TRAVEL BEHAVIOR INVENTORY 2018-2024 SUMMARY REPORT

December 2024

Prepared for Metropolitan Council by RSG, Inc. (rsginc.com)

Report Title:

Travel Behavior Inventory 2018-2024 Summary Report

Report Prepared by:

RSG, Inc. (rsginc.com)

Report Prepared for:

Metropolitan Council

For additional information regarding this report, or for questions about permissions or use of findings contained therein, please contact:

Metropolitan Council
390 Robert Street North
St. Paul, MN 55101
(651) 602-1000

CONTENTS

LIST OF FIGURES

LIST OF TABLES

FOREWORD

This report contains a high-level, comprehensive overview of the methodology and results from the 2018-2024 Travel Behavior Inventory program.

For additional information regarding this report, or for questions about permissions or use of findings contained therein, please contact:

Metropolitan Council
390 Robert Street North
St. Paul, MN 55101
(651) 602-1000
https://metrocouncil.org/

EXECUTIVE SUMMARY

The Travel Behavior Inventory is a study of household demographics, daily travel activities, and typical transportation patterns throughout the greater Twin Cities region.

From 1949 to 2010, the study was conducted every ten years. In 2018, the Metropolitan Council (Met Council) transitioned to a recurrent, every-other-year study program. The recurrent program collected data for approximately 12 months of data collection every other year (Table 1). This recurrent study introduced efficiencies and operational improvements enabling the Metropolitan Council to collect more current, detailed data for use in their travel models.

This report synthesizes data from the first six years of the recurrent program (2018-2024), encompassing three waves of survey data collection.

More than 37,000 people, comprising over 19,000 households, participated in the 2018-2024 TBI (Table 1).

TABLE 1. SAMPLE OVERVIEW BY SURVEY YEAR (UNWEIGHTED COUNTS)

SURVEY YEAR	FIRST TRAVEL DATE	LAST TRAVEL DATE	HOUSEHOLDS	PERSONS	DAYS	TRIPS
2019	2018-10-05	2019-09-29	7,516	15,434	79,556	335,284
2021	2021-06-22	2022-02-05	7,907	15,031	49,567	180,242
2023	2023-01-12	2024-01-16	3,749	7,280	28,838	108,413
		TOTAL	**19,172**	**37,745**	**157,961**	**623,939**

1.1 STUDY GOALS

Key objectives of the Travel Behavior Inventory include the following:

- Collecting core information on household demographics, typical travel behavior, and regional transportation patterns to support the Met Council's transportation modeling and planning needs.
- Capturing information about new transportation modes and behaviors to keep pace with rapid changes in the transportation industry.

- Employing a thorough, multi-pronged approach to reach a representative sample of the population in the final dataset.
- Leveraging new technologies and methods to reach a more complete, detailed, high-quality dataset.

1.2 METHODOLOGICAL HIGHLIGHTS

This section describes three methods employed by the Study to collect accurate and representative travel behavior data: smartphone-based data collection, online travel diaries, and calling into the survey call center. All three of these methods represented methodological improvements over previous iterations of the Travel Behavior Inventory.

Smartphone-based travel diaries improved data quality and quantity.

Households with smartphones were required (2019) and/or encouraged with higher survey incentives (2021, 2023) to complete their travel diaries using the rMove™ smartphone app for up to seven consecutive days. The number and share of household who completed their travel diaries using rMove™ is shown in Table 2. Households without smartphones or households who were not willing to participate using rMove™ participated by completing their travel diary online (rMove™ for Web) or by calling into the survey call center. These households reported travel for one day (Tuesday, Wednesday, or Thursday).

Compared to diaries completed online or by telephone interview that rely on recall, smartphone-based travel diary collection offered significant benefits for data quality and quantity (e.g., detailed trip paths, and lower degrees of under-reporting).

TABLE 2. PERCENT OF HOUSEHOLDS USING RMOVE™ SMARTPHONE TRAVEL DIARIES BY SURVEY YEAR

SURVEY YEAR	RMOVE HOUSEHOLDS	PERCENT OF ALL HOUSEHOLDS
2019	5,026	66.9%
2021	3,242	41.0%
2023	1,902	50.7%

Multiple days of data collection captured atypical travel behaviors.

While multi-day data collection increases the volume of data collected per household, it also helps capture a wider variety of behaviors.

Figure 1 and Figure 2 below show the variation of trip mode and purpose for rMove™ respondents by the number of reported days of travel. Increasing the number of days completed increases the number of non-walk, non-vehicle trip modes included in the dataset (Figure 1) and greatly increases the number of trip purposes other than home, school, or work included in the dataset (Figure 2).

FIGURE 1. NUMBER OF DISTINCT MODES CAPTURED BY INCREASING THE NUMBER OF COMPLETED DAYS.

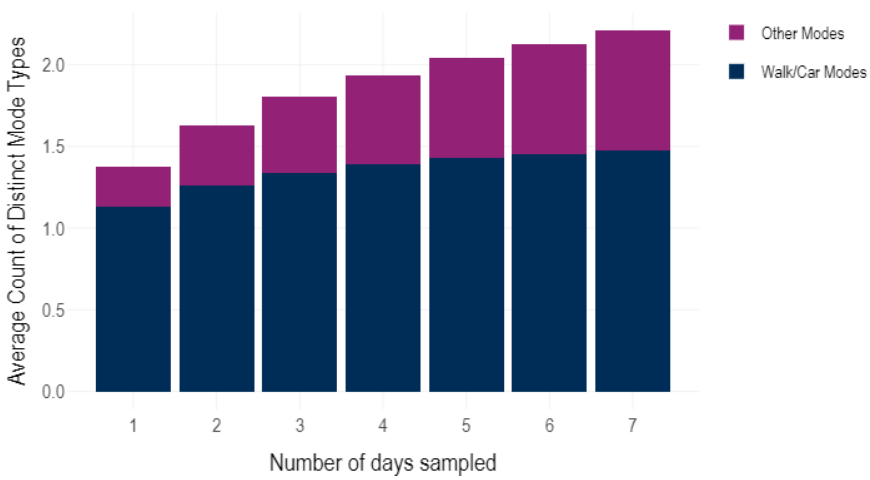
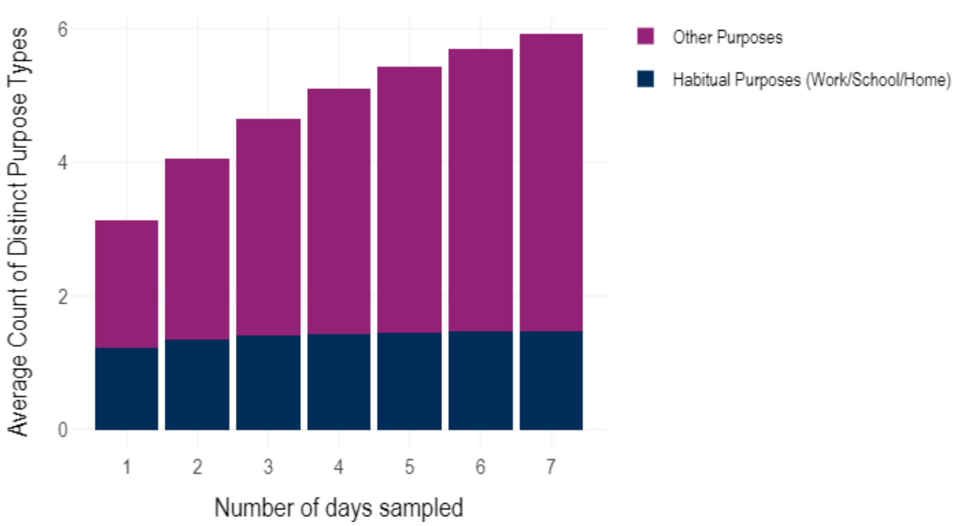


FIGURE 2. NUMBER OF DISTINCT PURPOSES CAPTURED BY INCREASING THE NUMBER OF COMPLETED DAYS



Address-based sample representativeness improved over time.

Most of the sample recruitment was accomplished through address-based sampling (ABS), a type of probability sampling, with a focus on reaching segment-level recruitment goals. The sample plan was reviewed and refined on an iterative basis, both within and between survey waves, to improve response from typically hard-to-survey groups. (For an example, see the 2021 Survey Management Plan).

Major efforts were made in the initial pilot (2018) and into the 2019 wave of data collection to recruit a large and representative sample. The team broadly disseminated information about the survey online, tested invitation materials and incentive amounts, and even tried door-to-door follow-up with survey non-respondents (See 2019 Survey Methodology Memorandum). Although the 2019 survey exceeded its sample size targets, the overall demographic composition of the survey was more biased than desired, especially related to race: only 2.5% of the people surveyed in 2019 identified as Black or African American, compared to 8.1% in the 2015-2019 American Community Survey (ACS; see Table 3).

TABLE 3. PARTICIPANT RACE BY SURVEY YEAR (UNWEIGHTED).

RACE	2019	2021	2023	2015-2019 ACS
American Indian or Alaska Native	0.4%	0.7%	0.7%	0.5%
Asian	3.3%	4.9%	5.1%	6.2%
Black or African American	2.5%	5.2%	4.2%	8.1%
Native Hawaiian or other Pacific Islander	<0.1%	<0.1%	0.2%	<0.1%
Two or more races	2.7%	3.1%	3.1%	3.2%
White	89.9%	84.8%	84.9%	77.7%
Other race, ethnicity, or origin	1.2%	1.2%	1.8%	4.2%
TOTAL	**100%**	**100%**	**100%**	**100%**

In the 2021 wave of data collection, the study team made several improvements to survey design, incentives, outreach, and supplemental sampling to recruit a more representative sample - with a special focus on race, ethnicity, and income (See 2021 Survey Methodology Memorandum). The percentage of the sample identifying as White decreased from 2019 to

2021, bringing the total sample composition more in line with Census estimates (Table 3), but the team felt there was more room for improvement.

Differential incentives and outreach efforts in 2023, however, did not yield further improvements to sample representativeness over those afforded in 2021: the percentage of respondents identifying as Black or African American remained at approximately half of Census estimates (Table 3).

Taken as a whole, these results point to a need for future waves of the Travel Behavior Inventory to implement further modifications to boost sample representativeness. Alternative methods of recruitment beyond mail, invitation redesign, panel sampling, and strategic oversampling are all possible avenues of exploration.

1.3 KEY SURVEY FINDINGS

Number of trips

In 2023, the average resident made 3.48 trips on a typical weekday. This trip rate was below that of 2019 (4.17 trips per day) but represented a rebound from the 2.85 trips per day in 2021 (Table 4).

TABLE 4. UNWEIGHTED AND WEIGHTED TRIP RATE BY YEAR.

	UNWEIGHTED			WEIGHTED		
SURVEY YEAR	TRIPS	DAYS	TRIP RATE	TRIPS	DAYS	TRIP RATE
2019	329,021	79,556	4.18 ± 0.02	15,348,964	3,682,918	4.17 ± 0.04
2021	176,093	49,567	3.48 ± 0.02	10,670,349	3,748,414	2.85 ± 0.04
2023	104,611	28,838	3.68 ± 0.03	13,038,925	3,750,006	3.48 ± 0.05

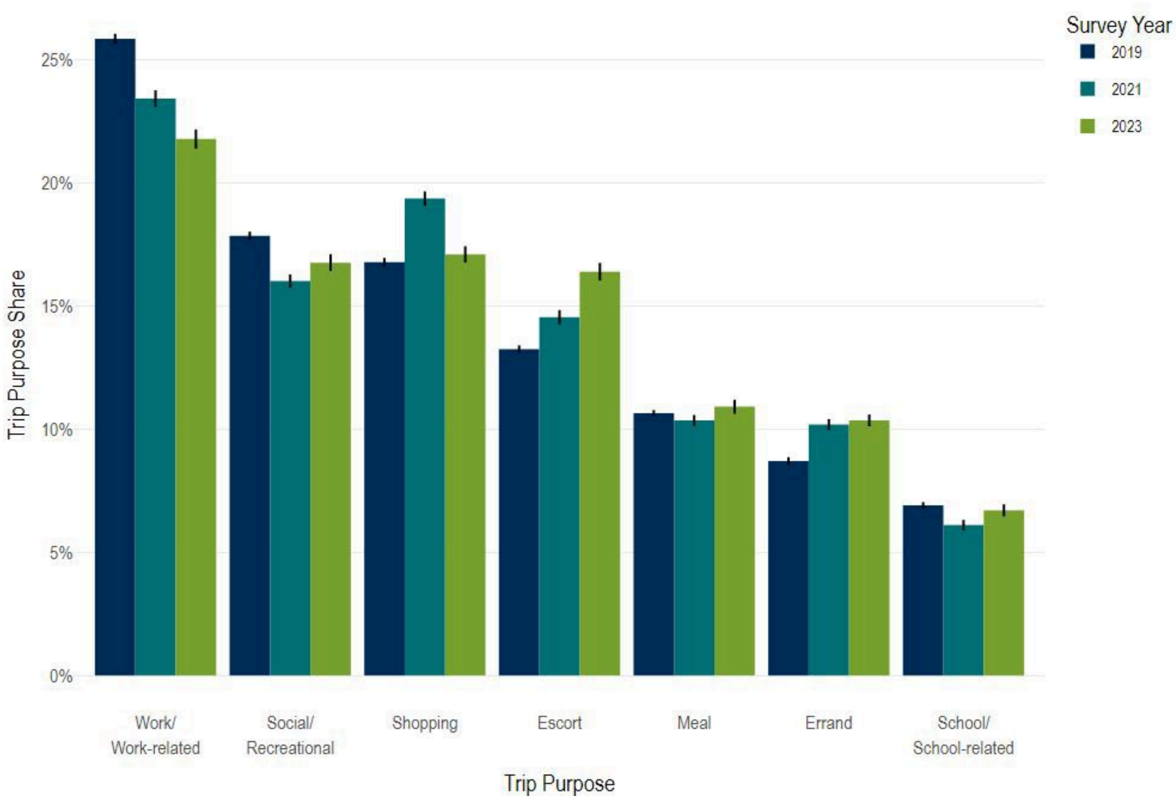
Error margins are +/- 1 standard error of the mean.

