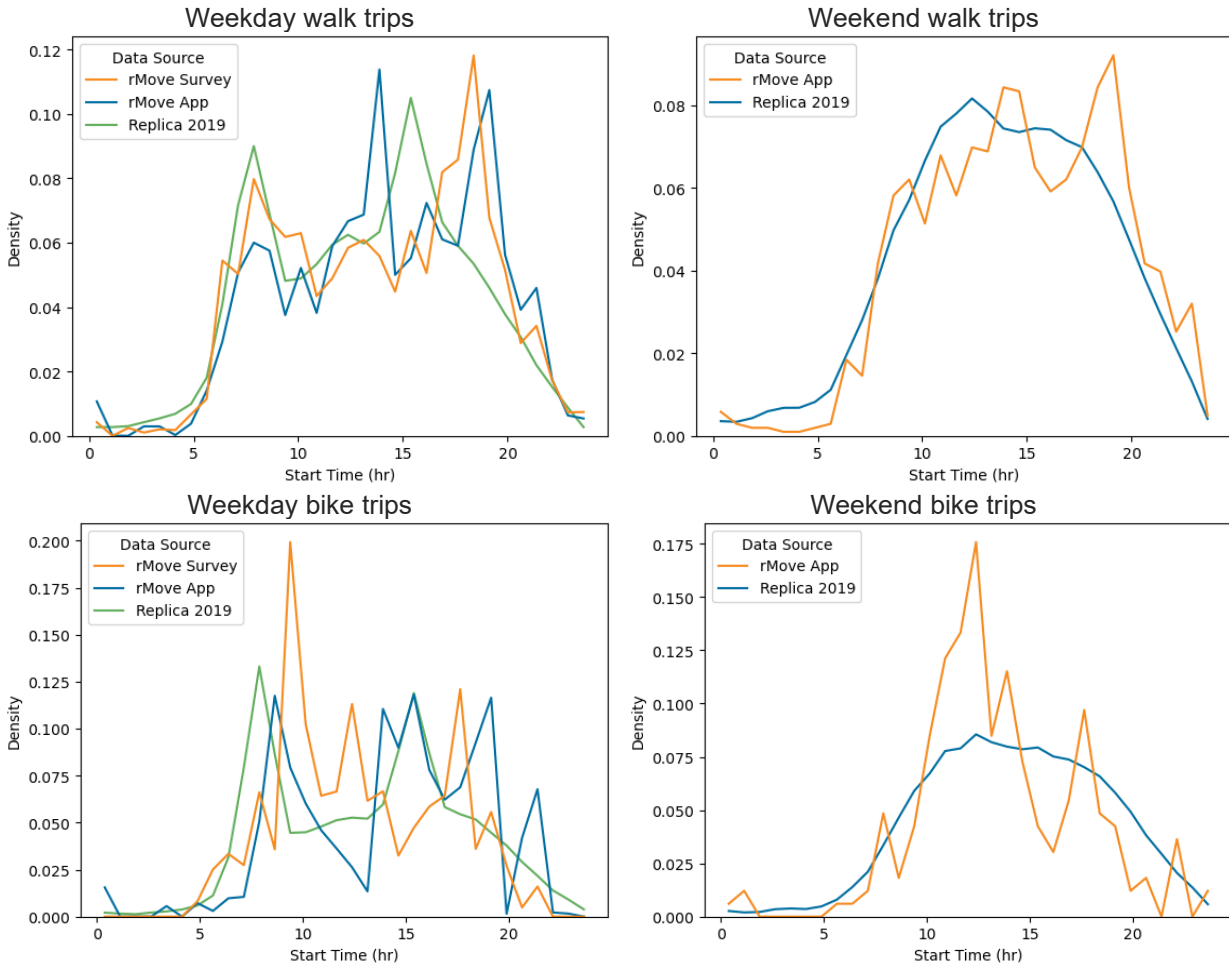
FIGURE 19. TRIP TIME-OF-DAY DISTRIBUTIONS, BY DATA SOURCE AND DAY TYPE FOR WALK AND BIKE TRIPS.

Finally, trip mode distributions are generally consistent between the two datasets (Table 8). As mentioned previously, Replica tends to over-estimate carpool trips and under-estimate auto driver trips, resulting in a higher-than-expected auto occupancy rate. Interestingly, Replica is close to rMove in walk and bike trip mode shares despite not modeling recreational walking and biking trips.

TABLE 8. TRIP MODE DISTRIBUTION, BY DATA SOURCE, MODE, AND DAY TYPE.

	Replica		rMove app		rMove diary	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Auto driver	62.9%	64.8%	66.4%	54.2%	72.0%	-
Carpool	25.8%	25.0%	20.7%	31.2%	12.4%	-
Walk	8.0%	7.3%	9.0%	9.7%	10.7%	-
Bike	1.2%	1.3%	1.5%	1.3%	0.0%	-
Transit	0.6%	0.3%	0.4%	0.4%	1.0%	
Other	1.3%	1.5%	2.3%	2.9%	2.6%	

1.5 CONCLUSIONS

The analysis presented in this memo focuses on the comparison of Replica's Fall 2019 simulated travel by D7 residents to rMove survey data for spring of that year. At a high-level, this comparison found that trip rates in the Replica data are low whereas trip distances tend to be higher than in the survey data. The closest alignment between the two datasets was found for work trips made in private automobiles. In general, the further from this "typical" trip the greater differences between Replica and the rMove survey. This may be due to richer ground truth data sources for work trips (e.g., Census journey to work data) and driving trips (e.g., count data).

The strongest aspects of the Replica simulation seem to be its ability to model trip departure time-of-day patterns and aggregate mode choice, and comparison of these aspects of the Replica data with rMove survey data provide high confidence in the fidelity of the simulation. This comparison also revealed several dimensions where there is lower confidence in the accuracy of Replica's simulation. First, destination choice for non-mandatory trips seems to have a bias towards longer-distance trips. This is true across all travel modes but seems particularly true for walking and biking trips. Second, Replica seems to be biased towards carpool trips, resulting in a high-than-expected auto occupancy rate implied by the Replica simulation. Finally, there are several instances where Replica diverges from common understanding of foundational travel behavior principles: trip rates do not seem to vary with household income, and zero-vehicle households generate trips at a similar rate as households with vehicles. While the latter may be expected in regions with dense transit networks and robust sidewalk and bicycle networks, zero-vehicle households in the Tampa region likely face transportation disadvantages that are not well-captured in the Replica simulation.