Trip purposes

Despite the extraordinary effects of the COVID-19 pandemic on daily life and work culture, the share of trips made by purpose stayed mostly consistent from 2019 to 2023 (Figure 3). Across all three years, the most common trip purpose was work or work-related, followed by social/recreational trips and shopping trips.

From 2019 to 2021, the share of trips made for work and work-related purposes declined from 26% of trips to 23% of trips, then declined again slightly to 22% of trips in 2023. Conversely, the share of trips made for shopping and errands increased from 2019 to 2021, then declined from 2021 to 2023.

FIGURE 3. SHARE OF TRIPS BY PURPOSE AND SURVEY YEAR (WEIGHTED).



Comparisons of the share of trip purpose by year must be made carefully, because overall trip rates declined during the COVID-19 pandemic and did not fully recover (Table 4). An alternative measure, the participation rate, can be used instead: this measure represents the share of people who make any given type of trip on a typical day (Figure 4), and as such accounts for both changes in the types of trips people make and the changes in the amount of travel people make.

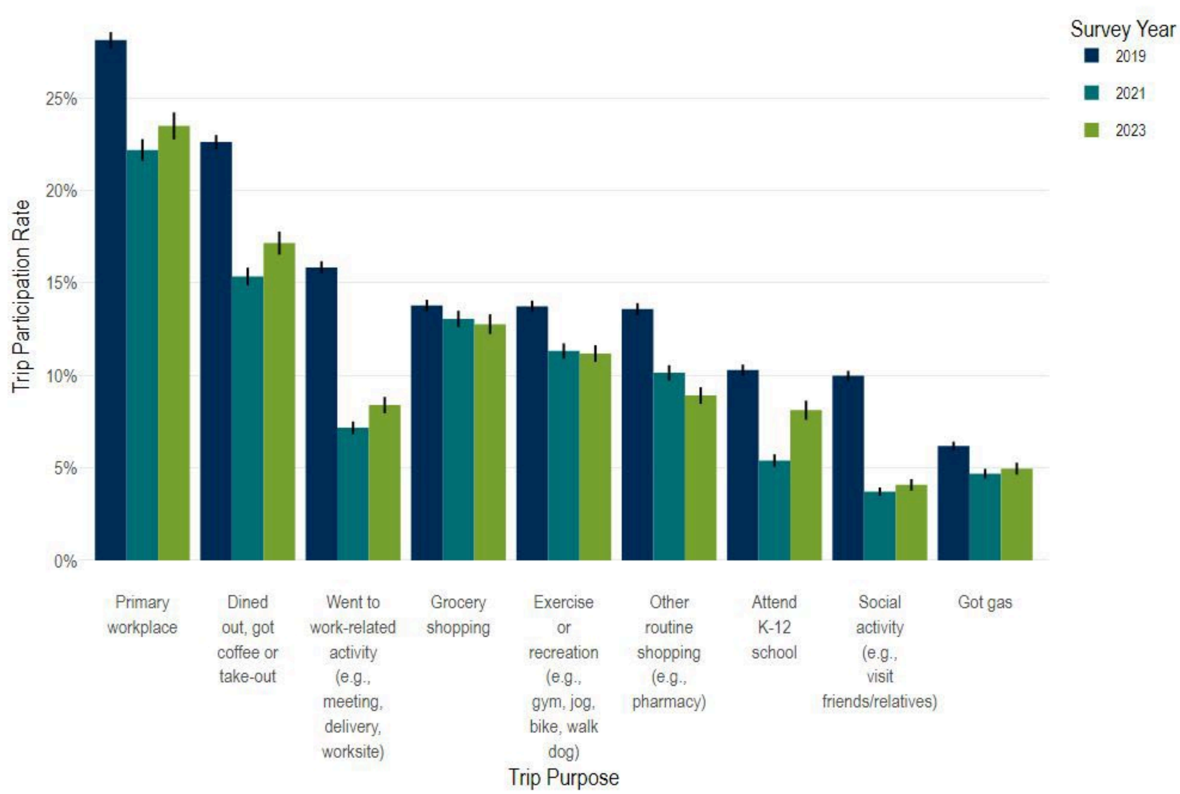
The trends in Figure 4 suggest a number of underlying processes at work.

The effects of telework on travel behavior are easy to spot: the share of people making work and work-related trips declined from 2019 to 2021, then slightly rebounded from 2021 to 2023 (Figure 4). Work-related trips (to meetings, deliveries, worksites) took a greater hit than trips to and from primary workplaces. From 2019 to 2023, the share of people making trips to work-related locations fell from 16% to 8%. Teleconferencing could be replacing trips that used to be made in person.

The fingerprint of e-commerce is also visible in the data. For example, the percent of people making a routine shopping trip on a typical day fell from 14% in 2019 to 10% in 2021, and then to 9% in 2023 (Figure 4).

Finally, the lingering effects of the pandemic on social life are also apparent. The share of people making a social visit trip more than halved from 2019 to 2021, then rebounded only slightly in 2023 (Figure 4) . Similarly, the share of people making trips to exercise or walk outside declined from 2019 to 2021.

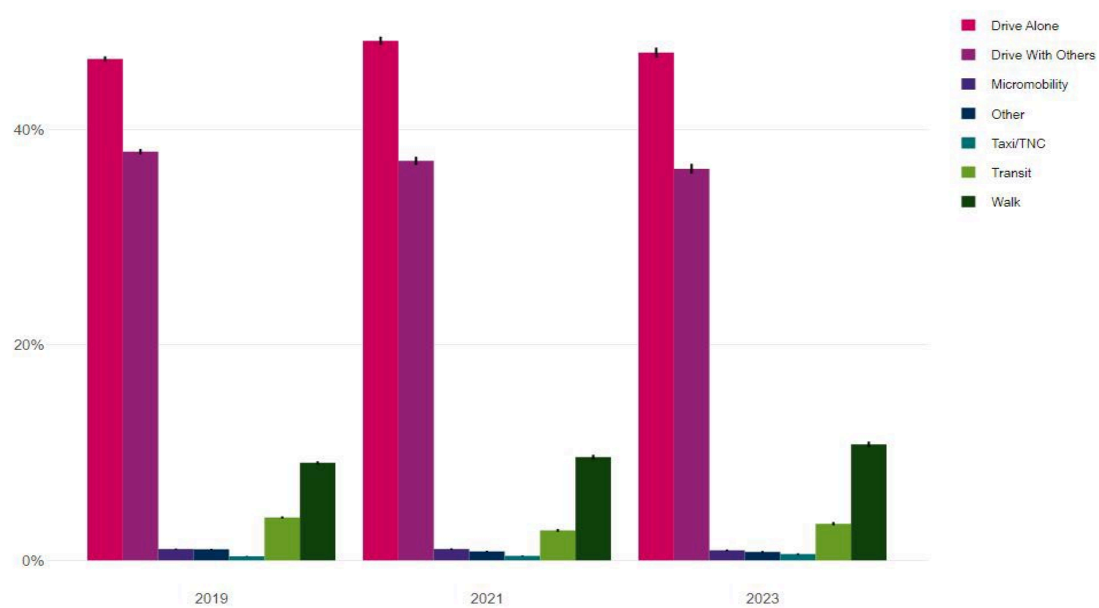
FIGURE 4. TRIP PURPOSE PARTICIPATION RATE BY YEAR (SELECTED PURPOSES, DETAILED CATEGORIES; WEIGHTED).



Travel choices

Driving remains the predominant mode of travel in the region, representing 83% of trips in 2023 (Figure 5). Apart from a decrease in transit mode share from 2019, the share of trips made by each mode was stable throughout the entire six-year study period.

FIGURE 5. TRIP MODE TYPE BY SURVEY YEAR (WEIGHTED).



1.4 EXPLORE THE DATA

To facilitate data exploration and analysis, the study team:

- Created a cross-wave dataset and codebook, available for download by the public (Met Council: link when dataset is published).
- Developed an interactive app for data exploration and crosstabs (link).
- Developed an open-source R package to support analysis of household travel survey data (link).

These resources are open-source and freely available. For support in accessing or using the data, or any elements of this report, please contact Met Council Public Information.

2.0 INTRODUCTION

The Travel Behavior Inventory is a study of household demographics, daily travel activities, and typical transportation patterns throughout the greater Twin Cities region conducted approximately every 10 years since 1949. The Met Council conducted the most recent study in 2010, with a household travel diary survey of approximately 12,000 households and supplemental surveys including a transit on-board survey, airport survey, and visitor surveys. In 2018, the project team developed a recurrent study program to collect study data more frequently, which is conducted by RSG on behalf of the Met Council and its partners the Minnesota Department of Transportation (MnDOT) and the Wisconsin Department of Transportation (WisDOT) This recurrent study program introduced administrative efficiencies and operational improvements for data collection enabling the Met Council to collect more current, detailed data for use in travel models and planning analysis.

2.1 STUDY OBJECTIVES

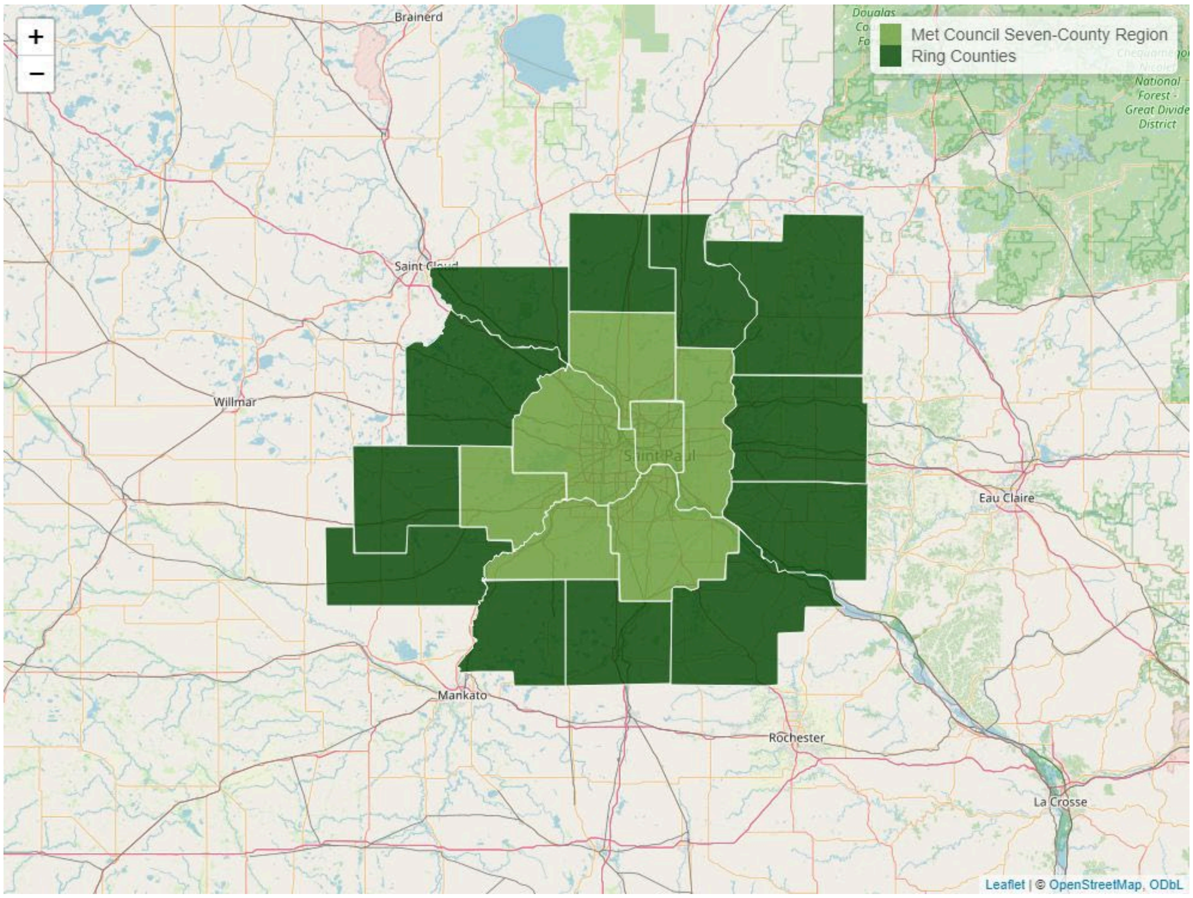
Key objectives of the Travel Behavior Inventory include the following:

- Collecting core information on household demographics, typical travel behavior, and regional transportation patterns to support the Met Council's transportation modeling and planning needs.
- Capturing information about new transportation modes and behaviors to keep pace with rapid changes in the transportation industry.
- Employing a thorough, multi-pronged approach to obtain a representative sample of the population in the final dataset.
- Leveraging new technologies and methods to obtain a more complete, detailed, high-quality dataset.

2.2 STUDY AREA

The study used a sample frame consistent with previous Travel Behavior Inventories: the seven-county Twin Cities metro area (which comprises the Met Council planning area), nine adjoining ring counties in Minnesota, and three counties in Wisconsin. The study area holds approximately 1.4 million households (2019 ACS).

FIGURE 6. STUDY AREA



2.3 STUDY TIMELINE

The scope of work for this project included the design and administration of a mixed-mode data collection approach, with smartphone-based retrieval of up to seven days of travel data as the primary methodology, supplemented with traditional online and call center-based data collection. The Travel Behavior Inventory planned for approximately 12 months of data collection every other year. Data collection in 2021 was delayed, and then the timeline compressed, due to the COVID-19 pandemic.

- 2019 Survey ("Wave 1") collected data from 2018-10-05 to 2019-09-29.
- 2021 Survey ("Wave 2") collected data from 2021-06-22 to 2022-02-05.
- 2023 Survey ("Wave 3") collected data from 2023-01-12 to 2024-01-16.

3.0 SURVEY SAMPLING

This section provides the sampling plan methodology for the three waves of the Travel Behavior Inventory. Below is a brief overview of the goals, methods and outcomes of each survey wave.

3.1 OVERVIEW BY WAVE

2018 pre-test

Before sending invitations for the first wave of the main study (2019), a pre-test was conducted to test survey administration processes, designs, instruments (rMove™ for Web and Smartphone), and resulting data to ensure that all platforms were optimized for the main survey effort. The pre-test tested two different study designs, "traditional" rMove™ and rMove™ "All-in-One" (AIO). The traditional design had all respondents recruit online, and get assigned to rMove™ if eligible, while AIO had respondents complete both the recruitment survey and travel diary on rMove™. During the pre-test, 50% of invited households were instructed to recruit using the traditional design and 50% of invited households were instructed to recruit using the AIO study design.

After completing the pre-test the consultant team and the project review team analyzed the study methodology and the pre-test results. As shown in Table 5, these results indicated that the AIO study design yielded a lower recruit rate and a lower response rate than the traditional study design.

TABLE 5. PRE-TEST PARTICIPATION METRICS BY STUDY DESIGN.

RECRUITMENT DESIGN	INVITED HHS*	RECRUITED HHS	COMPLETED HHS	RECRUIT RATE**	CONVERSION RATE**	RESPONSE RATE**
Traditional recruitment	10,250	423	227	4.10%	54%	2.20%
All-in-one recruitment	10,250	374	180	3.70%	48%	1.80%
Total	20,500	797	407	3.90%	51%	2.00%

*HHS = Households.

**Recruited households are those who complete the initial signup survey. Conversion rate is the percentage of recruited households who complete the travel diary.

In addition to testing effectiveness of study designs, the pre-test provided insight on response rates for hard-to-reach populations within the study region. A second finding from the pre-test showed that younger and lower-income households completed at a higher rate in the All-in-One study design. These populations are typically considered hard-to-reach groups in household travel surveys, so this finding indicated a promising potential use case for the AIO study design.

2019, “wave 1”

All data collected for the 2019 survey were collected through address-based sampling methods (ABS).

Based upon the pre-test findings, the project review team decided to split the sample of the main study such that 85% of participants were assigned to the traditional study design and 15% of participants were assigned to the AIO study design. Typically hard-to-survey populations received AIO invitations.

While representation in the sample was consistent with that observed in other household travel surveys, the unweighted response from hard-to-survey households was still lower than desired. The results of 2019 data collection resulted in recommendations later implemented in 2021, including:

- Increased differential incentives for households typically hard to survey or recruit.
- Increased the proportion of hard to survey households invited to participate through ABS.
- Revised the survey methodology to allow households in which all adults own a qualifying smartphone to choose whether they prefer to report travel using their smartphone, online, or through a call center.
- Implemented non-probability (convenience) sampling methods to increase the participation from hard-to-survey groups.

Read more in the Survey Methodology Report for 2019.

2021, “wave 2”

The 2021 survey faced several key challenges including the COVID-19 pandemic, declining trust in government, and postal delays. Despite these historic challenges, the 2021 survey surpassed the overall survey participation target of 7,500 households.

The 2021 survey departed from the 2019 survey design in utilizing an opt-in approach where households where all adults had a smartphone could choose to report their travel by smartphone, online, or through the call center, rather than being assigned to participate by smartphone based on smartphone ownership. This resulted in 41% of participants completing the survey by smartphone, 53% online, and 6% by call center[1].

The 2021 survey also included a substantial set of data collected using supplemental non-probability methods to increase the proportion of hard-to-survey households in the final dataset. ABS efforts yielded 90% of the complete households, while supplemental sampling efforts yielded 10% of the complete households. Non-probability sampling methods included outreach through community-based organizations and leveraging Metro Transit's Transit Assistance Program (TAP) email and text lists.

RSG and the Met Council worked closely with a public outreach firm, NewPublica to coordinate an effort to invite community-based organization (CBO) members to participate in the Travel Behavior Inventory with a focus on CBOs that are primarily composed of Black, Indigenous, people of color (BIPOC) community members that were underrepresented in the 2019 survey. NewPublica coordinated with CBOs to determine the best means to invite CBO members and facilitate members invitation to participate in the survey. The CBO sample were offered higher differential incentives.

Working closely with Metro Transit, RSG invited Metro Transit customers to participate in the Travel Behavior Inventory. The goal of this method was to leverage a reasonably low-cost method to try to improve response among certified low-income populations who are known to be hard-to-survey and for whom there is some overlap with underrepresented races and ethnicities. Of specific interest was the population of riders who have been certified as low-income, receive a form of transit subsidy, and for whom an email address or mobile phone number is available. For additional details see the Survey Methodology Report for 2021 Appendix.

2023, "wave 3"

As in 2021, the 2023 survey employed an opt-in approach to recruiting households to the smartphone-based travel diary. This resulted in 48% of participants completing the survey by smartphone, 39% online, and 12% by call center.

The 2023 sample plan aimed to improve recruitment of demographic groups that were underrepresented in 2019 and 2021. RSG implemented a combination of ABS and non-probability sample methods (community-based organization outreach). Due to the lower quality of the CBO outreach survey data, these records were not weighted in the final dataset. For additional details see the Outreach Report for 2023.

[1] The 2019 survey used an assignment approach resulting in a higher share of smartphone completes (68%) than seen in 2021 data collection. Analysis of the 2019 survey determined that a higher proportion of hard-to-survey households that recruited online (or through the call center) did not complete the survey if they were required to report travel by smartphone. In addition, the assignment approach requires more mailings, which is more costly. For these reasons, an opt-in smartphone approach was used in 2021.

3.2 SAMPLING METHODS

Sampling Goals

The 2019 Travel Behavior Inventory aimed to sample 7,500 households, which equates to a 0.20% sample rate according to data from the 2016 ACS 5-year estimates. Beyond achieving the overall sample target, the survey also aimed to ensure that the sample was representative across key demographics and behaviors, as discussed below.

Sampling Frame

The Travel Behavior Inventory region is comprised of the seven-county Twin Cities metropolitan area, nine adjoining ring counties in Minnesota, and three bordering counties in Wisconsin. RSG used ABS to select a random sample of addresses from all residential addresses in the study area. Using this method, all households within each defined area have an equal chance of selection for the sample. The sampled addresses were purchased from Marketing Systems Group (MSG), which maintains the Computer Delivery Sequence file from the U.S.

When purchasing the addresses, RSG also purchased the estimated household income for the list of addresses. Typically, MSG provides an estimated income for about 85-90% of the total addresses at a cost of $0.01 per address. The estimated household income data was used to aid address selection.

Address-Based Sampling

Sample Segmentation

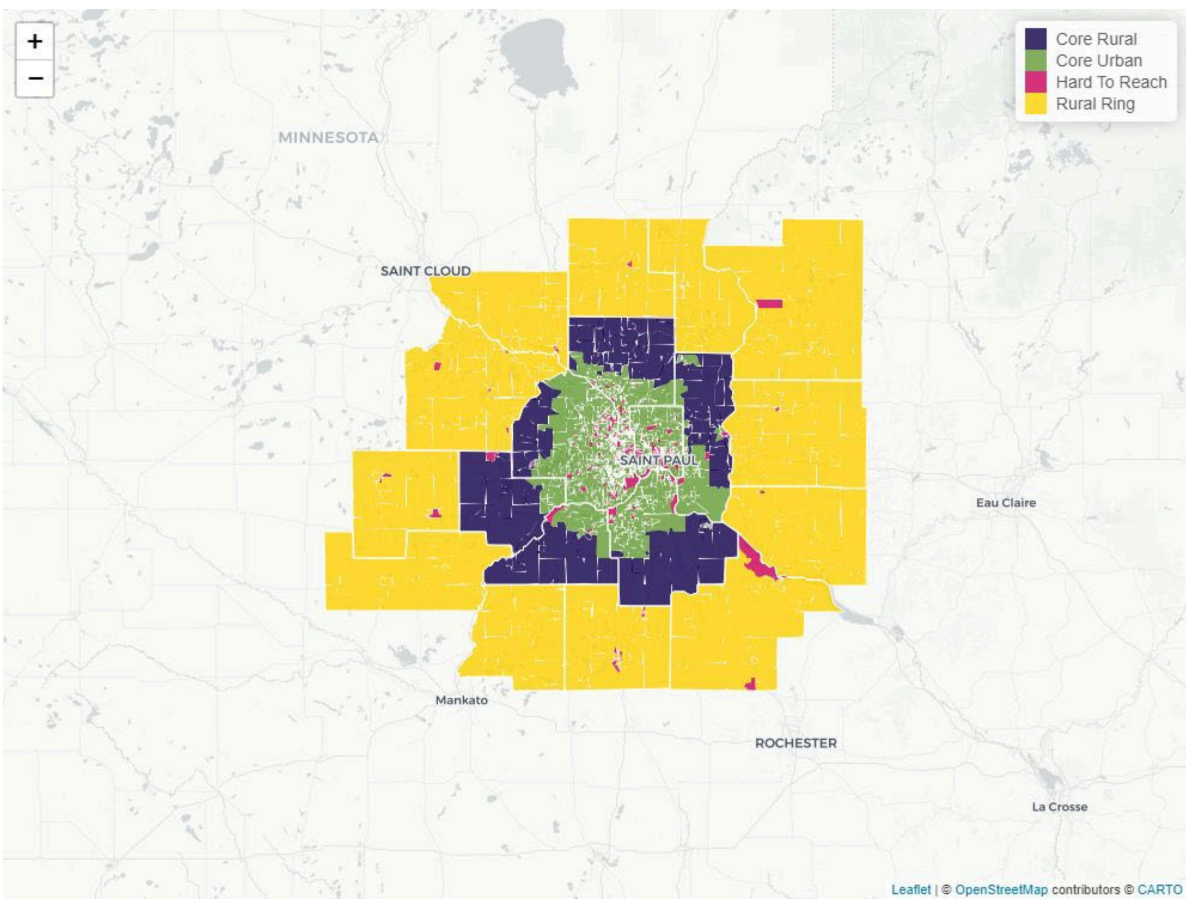
RSG stratified the sample using census Block Group (BG) data from the most recently available 2012–2016 American Community Survey 5-year estimates (ACS). The most detailed way to stratify the sample is to use census BGs, which are the smallest geography for which most census and ACS tables are publicly available. Each BG generally contains between 600 and 3,000 people. According to this ACS data, the study region contains 1.4 million households and 3.6 million persons. Group Quarters, excluded from the sample frame, are a relatively small segment of the population at 2%.

Sampling planning methodologies employed in the Travel Behavior Inventory mitigated non-response bias and other known biases to meet study targets.

For the 2019 survey, RSG coordinated with the Met Council to determine four key groups to sample. Based upon the Met Council's interest in proportional representation from low income, rural, and limited English-speaking households for the main study, RSG proposed the following mutually exclusive and collectively exhaustive sample segments (Figure 7):

1. **Core-Urban BGs:** Comprised of the BGs in the Twin Cities seven-county metropolitan area that do not qualify for the hard-to-reach segment below and which are designated as Urban in the Thrive MSP 2040 Community Designations.
2. **Core-Rural BGs:** Comprised of the BGs in the Twin Cities seven-county metropolitan area that do not qualify for the hard-to-reach segment below and which are designated as Rural in the Thrive MSP 2040 Community Designations.
3. **Rural Ring BGs:** Comprised of the BGs in the twelve ring counties surrounding the seven-county metropolitan area that do not qualify for the hard-to-reach segment below.
4. **Hard to reach BGs:** Comprised of BGs in the nineteen-county study region that were in the 90th percentile of BGs with the highest percentage of households with annual incomes below $25,000 and/or the 90th percentile of BGs with the highest percentage of limited English-speaking households.

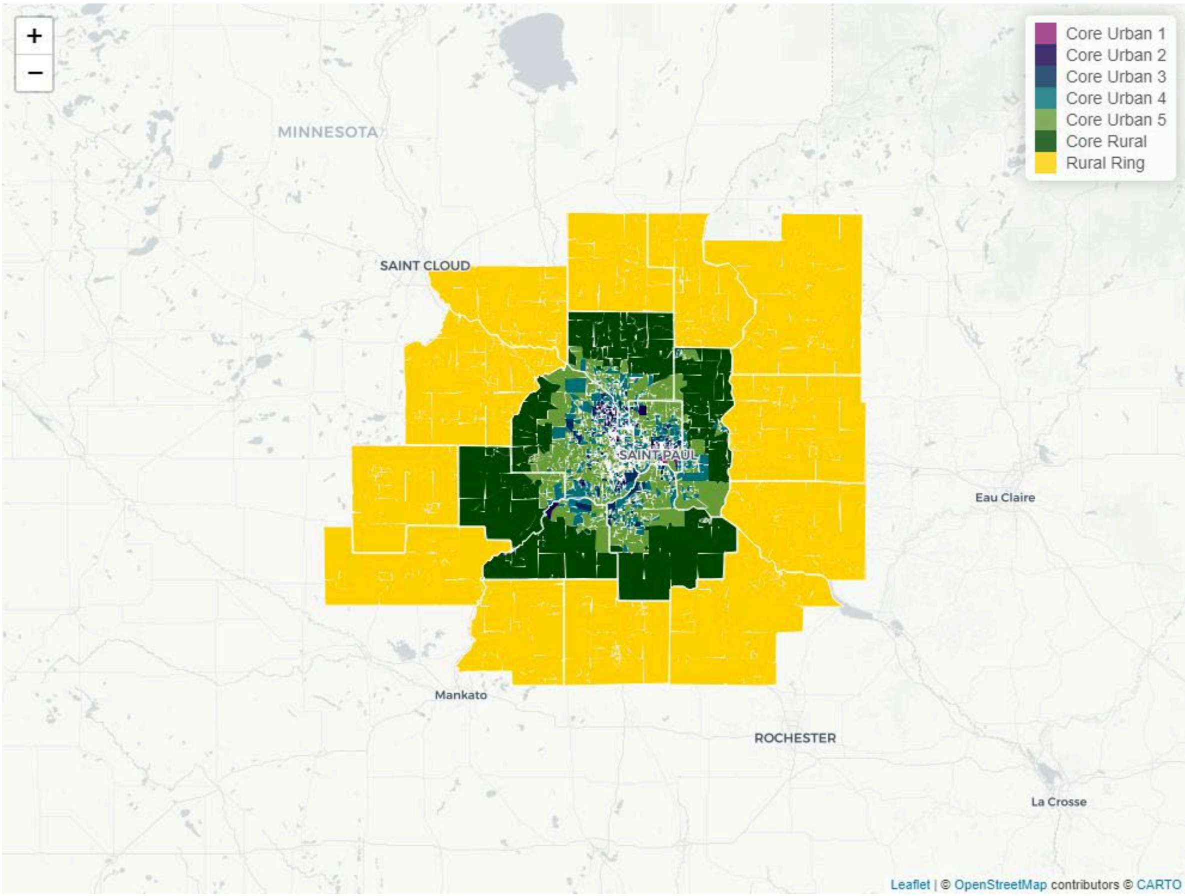
FIGURE 7. 2019 SAMPLE SEGMENTS.



Starting in 2021, RSG proposed a more specific focus on sampling residents who are Hispanic and/or Black, Indigenous, and people of color (BIPOC). Sample segments for the 2021 and 2023 surveys built upon the 2018-2019 Travel Behavior Inventory, but dropped the “Hard-to-Reach” segment in favor of sub-segmenting the Core Urban geography into five groups, making the final segments (Figure 8):

1. **Core-Urban BGs – Group 1:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Urban in the Thrive MSP 2040 Community Designations and whose population is at least 80% Hispanic and/or BIPOC.
2. **Core-Urban BGs – Group 2:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Urban in the Thrive MSP 2040 Community Designations and whose population is 60%-80% Hispanic and/or BIPOC.
3. **Core-Urban BGs – Group 3:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Urban in the Thrive MSP 2040 Community Designations and whose population is 40%-60% Hispanic and/or BIPOC.
4. **Core-Urban BGs – Group 4:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Urban in the Thrive MSP 2040 Community Designations and whose population is 20%-40% Hispanic and/or BIPOC.
5. **Core-Urban BGS – Group 5:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Urban in the Thrive MSP 2040 Community Designations and whose population is less than 20% Hispanic and/or BIPOC.
6. **Core-Rural BGs:** Comprised of the BGs in the Twin Cities seven-county metropolitan area which are designated as Rural in the Thrive MSP 2040 Community Designations.
7. **Rural Ring BGs:** Comprised of the BGs in the twelve ring counties surrounding the seven-county metropolitan area.

FIGURE 8. 2021 AND 2023 SAMPLE SEGMENTS.



Oversampling

Compensatory oversampling is a sampling method where more invitations are sent to regions with lower estimated response rates. RSG typically uses compensatory oversampling so that the geographic distribution in the final sample is closer to population proportional to the study region. While oversampling may not result in the study matching weighted ACS demographics for race, ethnicity, and income for the region, it can increase the overall number of hard-to-survey households.

To increase representation from households with BIPOC and Hispanic members, RSG used targeted oversampling in addition to compensatory oversampling. RSG sent additional invitations (in addition to the invitations sent as part of compensatory oversampling) in core-urban block groups whose population is greater than 20% BIPOC and/or Hispanic (Table 6).

Targeted oversampling increased the number of completed households from these block groups and reduced the number of completed households from other block groups.

TABLE 6. OVERSAMPLING RATE BY STUDY SEGMENT (2021 AND 2023).

SAMPLE SEGMENT	PERCENT BIPOC AND/OR HISPANIC	OVERSAMPLING RATE
Core Urban Group 1	88%	300%
Core Urban Group 2	71%	200%
Core Urban Group 3	49%	150%
Core Urban Group 4	29%	125%
Core Urban Group 5	11%	--
Core Rural	8%	--
Rural Ring	8%	--
Study Region Total	24%	--

3.3 SUPPLEMENTAL SAMPLING

A variety of supplemental sampling methods were employed in 2021 and 2023 to boost response by households that were harder to recruit using Address-Based Sampling.

Recruitment Through Metro Transit rider email lists (2021)

Working closely with Metro Transit, RSG coordinated permissions, invitation copy, and conduct of inviting a sample of Metro Transit customers. Of specific interest was the population of riders who had been certified as low-income and receive a form of transit subsidy (the “TAP,” or Transit Assistance Program) and for whom an email address was available.

Metro Transit issued email invitations and reminders working in conjunction with RSG. The goal of this approach was to leverage a reasonably low-cost method to try to improve response among certified low-income populations who are known to be hard-to-survey and for whom there is anticipated some overlap with underrepresented races and ethnicities.

Invitations were distributed over the course of 6 weeks and went out to four invitation groups to spread the sample's survey participation across several weeks in the Fall. The team obtained 400 complete household responses from Metro Transit rider email lists.

Recruitment Through community-based organizations (2021)

Working closely together, New Publica, RSG, and the Met Council coordinated an effort to recruit participants through community-based organizations (CBOs) in Fall of 2021. These efforts did not yield participants in the amounts hoped for, however, primarily due to difficulties in reaching community members in the early years of the COVID-19 pandemic.

Recruitment Through community-based organizations (2023)

In 2023, RSG contracted SDK Communications to lead the equity cohort of the Travel Behavior Inventory as part of the Met Council's Transportation Policy Plan. SDK's specific charge was to obtain between 300 and 450 survey responses from the under-represented African American, African Immigrant and LatinX communities of the Twin Cities metro region. SDK collaborated with a cohort of partner organizations to achieve these numbers. Between our team's direct outreach and partner submissions, SDK recruited 437 participants.

Community partners engaged to administer surveys were:

- Latino Chamber of Commerce
- St. Paul Promise Neighborhood
- The Lift Garage
- Greater Mount Vernon Missionary Baptist Church
- ACER, Inc.
- Urban Strategies, Inc. (Heritage Park Neighborhood, Highway 55 in North MPLS)

In addition, SDK leveraged a history of relationships with affordable housing developers and managers to hold lunch events where people took the survey at an apartment building's community room.

Although the data did not meet the requirements for weighting, the unweighted data provided many anecdotal insights into the way that people who are typically not recruited using address-based sampling move around the region. These are detailed in SDK's final report.

4.0 SURVEY DESIGN

The Travel Behavior Inventory employed a mix of data collection methods, including smartphone, online, and telephone. The study design balanced the strengths of innovative technologies with pragmatic best practices derived from traditional market research.

The survey occurred in two parts. Part one (the “recruitment survey”) gathered data on each household’s demographic composition and typical travel behaviors. The recruitment survey was conducted online through the study web page, over the phone through the study call center or using the rMove™ app (starting in 2021).

Part two (the “travel diary”) gathered travel data for all related household members during a designated travel period. Households where all adults had smartphones were required (2019) or encouraged with incentives (2021, 2023) to download the rMove™ app and complete a seven-day travel diary.

If one or more adults in the household lacked a smartphone (all study years) or if the household opted out of using rMove™ (2021, 2023) the household completed a one-day travel diary. The one-day travel diary could be completed online or by calling the study call center. Call center operators used the same online travel diary instrument as survey participants. Non-app households were assigned a travel date of Tuesday, Wednesday, or Thursday to capture typical weekday travel behavior.

4.1 RECRUITMENT

Participants were recruited using a series of invitations mailed to their home addresses. Each mailing contained information about the study and a unique password for their household. See Section 5.1 for examples.

RSG leveraged contact information captured in the recruitment survey to provide customized, strategically limited, and well-timed reminders to recruited households by email, telephone, or within the rMove™ smartphone app to ensure nearly real-time completion of travel surveys.

4.2 LANGUAGE OPTIONS

Throughout the survey stages (e.g., recruitment and travel diary) and participation modes (e.g., app, online, or call center), several language options were made available to participants. The call center employed operators fluent in English and Spanish. Print materials included translated text with instructions on how to participate in six languages: English, Spanish, Hmong, Karen, Oromo, and Somali. The survey instruments were also available fully translated from English to Spanish, Hmong, Karen, Oromo, and Somali. In addition, the study website included a Google Translate bar to accommodate approximately 100 other languages. Table 7 shows the number of recruited households who took the survey in each offered language option in 2021 and 2023. Data was not available for the 2019 survey effort.

TABLE 7. RECRUITED HOUSEHOLDS BY LANGUAGE BY YEAR.

YEAR	ENGLISH	HMONG	KAREN	OROMO	SOMALI	SPANISH
2021	14,645	1	0	3	3	86
2023	5,660	1	1	1	1	95

Includes all recruited households (those who completed the signup survey). Some households did not go on to complete the travel diary and are not in the final dataset.

4.3 INCENTIVES

As compensation for their time and efforts, gift card incentives were offered to households for completing at least five complete travel days (smartphone diary participants) or a complete online travel diary (online participants). Otherwise incomplete households did not qualify to receive the incentive. Many surveys, including most household travel surveys and the National Household Travel Survey (NHTS), offer incentives because:

- Incentives increase response rates significantly, often doubling participation rates.
- Incentives help reduce participant bias.
- Historically low response rates in the United States and abroad mean that incentives are now often required to obtain a sufficient sample.
- By increasing study response rates, incentives help save the project money overall and encourage the most cost-effective use of public research funding.
- Incentives represent fair payment for the expertise participants provide on their own lived experience.

The study offered gift cards from Caribou Coffee, Target, and Amazon.com as incentives; participants could also waive their right to an incentive (in 2023, 9% of participants chose this option). These vendors were chosen for their availability within the travel study region, wide selection of products, widespread name recognition within the travel study region, and administrative ease of use to obtain and send gift cards to study participants. Table 8 shows the incentive offerings in each survey year, which varied based on diary platform and if the household was determined to be a “hard-to-reach” household. In 2019, all households were offered the same incentive (differential incentives were not offered).

TABLE 8. INCENTIVE OFFERING BY YEAR.

YEAR	RMOVE HARD-TO-REACH	RMOVE GENERAL	ONLINE HARD-TO-REACH	ONLINE GENERAL
2019	$15	$15	$10	$10
2021	$30	$20	$20	$10
2023	$30	$20	$20	$10
rMove™ incentives were per adult while online incentives were per household.				

In 2021 and 2023, “general” households (i.e., not hard-to-reach) that completed with the one-day online travel diary were offered a $10 e-gift card to their vendor of choice, or a physical gift card for households without email. Households that used rMove were offered e-gift cards of $20 per adult rMove participant to their vendor of choice. Thus, if a household had three rMove™ participants (each at least 18 years or older), then they received $45 in gift cards.

This incentive structure was designed to encourage larger households to participate using rMove™, which required more effort on behalf of each participant, and collected data for a greater number of days (seven travel days vs. one travel day).

TABLE 9. AVERAGE INCENTIVE PAYOUT BY DIARY PLATFORM IN 2023.

DIARY PLATFORM	PAYOUT PER HOUSEHOLD
rMove™	$34.26
Online	$10.81

4.4 QUESTIONNAIRE

Initial Design (2018)

In spring 2018, an initial data variable list was developed for the Travel Behavior Inventory Pre-test study. As part of this process, the Met Council obtained input from Minnesota Department of Transportation and Metro Transit. A two-fold approach was used: building upon the work performed in the 2014-2015 scoping project and building upon the data needs from prior smartphone-based household travel survey projects conducted by the consultant, RSG, in the U.S. and elsewhere.

The data needs for 2018-2019 data collection were finalized in August 2018 after the project team members reviewed pre-test results and identified necessary updates. The final data variable list summarized below aimed to satisfy core household travel survey data elements and select supplementary data elements that will benefit the Met Council's planning needs. Those key data elements included the following:

Core Household Travel Survey Data Elements

- Household-level (e.g. household composition, income, housing characteristics)
- Person-level (demographics, employment, and student status)
- Vehicle-level (year/make/model, ownership)
- Trip-level (activity, mode, costs, travel party composition, etc.)
- Location-level (GPS points, path data, etc.)

Supplemental Data Elements

- New travel mode usage including smartphone-app ride services (e.g., Uber, Lyft), peer-to-peer car rental, bikeshare (both docked and dock-less), and scooter share (electric scooters and mopeds)
- Trip replacement behavior such as package delivery (to participant's home, work, and package lockers), meal delivery, and service work (e.g., Electrician visit), telecommute hours, and online shopping behavior
- Interest in autonomous vehicle (AV) use
- Land-use questions requested by the Met Council's planning staff (e.g., monthly household rent)

Updates in 2021 and 2023

> Note
>
> This list of updates is not comprehensive. For a complete overview, consult the combined codebook or the questionnaires.

In 2021 and 2023, several updates were made to questionnaire language and content. Some highlights of these updates are provided below.

Race and ethnicity. In 2019, respondents were asked about their race and ethnicity in a single, multiple-response ("select-all-that-apply") question. In 2021 and 2023, race was asked separately from ethnicity. This allowed the survey to gather more detailed responses from those who identified as Hispanic. The survey furthermore asked follow-up questions for those who identified as African American or Asian/Asian American. These updates were made to align the Travel Behavior Inventory with other regional and national efforts, including the National Household Travel Survey.

Gender. In 2019 and 2021, "transgender" was listed as a response option for gender separate from "Male" and "Female." This way of soliciting gender identity conflate gender identity with transgender status; in other words, the identity of "woman" belongs to both cis- and transgender women. To be trans-inclusive, the response options for gender were updated in 2023 to read:

- Female/Woman/Trans woman/Girl
- Male/Man/Trans man/Boy
- A gender other than singularly male or female (e.g., non-binary, genderfluid, agender, culturally specific gender)
- Other
- Prefer not to answer

Despite concerns that the new trans-inclusive language might turn away some respondents, only five participants dropped out of the survey at the "gender" question screen. For comparison, the same number of participants dropped out when asked to verify which household vehicles used a toll transponder. By contrast, 182 participants dropped out when asked to provide their home address.

New supplemental data elements

From one survey year to another, some questions were added while others were dropped. A handful of these are listed below, along with their corresponding variable_names in `monospaced` text.

- 2021, 2023: Questions about typical electric vehicle charging behavior (`ev_typical_charge_[1-6]`), charging behavior on trips (`ev_charge`, `ev_charge_time`, `ev_charge_station_decision`, `ev_charge_station_level_[1-3]`, and attitudes towards EV purchasing) were added.
- 2021, 2023: Proxy reporters for school-aged children were asked if their child traveled to school, on days when no school trips were otherwise present in the travel diary (`attend_school`). When the child did not travel to school, the proxy reporter was asked to select reason(s) they did not travel to school on that day (`attend_school_no`).

- 2021, 2023: Additional questions about bicycle types, micro-mobility devices (`micromobility_devices`), and bike storage locations (`bike_store`) were asked of each household.
- 2019, 2021, 2023: Travel modes (`mode_type_detailed`) and shared mobility services (`share_[0-7]`) were verified each year for their relevance to the region.
- 2021, 2023: Participants were asked if one or more barriers to travel impacted them in the past week (variable `transportation_barriers`).

Questionnaires

The complete questionnaires are attached in the appendix.

> Note
>
> In 2019, the web browser-based and smartphone-based questionnaires had slightly different structures. In 2021, RSG updated the technology underlying the rMove app, such that the web- and smartphone-based surveys drew from the same source. For this reason, the 2019 web-based and smartphone surveys slightly differ in their structure.

- 2019 Web-based Survey Questionnaire
- 2019 Smartphone Questionnaire
- 2021 Questionnaire
- 2023 Questionnaire

5.0 SURVEY BRANDING, COMMUNICATION AND ADMINISTRATION

RSG developed the study branding collaboratively with Met Council, including the study name, color scheme, and font selections. The study logo was used broadly in the print materials, online survey, outreach efforts, and on the study web page.

5.1 INVITATION MATERIALS

Each invited household received two types of mailings:

Invitation packet: A formal study invitation packet was mailed to prospective participant households for each Sample Order of Wave 1. The cover letter explained the study purpose, described the steps necessary to complete the study, and included an FAQ sheet on the reverse. The packet also contained a translated insert where instructions on how to participate were translated into all language offerings (see Section 4.2).

Reminder postcards: Reminder postcards were mailed after the invitation packet to encourage every household to complete the study. In 2019 and 2021, two postcards were sent: the first card 2-3 days after the invitation packet, and the second card 2-3 days later. In 2023, only one reminder postcard was sent, seven days after the invitation packet. Postcards included the study phone number, web page, and participant login information.

Study Website

A project web page (MSPtravelstudy.org) was developed to describe the study and facilitate participation of invited households. This website provides information about the project, including FAQs, quotes of support, study news, the toll-free study telephone number, and a contact form for online participant support. The website served as a resource for participants as well as a clear entry point for invited households to enter their unique access code and complete their surveys.

Public Outreach

Especially in 2019, the project team implemented a community outreach program for the Travel Behavior Inventory.

Outreach efforts were aimed at both the general population (to increase overall participation) and at hard-to-reach populations such as rural, minority, and low-income households (to increase sample representativeness). General population outreach was conducted primarily through unpaid content on social media, the Met Council website, emails, and where possible, through earned media coverage. Outreach to hard-to-reach populations including rural, minority, non-English speaking, and low-income households were conducted through a mix of paid and unpaid content placement as well as engagement with key community groups and stakeholders.

Efforts included:

1. Posts on social media at Met Council accounts.
2. Updates the study website and page on the Met Council website.
3. Press releases.
4. Newsletters and emails to available listservs.
5. Advertising content formatted for multicultural print, digital, and radio buys.

Broad-based public outreach was found to be less effective than desired; as a result, the 2021 and 2023 surveys dropped most of these broad-based public outreach efforts while maintaining a presence at the Met Council website to legitimize the study for those seeking information online.

5.2 PARTICIPANT SUPPORT

Inbound Participant Support

The study maintained three inbound communications channels to support participants before, during, and after their data collection periods. First, a call center with a toll-free number helped participants answer their online surveys and to help answer questions from participants or curious nonparticipants. Second, an email inbox fielded similar inquiries. Third, rMove™ allowed participants to submit feedback using the app, which was then responded to over email. All three communications channels were staffed with the intent to respond to inquiries within one business day.

Outbound Participant Support

The call center also provided outbound reminder calls to select households that used the online travel diary to remind them about their surveys before and after their assigned travel day. All households that used rMove™, and most households that used the online diary, received their travel period reminders with group-specific and timely emails.

6.0 DATASET PREPARATION, QUALITY ASSURANCE, AND QUALITY CONTROL

6.1 OVERVIEW

RSG conducted dataset preparation and quality control procedures at every stage of the study (before, during, and after data collection). These procedures were designed to validate survey logic, review participant experience, and confirm consistent data coding in the survey database. The following sections summarize the various dataset preparation and quality control steps. RSG provided a separate QAQC Plan to the Met Council for each wave of survey collection; these plans include data cleaning details for key elements.

Database Setup and Real-Time Quality Controls

Prior to a survey launch, RSG and the Met Council reviewed the survey instruments to ensure that the survey interface was clear and easy to use, questions were understandable, and variables wrote out to the database as expected. To reduce survey burden and improve final data quality, the survey also included real-time data checks and logic. Examples of these checks include the following:

- Validation logic to prevent skipped questions.
- Logic checks to hide irrelevant questions and answers (e.g., employment questions for children).
- Spatial and temporal checks within trip rosters to prevent overlapping trips.

These real-time data checks do not eliminate every inconsistency, but they do significantly reduce reporting errors and re-coding requirements after data collection.

Geographic Data Checks

During data collection, the survey instruments used the Bing Maps API to geocode the coordinates for reported home, work, school, and trip addresses.

Following data collection, RSG also coded home location points to block groups and broader regional definitions.

Trip Derivation for Nonparticipating Household Members

Household travel surveys require data for all household members to assess complete household travel patterns. However, some exceptions are allowed in the data collection process where travel can be reported by proxy, particularly for children.

Household adults were asked to report travel for the children in the household (under age 18). Participants could also report children of all ages as travel party members on their own trips. RSG used these records to derive diary records for children under age 18.

Completion Criteria

The last step of dataset preparation involved reviewing all data records to confirm that they met survey, travel day, and household completion criteria. “Complete” households met the following conditions:

1. The household completed the online recruitment/demographic survey.
2. All ABS household members provided complete travel diary information (i.e., answered all surveys and reported all trips). Online panel members provided complete travel diary information for themselves (person 1 in the household).
3. The household reported a home address within the study region.

In 2023, outreach segment households were marked as incomplete because they did not meet criteria 1 and 2: outreach participants completed the survey for themselves, but did not report complete information for their household.

Imputation

Departure Time

In some cases, the rMove™ app may have detected the start of a trip after its true start time, which can yield invalid or extreme values for trip duration and speed. In these cases, the fields depart_date, depart_hour, and depart_minute were adjusted for “late pickup” conditions using the following approach:

- Departure time was imputed using the median speed between all locations along the trip, excluding the origin point, and the distance between the origin and the next point on the trip. For trips with fewer than three recorded locations, imputed departure time is set three minutes earlier than the original departure time to compensate for rMove’s 3-5-minute ping interval. Note that some trips that are the result of split loop trips may only have three or fewer points but will use the imputed depart time from before the loop trip was split and thus may not be included in this rule.
- If the imputed departure time overlaps with the previous trip’s arrival time, the previous trip’s arrival time was instead used as the departure time. Regardless of the number of locations along a trip, if the imputed departure time was later than the initially reported departure time, the imputed departure time is set to the original departure time. User-added trips as well as long distance passenger mode trips are also set to the original departure time, as user-added trips are not subject to “late pickup” conditions, and long-

distance passenger modes are often plane trips where all collected traces contain speed information from other modes and thus are less reliable (as rMove™ cannot collect locations when a phone is in “airplane mode”).

Duration and speed are calculated based on the imputed departure time.

Purpose

Respondents report the purpose of the trip destination in each trip survey. The origin purpose is derived from the destination purpose of the previous trip, except for the first trip in the travel period or where an rMove™ trip occurs after a trip with item non-response. For the first trip in the travel period, the origin purpose can be inferred from “begin_day” in the day table.

When purpose was not asked because an analyst split a user-reported trip during data cleaning (creating a new destination along a trip), purpose values are derived where possible based on proximity (within 150 meters) to estimated home, work, or school locations. If the location is not proximate to home, work, or school locations, the purpose is set to “other.”

The purpose category variables (o_purpose_category, d_purpose_category) contain aggregated purpose values based on the type of purpose at the origin/destination of each trip. Dataset users are welcome to perform their own recoding of the purpose categories as well.

Trip purposes have been imputed in cases where a purpose reported by the user is assumed to be inaccurate based on information about that person’s reported habitual locations and other trips (primarily to home, work, and school locations). The trip purpose imputation approach was applied to all rMove™ trips in person-days with at least 1 complete trip and no more than 10 incomplete trips. (“Incomplete” trips are trips for which the respondent did not answer the trip-specific survey questions about purpose, mode, etc. for the given trip.)

The approach was to apply various “tests” in logical sequence to trips for which the stated purpose is not consistent with the location type based on the reported habitual locations. In general terms, the tests were designed to:

- Check the respondent’s reported destination purpose when it conflicts with the destination location type. (The details of the tests depend on the trip purpose, with different criteria used for change-mode trips, escort trips, linked transit trips, trips with home destinations but other reported purposes, etc.)
- Identify cases where respondents swapped the order of two or more trips when reporting their details.
- Identify cases where respondents may have omitted a trip and shifted remaining reported trip details by one trip when reporting the rest of their trips.
- Fill in missing data by sampling destination purposes from other trips made to the same locations, either by the same respondent or by other respondents.

Mode type (`mode_type`)

The variable `mode_type` synthesizes `mode_1` to `mode_3` down to a single, easier-to-use variable for analytical purposes (so that data users can avoid always referencing all modes on multimodal trips). Table 10 shows the full crosswalk of which detailed modes correspond to which `mode_types` in the 2023 data. Higher values of `mode_type` are prioritized over lower `mode_type` values in the derivation. For example, transit trips, with `mode_type` 13, are prioritized over walk trips, with `mode_type` 1. When transit trips were unlinked using the Google API during cleaning, the non-transit legs of the trip were recoded using Google's suggested mode (most frequently "walk" or "bike") and do not have a reported `mode_1`, `mode_2`, or `mode_3`.

TABLE 10. MODE TYPE HIERARCHY (2023).

DETAILED MODE VALUE	DETAILED MODE VALUE	MODE TYPE VALUE	MODE TYPE LABEL
1	Northstar	1	Rail
2	Light rail (e.g., Blue Line, Green Line)		
3	Other rail		
4	School bus	2	School Bus
5	Bus rapid transit (e.g., A Line, C Line, Red Line)	3	Public Bus
6	Express/commuter bus		
7	Local bus		
8	Dial-A-Ride (e.g., Transit Link)		
9	Metro Mobility		
10	SouthWest Prime or MVTA Connect		
11	Employer-provided shuttle/bus	4	Other Bus
12	University/college shuttle/bus		
13	Other private shuttle/bus (e.g., a hotel's, an airport's)		
14	Vanpool		
15	Other bus		
16	Intercity rail (e.g., Amtrak)	5	Long distance passenger mode
17	Intercity bus (e.g., Greyhound, Jefferson Lines)		
18	Airplane/helicopter		
19	Uber, Lyft, or other smartphone-app ride service	6	Smartphone ridehailing service
20	Regular taxi (e.g., Yellow Cab)	7	For-Hire Vehicle
21	Other hired car service (e.g., black car, limo)		
22	Household vehicle 1	8	Household Vehicle
23	Household vehicle 2		
24	Household vehicle 3		
25	Household vehicle 4		
26	Household vehicle 5		
27	Household vehicle 6		
28	Household vehicle 7		
29	Household vehicle 8		
30	Other vehicle in household		
31	Other motorcycle in household	9	Other Vehicle
32	Other motorcycle (not my household's)		
33	Car from work		
34	Friend/relative/colleague's car		
35	Rental car		
36	Carpool match (e.g., Waze Carpool)		
37	Carshare service (e.g., Zipcar)		
38	Peer-to-peer car rental (e.g., Turo)		
39	Other vehicle (not my household's)		
61	Electric vehicle carshare (e.g., Evie)		
40	Electric bicycle (my household's)	10	Micromobility
41	Standard bicycle (my household's)		
42	Borrowed bicycle (e.g., a friend's)		
43	Bike-share - standard bicycle		
44	Bike-share - electric bicycle		
45	Other rented bicycle		
46	Personal scooter or moped (not shared)		
47	Scooter-share (e.g., Bird, Lime)		
48	Moped-share (e.g., Scoot)		
49	Segway		
50	Other scooter or moped		
51	Skateboard or rollerblade		
52	Other boat (e.g., kayak)	11	Other
53	Vehicle ferry (took vehicle on board)		
54	Other public ferry or water taxi		
55	Golf cart		
56	Snowmobile		
57	ATV		
58	Medical transportation service		
59	Other		
60	Walk (or jog/wheelchair)	12	Walk

This mode type hierarchy table contains the values from the 2023 dataset; some names for mode types have changed slightly since the 2019 and 2021 surveys. For more information, consult the combined codebook.

iOS Trip Trace Irregularities (Wave 3 2023 Data Only)

The release of iOS 16.4 by Apple on March 27, 2023, brought about significant changes to background location tracking, affecting apps such as rMove, which rely on collecting location information. Consequently, iPhone users with iOS 16.4 or later experienced irregular trip traces within the rMove™ app, impacting data accuracy for Spring 2023.

To address this issue, RSG swiftly updated the rMove™ app and monitoring scripts to mitigate inconsistencies in future data collection. Despite these efforts, the 2023 dataset remained affected. To manage the impact on the dataset and downstream processes, RSG developed a series of criteria to identify suspect trips and flag individuals or households accordingly. Additionally, adjustments in weighting and dataset delivery were made to ensure maximum data utility.

Any suspect trip trace records were identified and the dataset was provided with multiple weights. One set of weights with the full dataset and one set of weights to use if applying this strict criteria, so that any trip analysis metrics could exclude potential trip trace irregularities.

6.2 COMBINED DATASET (2019, 2021 AND 2023)

To facilitate analyses across waves of the survey, RSG developed a cross-wave combined dataset and codebook.

Combined Codebook

> ⓘ Note
>
> Download an Excel version of the combined codebook by clicking here.

RSG typically delivers data in its raw form, with the numeric codes that correspond to survey entries instead of the text seen by participants. The codebook allows data users to translate survey results into human-readable format, and is comprised of two parts:

- A variable list, which includes attributes at the level of individual survey questions; and
- A value labels table, which corresponds to attributes of survey responses.

A combination of manual and scripted processes were used to create a combined codebook:

1. **Variable crosswalk (manual process performed in Excel).** First, a variable crosswalk table was constructed by aligning variable names and survey questions across years. Where variable names differed, but survey question meaning stayed the same, a "unified" variable name was chosen. Logic, variable descriptions, location of the variable in the database, and other attributes were inspected manually and with a combination of processes in an Excel workbook (i.e., using VLOOKUP processes and other formulas).

Example: The variable "fuel" in 2019 is renamed "fuel_type" in 2023. The "unified" variable name becomes "fuel_type."

2. **Value label crosswalk (manual process performed in Excel).** Next, a value label crosswalk was constructed in a similar manner. A crosswalk for numeric value inputs was created for where value labels differed for the same numeric value entry. These "unified" values and value labels were then used to construct "upcoded" values and value labels, which consolidated across disparate categories with similar meanings.

   Example: "Plug-in hybrid (PHEV)" and "Hybrid (HEV)" vehicle fuel types, used in 2019 data, are upcoded to "Hybrid" (2021, 2023 data).

TABLE 11. COMBINED CODEBOOK EXAMPLE: FUEL TYPE.

2019	2021	2023	UNIFIED	UP-CODED	2019	2021	2023	UNIFIED	UP-CODED
-9998	--	--	-9998	995	Missing: Non-response	--	--	Missing	Missing
1	1	1	1	1	Gas	Gas	Gas	Gas	Gas
--	2	2	2	2	--	Hybrid (HEV)	Hybrid (HEV)	Hybrid (HEV)	Hybrid
3	3	3	3	2	Hybrid	Plug-in hybrid	Plug-in hybrid	Plug-in hybrid	Hybrid
4	4	4	4	3	Electric	Electric (EV)	Electric (EV)	Electric (EV)	Electric (EV)
2	5	5	5	4	Diesel	Diesel	Diesel	Diesel	Diesel
--	6	6	6	5	--	Flex fuel (FFV)	Flex fuel (FFV)	Flex fuel (FFV)	Other
997	7	7	7	5	Other	Other (e.g., natural gas, bio-diesel)	Other (e.g., natural gas, bio-diesel)	Other (e.g., natural gas, bio-diesel)	Other
995	995	--	995	995	Missing: Skip logic	Missing	--	Missing	Missing

Combined Dataset

A scripted process, written in R and relying on the combined codebook, was used to create a single dataset containing all three waves of survey data. The scripted process:

1. Renamed dataset columns (variables) from their year-specific names to the “unified” names chosen in the combined codebook.
2. In each column (for each variable), replaced year-specific numeric response codes with their “unified” response codes.
3. Repeated step 2 with “upcoded” response codes, to create an “upcoded” dataset.

Special Output: Trip Purpose Table

The trip_purpose table was derived from the upcoded and unified trip tables. Its purpose is to aid data analysis of overall trip purposes. This table was developed because the origin and destination "purpose" categories in the trip table can contain non-intuitive classifications. For example, a summary of destination purposes will have many trips "home" but the overall trip purpose for that trip home might actually correspond to the non-home trip end (work, school, etc).

Removing trips with a destination of "home" from the overall analysis is one option, but this can lead to some place types being missing from the final dataset when the trip roster for a person's day is incomplete. For example, if a person's trip diary for a day consists of a trip from a friend's house to their home, the place type "friend's house" will be missing from the final summary of trip purposes.

To account for all place types – both origin and destination – in the final trip purpose summaries, the following steps were used to create the trip purpose table:

1. Transit trips that had been unlinked into access, transit and egress legs were re-linked, by consolidating multiple legs of transit trips into a single record. This removes "change mode" trips from the table, except for long-distance trips. The trip weight for each linked trip was set to the maximum trip weight its composite unlinked trips.
2. Trips were placed in two categories: home-based (having one trip end at home) and non-home-based trips.
3. Home-based trips' purposes were classified as the non-home end. The weight for this trip purpose record is equal to the original trip weight.
4. Non-home-based trips were split into two records for each trip: one for the origin end, and a second for the destination end. The weight for each record was set to half of the original trip weight.

The tables below show a hypothetical example of this process for the travel diary corresponding to `day_id` `199885710201`.

In the trip table, there are four records that correspond to this day (Table 12). The person left home, went to work, went on an exercise trip or to the gym, picked up someone from school, and finally returned home.

TABLE 12. TRIPS ON DAY 199885710201.

TRIP_ID	O_PURPOSE	D_PURPOSE	TRIP_WEIGHT
1998857102001	Went home	Primary workplace	233.4265
1998857102002	Primary workplace	Exercise or recreation (e.g., gym, jog, bike, walk dog)	363.7046
1998857102003	Exercise or recreation	Pick-up/drop-off to/from K-12 school or college	363.7046
1998857102004	Pick-up/drop-off to/from school	Went home	363.7046

An analysis of trip purpose by destination place types (`d_purpose`) would yield an overall trip purpose share of 16.7% trips to work, 33.4% of trips to exercise, 30.2% of trips to escort others to school, and 19.8% of trips to home (Table 13).

TABLE 13. TRIP DESTINATION PURPOSE SHARE, DAY 199885710201.

D_PURPOSE	TRIP_WEIGHT	PURPOSE_SHARE
Primary workplace	233.4	17.6%
Exercise or recreation	363.7	27.5%
Pick-up/drop-off to/from K-12 school	363.7	27.5%
Went home	363.7	27.5%
Total	**1,324.5**	**100.0%**

In the trip purpose table for the same day ID, there are six rows instead of four – the trip from work to exercise, and from exercise to pick-up someone from school, have both been expanded to two rows to allow trip weight to be distributed across them. For the home-based trips, the trip purpose has been assigned to the non-home end of the trip (work and escort, respectively) (Table 14).

TABLE 14. TRIPS PURPOSE RECORDS FOR DAY 199885710201.

TRIP_PURPOSE_ID	PURPOSE	TRIP_PURPOSE_WEIGHT
181704	Primary workplace	233.4265

181705	Pick-up/drop-off to/from school	363.7046
480995	Primary workplace	181.8523
480996	Exercise or recreation	181.8523
732042	Exercise or recreation	181.8523
732043	Pick-up/drop-off to/from school	181.8523

Summarizing the trip purpose table yields 33.4% of trips "for" work (i.e., one-third of trips are work-related), 31.8% of trips for exercise or recreation, and 34.9% of trips to pick up others from school (Table 15).

TABLE 15. TRIP PURPOSE SHARE FROM TRIP PURPOSE TABLE, DAY 199885710201.

	TRIP_PURPOSE_WEIGHT	PURPOSE_SHARE
Primary workplace	233.4	17.6%
	181.9	13.7%
Pick-up/drop-off to/from school	363.7	27.5%
Exercise or recreation	181.9	13.7%
	181.9	13.7%
TOTAL	**1,324.5**	**100.0%**

Alternatively, the data user could remove trips "home" and calculate purpose share using the destination purpose, using only the subset of trips that do not end at home. For this travel day, calculating purpose share from the `d_purpose` for the subset of trips that do not end at home yields a greater share of trips for exercise and pick-up relative to the calculations from the trip purpose table (Table 16).

TABLE 16. TRIP PURPOSE SHARE FROM TRIP TABLE, NON-HOME TRIPS, DAY 199885710201.

D_PURPOSE	TRIP_WEIGHT	PURPOSE_SHARE
Primary workplace	233.4	24.3%
Exercise or recreation	363.7	37.9%
Pick-up/drop-off to/from school	363.7	37.9%
TOTAL	**960.8**	**100.0%**

Table 17 below shows how using destination purpose in the trip table compares to using the overall trip purpose in the trip purpose table. The vast majority of trips “home” have been re-categorized (a small number remain, where both origin and destination were “home,” i.e., loop trips without an intermediate stop point). The estimate of total number of trips differs across the two tables as well, because the trip purpose table has consolidated “change mode” trips into a broader linked trip purpose.

TABLE 17. TRIP PURPOSE SHARE FROM TRIP AND TRIP PURPOSE TABLE (2023 ONLY)

	TRIP TABLE		TRIP PURPOSE TABLE	
PURPOSE/D_PURPOSE	**SHARE**	**WTD TOTAL**	**SHARE**	**WTD TOTAL**
Home	32.4%	4,339,661	0.1%	10,526
Shopping	10.8%	1,447,777	16.3%	1,927,275
Social/Recreation	9.7%	1,305,811	16.0%	1,889,502
Escort	9.6%	1,285,127	15.6%	1,847,927
Work	8.2%	1,096,092	13.4%	1,587,782
Meal	6.9%	929,880	10.4%	1,229,728
Errand	5.5%	731,313	9.9%	1,167,337
Work related	5.2%	694,438	7.3%	867,563
School	3.5%	468,666	5.7%	673,190
Change mode	3.4%	451,611	NA	NA
Overnight	2.5%	335,107	3.4%	398,291
Other	1.9%	258,319	1.2%	146,963
School related	0.5%	65,032	0.7%	82,248
TOTAL	**100.0%**	**13,408,835.3**	**100.0%**	**11,828,331.3**

Dataset Composition

The final unweighted datasets includes seven distinct data tables. These tables include all user-input survey variables, certain survey metadata (e.g., survey completion mode), and variables derived to support data analysis.

- Household: 19,170 complete households
- Person: 38,691 people
- Vehicle: 30,239 vehicles
- Day: 157,947 days
- Trip: 623,926 unlinked trips
- Location: 12,345,002 points
- Trip Purpose: 343,372 linked, single-ended trips and 251,366 unlinked, two-ended trips

7.0 EXPANSION AND WEIGHTING

7.1 INTRODUCTION

This section summarizes data weighting for Travel Behavior Inventory datasets.

The Travel Behavior Inventory is tasked with recording a representative snapshot of regional travel behavior. Two mechanisms help the TBI team accomplish this task: the first is **randomized sampling** of addresses in the region, and the second is weighting to account for non-response bias.

Non-response bias occurs when certain demographic groups (for example, young people or large households) are harder to survey than others. RSG corrects for a portion of non-response bias in the sampling phase, by increasing the number of randomly selected addresses in geographic areas where hard-to-recruit groups tend to live, but some bias in the final survey sample will always persist.

Weighting corrects for non-response bias by expanding the representation of the survey from only the sample itself to the population from which it is sampled along multiple key dimensions (sourced from Census data). RSG has historically prioritized adjustments for representation across the following dimensions:

- Household size, income, vehicles, workers, and presence of children
- Person sex, age, worker status, typical work mode, education, race, ethnicity
- Aggregated geography (e.g., PUMA or groups of PUMAs)

Near the conclusion of the six-year project, RSG decided to re-weight both 2019 and 2021 datasets while weighting the 2023 dataset to uniformly apply weighting methodologies across all three waves of data collection (See Re-Weighting Memo in the appendix). As a result, all weighting procedures were nearly identical across all three waves, with the exceptions that the study segmentation changed from 2019 to 2021. Namely, the “core urban” segment was sub-divided into five sub-segments, while the “hard-to-reach” segment was absorbed into other segments (urban, rural, rural ring). For details on the sample segmentation, see the Survey Sampling section of this report.

The weighting process generates four types of weights:

1. **A household-level weight.** The sum of these weights reflects the total households in the survey region.
2. **A person-level weight.** The sum of these weights reflects the total persons in the survey region.

3. **A day-level weight.** The sum of these weights also reflects the total persons in the survey region (and matches the sum of the person-level weights). The person weights are spread evenly across the number of complete weekdays, so the table represents the sum of one average weekday for each person in the study.
4. **A trip-level weight.** The sum of these weights represents the total number of trips all persons residing in the region make on a typical weekday (i.e., Tuesday, Wednesday, and Thursday). This weight should be used for trip-level analyses.

> Note
>
> Note: this differs from the number of trips made in the survey region on a typical day, given that some residents make trips outside the region.

7.2 WEIGHTING PROCESS

The survey weighting process includes five primary steps:

1. **Initial expansion**: Calculating an "initial weight" based on the probability of selection in the sample design. This step essentially "reverses" the sample plan, providing higher initial weights to areas where less sampling occurs. Each household is assigned an initial weight based on its probability of being invited to the survey. For example, if 5 households respond in a geography with 100 total households, each of the 5 households will receive an initial weight of 20 (100 / 5 = 20).
2. **Target-optimized weighting**: After the initial expansion, household weights are adjusted to simultaneously fit selected household- and person-level targets. For example, if 20% of households in the state are one-person households, but 25% of the sample households are one-person households, RSG adjusts the weights to better match the household size distribution of the population. This process leverages an open-source application, PopulationSim, which optimizes the household weights against all target control variables simultaneously (e.g., household size, vehicles, age, race, ethnicity, gender, employment, education). This step is performed twice in the weighting routine: once after the initial expansion (step 1) and again using additional targets estimated from the day-pattern (step 3).
3. **Re-weighting with day-pattern adjustment to account for multi-day survey data:** Some survey respondents will provide data for one travel day while others will provide data for two or more days. To ensure multi-day respondents are not over-represented in the final data, RSG estimates a simple day-pattern model using the initially weighted data, normalized by the number of complete weekdays per person. The normalized model then estimates the frequency of each daily-activity type (mandatory trips, non-

mandatory trips, no trips) to account for the multi-day bias. The aggregated weighted frequencies are then included as additional control target variables and the data are re-weighted in PopulationSim a second time to ensure the households are weighted to account for this multi-day bias so that households with more days are not equal weight with households with just one day.

4. **Calculating person, day, and trip weights:** The household weight is assigned to subsequent person, day, and trip weights.
5. **Adjusting for non-response bias in day-pattern and trip rates**: During this final step, RSG adjusts the trip-level weights to account for survey biases based on the method respondents use to report their travel. For example, if respondents who report their travel over the phone report fewer non-home-based-work trips compared to respondents who report their travel in the smartphone app (after correcting for differences in demographics), the day- and trip-level weights for respondents who report their travel over the phone are adjusted to align with the smartphone app respondents more closely. Travel reported by respondents who use the smartphone app is considered more accurate because respondents are not required to recall their travel and are therefore less likely to under-report trips.

7.3 WEIGHTING UPDATES

Over the course of the six-year TBI program contract between the Council and RSG, the survey team made a number of substantial improvements to weighting. We expect weighting to be under continuous development, both as new challenges arise (e.g., incorporating new types of samples, like panel samples) and as new methods are developed to correct for bias (e.g., developing new statistical models to account for survey bias).

To ensure compatibility of weighted data over time, the survey team has re-weighted data when new methods are developed. These re-weighting instances are listed below.

2021: Transition from Iterative Proportional Fitting (IPF) algorithm to PopulationSim Entropy Maximization (EM) algorithm. RSG implemented an improved method that matches sample (survey) data to target (Census) data during the second wave of the Travel Behavior Inventory. This method resulted in more reliable weighting results and was less prone to the influence of small sample size or outlier data. To bring Wave 1 (2019) data into line with the new methods, RSG re-weighted Wave 2.

2024: Use of a transit trip target. After examining weighted data for all three waves of the survey (2019 – 2021 – 2023), the team decided to incorporate a new type of target in weighting: transit boardings. Using published data on boardings from the NTD and Metro Transit, RSG developed a method to tie the final, household-level transit trip rates to a known number. This allowed the team to correct for a bias towards transit users in the sample, which was not sufficiently addressed with Census demographic targets.

2024: Improvements to handling of non-related persons. In 2021 and 2023 (the second and third waves of the survey program), the survey team stopped requiring full travel diaries from roommates and household help. Because the Census defines a household as all those living under one roof, this necessitated adjustments to the way RSG matches sample (survey) data to Census targets. In 2021, RSG used a method whereby raw Census PUMS data was manipulated to account for the "missing" travel diaries of those non-related householders. In 2024, RSG revised this method, using a correction during final weight-setting instead. This preserved the integrity of PUMS data and increased the transparency of methods. Along the way, RSG also updated its methods for imputing household income to include the "unreported" household income of those non-related householders.

7.4 USING THE WEIGHTS

> Note
>
> For a detailed explanation of how to use the weights with examples in R, consult the data user's guide in the appendix.

Analyses designed to draw conclusions about travel behavior in the region (as opposed to just the survey respondents) should use weighted data.

When applied, the weights make the dataset representative of travel for residents within the study region for the time period studied.

Just a reminder that the dataset does not include:

- Commercial vehicle travel
- Travel for persons residing in group quarters outside of the address-based sample frame (e.g., college dorms, institutional housing)
- Travel from non-residents (i.e., visitors to the region)
- Seasonal/holiday travel outside of the survey fielding period.

Data users should also keep in mind the following when creating weighted statistics and summaries from HTS data:

- **Filter to the data relevant to your analysis.** Note that not all people are asked every question, so understanding the 'missing value' codes and 'survey logic' in the data dictionary are important.

- **Remember the survey design when using and interpreting weighted values.** For example, the Travel Behavior Inventory included both one-day online and call center participation and seven-day smartphone participation. Therefore, it is best to avoid filtering by day of week since not all participants traveled on all days.

Calculating Weighted Summaries

In general, household weights should be used for household- and vehicle-level analyses; person weights for person-level analyses, day weights for day-level analyses, and trip weights for trip-level analyses. To calculate weighted summaries or descriptive statistics, sum the weights for that table.

In the case of an analysis that requires variables from two or more tables, the weight from the lowest level of the data hierarchy should be used (trip weights, then day weights, person weights, and finally household weights). For example, a weighted summary of race (a person-level variable) by household income should use person-level weights.

Calculating Trip Rates

Special considerations apply when generating weighted trip rates.

To calculate a weighted trip rate –the number of trips per day–data users must divide the number of weighted trips by the number of weighted travel days.

For example:

- If there are 300,000 weighted person-trips across 75,000 person-days, then the average person-trip rate is 4.0 per day.
- If there are 225,000 person-trips by car across 75,000 person-days, then the person-trip rate for car trips is 3.0.

> ⓘ Note
>
> Data users should always calculate the number of weighted travel days using the day table rather than the trip table given that persons with zero-trip travel days do not have any records in the trip tables for those days.

Variance Estimation

There are two common approaches for variance estimation using weights. The first, **replicate weights**, requires the data user to leverages multiple sets of weights for the same set of observations. Analyses using each set of weights are “averaged” to create variance estimates. This procedure has major drawbacks, because the replicates increase dataset complexity and add require complex weighting procedures.

The preferred approach is an approximation using **Taylor-series linearization**. This procedure approximates the variance using a simpler to implement formula, simplifying the dataset and weight generation. It is easily implemented in the R library {survey} or the {samplicspackage} in Python, both of which are well-sourced packages with deep use among survey researchers.

Throughout this report, standard error bars are estimated using the Taylor-series linearization approach in the R package {survey}.

Combining weights

To align with the ACS multi-year weighting approach[1], RSG has delivered single-year weights and recommends calculating multi-year weights by taking a simple average of weights. For example, to combine the 2019, 2021 and 2023 data into a single sample, the weights would be divided by three, resulting in a sum of weights (e.g., for households) that is an average of the population between 2019 and 2023. Note, however, that the effects of the COVID-19 pandemic on travel in 2021 could make such a combined dataset less useful in some regards.

[1] American Community Survey 2010, page 11-16: https://www.census.gov/content/dam/Census/library/publications/2010/acs/Chapter_11_RevisedDec2010.pdf

8.0 SURVEY RESULTS

8.1 SAMPLE PLAN EVALUATION

Overall response rates declined over the course of the study, from 2.54% in 2019 to 2.16% in 2023.

TABLE 18. RESPONSE RATES BY SURVEY YEAR.

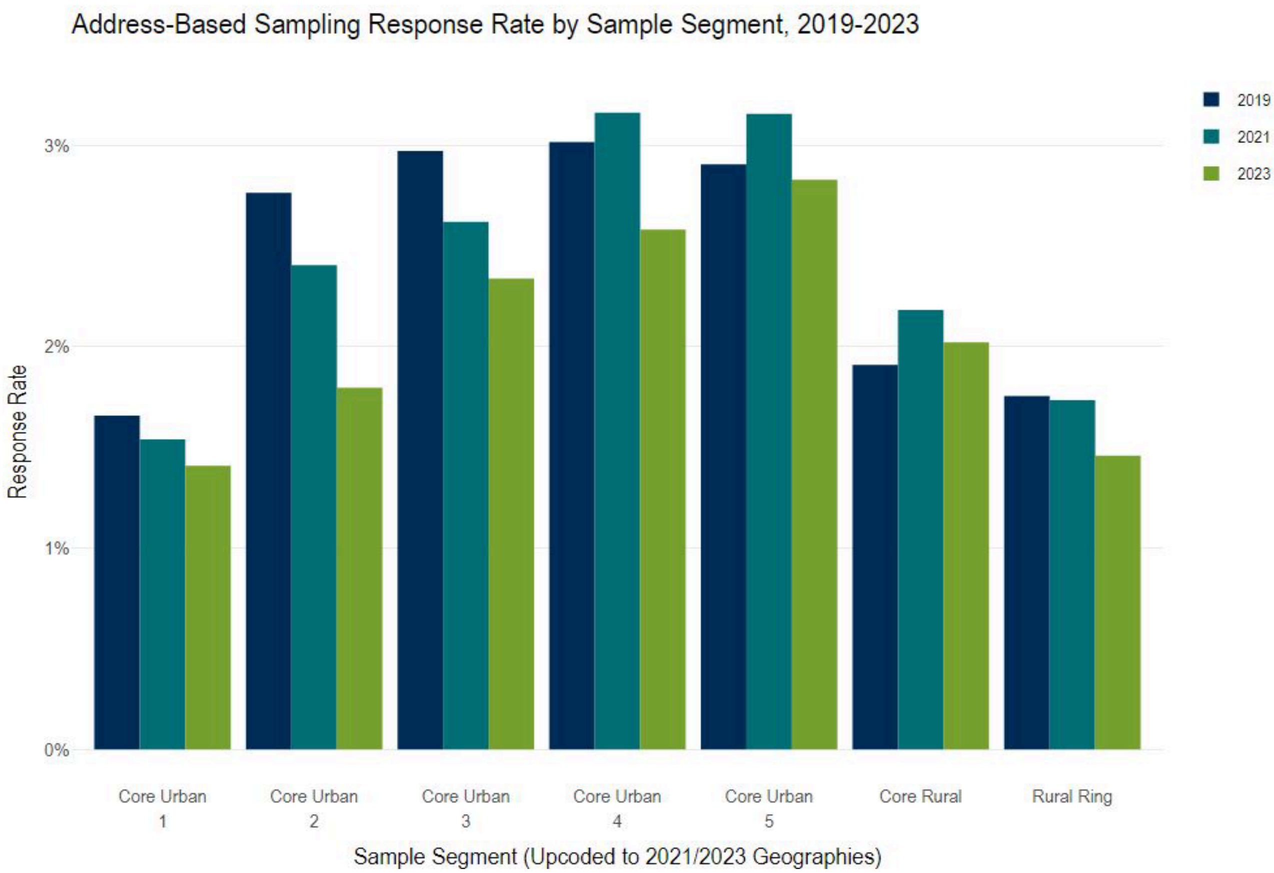
	ADDRESS-BASED SAMPLING			SUPPLEMENTAL SAMPLING
Survey Year	Invited Households	Complete Households	Response Rate	Complete Households
2019	296,465	7,516	2.54%	0
2021	282,698	7,136	2.52%	771
2023	173,701	3,749	2.16%	437*
*Households did not meet data requirements for weighting.				

Figure 9 shows the Address-Based Sampling (ABS) response rates by sample segment from 2019 to 2023, divided by sample segment. To aid comparisons across years, households in 2019 were assigned the sample segment assigned to their home block group in 2021 and 2023.

Urban areas tended to have higher response rates than rural areas across all years. In 2021, some segments saw an increased response rate, perhaps due to the change in survey methodology that year, which allowed households to complete using a one-day online diary even if they were eligible for a seven-day smartphone diary.

In 2023, the second reminder postcard was dropped as a cost-savings measure. Small but noticeable declines in response rates followed, and were especially noticeable in urban segments, and less pronounced in Core Rural and Rural Ring segments, where response rates were already quite low.

FIGURE 9. ABS RESPONSE RATE BY SAMPLE SEGMENT.



8.2 PARTICIPANT DEMOGRAPHICS

This section analyzes the demographic composition of the final dataset. Unless otherwise noted, all analyses use unweighted data, for an evaluation of how closely the dataset matched the ACS data for the region.

The demographic composition of survey participants changed substantially from 2019 to 2021, in many cases reflecting four changes to study design initiated during the 2021 survey:

1. A change to sample strata in order to better target hard-to-survey households;
2. Higher incentive offerings in invitations sent to block groups with a greater share of low-income households;
3. Higher incentive offerings at the conclusion of the signup survey, for households making less than $50,000 per year; and

4. Incorporation of supplemental sampling through community-based organizations (2021) and transit rider lists (2021). Additional community-based outreach was performed in 2023, but are not included in these analyses as the data were not eligible to be weighted.

In 2021 and 2023, the survey had a higher share of lower-income (i.e., 50K or under) households participate in the survey than are found in the region, indicating that oversampling, supplemental sampling and differential incentives were effective (Table 19).

TABLE 19. PARTICIPANT HOUSEHOLD INCOME BY SURVEY YEAR (UNWEIGHTED)

HOUSEHOLD INCOME	2019	2021	2023	2015-2019 ACS
Less than $15,000	4.6%	9.7%	7.7%	6.2%
$15,000-$24,999	5.6%	7.8%	7.4%	5.7%
$25,000-$34,999	6.4%	8.5%	6.7%	6.2%
$35,000-$49,999	11.1%	10.7%	9.7%	9.8%
$50,000-$74,999	19.8%	18.6%	17.0%	15.1%
$75,000-$99,999	17.1%	14.8%	16.5%	13.0%
$100,000-$149,999	19.8%	17.2%	18.9%	17.6%
$150,000-$199,999	8.8%	7.6%	8.0%	8.6%
$200,000 or more	6.7%	5.2%	8.1%	17.9%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Across all three waves, the survey had a higher share of one-person and two-person households participate than are found in the region (Table 20).

To boost sample representation for large households, the 2021 Travel Behavior Inventory implemented incentive structure changes for those households in the final months of fielding: households with four or more members were offered higher incentives after completing the signup survey (when they provided info on household size). Overall, this offering was effective at increasing the conversion rate (thus decreasing attrition) for large households for whom it is more burdensome to complete the survey but was not sufficient to bring the share of large households in the survey to closer equivalence with ACS percentages.

The weighting process adjusted for this discrepancy, which is commonly seen in household travel surveys. Survey burden is often provided as a reason for lower response rates in large households. Future iterations of the Travel Behavior Inventory might seek to implement substantive changes to the survey itself to reduce burden, rather than try to correct the problem through sampling and incentive structures. For example, the survey currently asks households with children to report children's trips on one assigned day, instead of all seven days. Large households could be asked to complete only one or two days of the smartphone app survey, whether or not they have children.

TABLE 20. PARTICIPANT HOUSEHOLD SIZE BY SURVEY YEAR (UNWEIGHTED).

HOUSEHOLD SIZE	2019	2021	2023	2015-2019 ACS
1 person	37.8%	40.8%	40.8%	26.2%
2 people	39.1%	36.9%	37.2%	32.2%
3 people	10.1%	10.8%	10.2%	13.9%
4 people	8.6%	7.5%	7.5%	12.8%
5 people	3.0%	2.8%	2.6%	5.5%
6 people	0.9%	0.9%	1.1%	2.0%
7 or more people	0.6%	0.4%	0.7%	7.5%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

In 2021 and 2023, the survey had a higher share of zero-vehicle households in the survey than are found in the region (Table 21). Although zero-vehicle households were not specifically recruited with differential incentives or special address-based sampling, the strategies to reach lower-income households likely played a role in recruiting a greater number of these households over time.

TABLE 21. PARTICIPANT HOUSEHOLD VEHICLES BY SURVEY YEAR (UNWEIGHTED).

HOUSEHOLD VEHICLES	2019	2021	2023	2015-2019 ACS
0 (no vehicles in household)	6.9%	11.3%	10.6%	6.5%
1 vehicle	39.1%	43.3%	42.8%	28.4%
2 vehicles	38.4%	34.9%	35.2%	38.7%
3 vehicles	10.7%	7.8%	8.3%	14.4%
4 vehicles	3.3%	1.9%	2.0%	4.4%
5 or more vehicles	1.7%	0.7%	1.0%	7.7%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

The oversample of zero-vehicle households is apparent even when broken down by household size (Table 22). For example, in 2023, 23.5% of one-person households in the survey had no vehicle available, compared to 17.4% of one-person households in the region. Low-vehicle households were similarly oversampled: for example, in 2021 and 2023, the share of four-or-more person households in the survey with only one vehicle available was nearly twice that expected from census estimates.

TABLE 22. PARTICIPANT HOUSEHOLD VEHICLES BY HOUSEHOLD SIZE AND SURVEY YEAR (UNWEIGHTED).

VEHICLE SUFFICIENCY	2019	2021	2023	2015-2019 ACS
Zero Vehicles	6.9%	11.3%	10.6%	6.9%
Insufficient	11.2%	14.9%	14.2%	33.5%
Sufficient	81.9%	73.8%	75.2%	59.6%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

In terms of age, the biggest discrepancy between the ACS and the survey data was that 65-74 year-olds were over-represented, while children 5-17 were underrepresented each year Table 23. The same factors that led to over-representation among small households likely contributed to this trend. Additionally, over-representation of retired and older adults in surveys is a common phenomenon outside of household travel surveys.

TABLE 23. PARTICIPANT AGE BY SURVEY YEAR (UNWEIGHTED)

AGE	2019	2021	2023	2015-2019 ACS
Under 5	5.8%	5.1%	5.2%	5.5%
5 to 17	13.2%	11.2%	11.8%	29.4%
18 to 24	3.8%	6.0%	5.4%	7.3%
25 to 34	13.2%	17.5%	16.0%	12.3%
35 to 44	14.3%	16.1%	15.7%	11.3%
45 to 54	11.5%	10.6%	12.2%	11.3%
55 to 64	16.9%	14.6%	13.0%	11.1%
65 to 74	14.4%	13.6%	14.7%	6.9%
75 or older	7.0%	5.2%	6.0%	4.8%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

The percentage of Asian participants went up each survey year, moving closer to the ACS percentage (Table 24).

TABLE 24. PARTICIPANT RACE BY SURVEY YEAR (UNWEIGHTED).

RACE	2019	2021	2023	2015-2019 ACS
American Indian or Alaska Native	0.4%	0.7%	0.7%	0.5%
Asian	3.2%	4.8%	5.1%	6.2%
Black or African American	2.5%	5.1%	4.2%	8.1%

Native Hawaiian or other Pacific Islander	<0.1%	<0.1%	0.2%	<0.1%
Two or more races	2.6%	3.1%	3.1%	3.2%
White	88.6%	83.1%	84.9%	77.7%
Other race, ethnicity, or origin	1.2%	1.2%	1.8%	4.2%
Don't know	1.4%	2.0%	NA	NA
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

The percentage of Hispanic or Latino participants increased each survey year, moving toward the ACS percentage (Table 25).

TABLE 25. PARTICIPANT ETHNICITY BY SURVEY YEAR (UNWEIGHTED).

ETHNICITY	2019	2021	2023	2015-2019 ACS
Hispanic or Latino	2.5%	3.2%	5.0%	19.9%
Not Hispanic or Latino	97.5%	96.8%	95.0%	80.1%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

The distribution of genders was fairly consistent between survey years. Women were slightly over-represented in the survey, consistent with a female bias in many types of surveys (Table 26).

TABLE 26. PARTICIPANT GENDER BY SURVEY YEAR (UNWEIGHTED), SELECTED CATEGORIES.

GENDER*	2019	2021	2023**	2015-2019 ACS
Female	52.8%	51.9%	51.9%	50.4%
Male	46.6%	46.6%	46.5%	49.6%
Transgender/Non-binary/Other/prefer to self-describe	0.6%	1.5%	1.6%	NA
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

*Respondents who identified as a gender other than Male or Female are excluded from this table.
**In 2023, the option for 'Female' was changed to 'Female/Woman/Trans woman/Girl', and 'Male' was updated to 'Male/Man/Trans man/Boy'

The data also show an increase in the percentage of respondents who decline to identify their gender or identify as something other than “Male” or “Female” (Table 27). In 2023, 1.2% of respondents identified as “A gender other than singularly male or female”, compared to 0.9% and 0.3% of respondents who identified as “Non-binary/third gender” in 2021 and 2019, respectively.

TABLE 27. PARTICIPANT GENDER BY SURVEY YEAR (UNWEIGHTED), ALL RESPONSE OPTIONS.

GENDER	2019	2021	2023
Female*	52.8%	51.9%	51.9%
Male*	46.6%	46.6%	46.5%
A gender other than singularly male or female (e.g., non-binary, genderfluid, agender, culturally specific gender)	---	---	1.3%
Non-binary/third gender	0.3%	0.9%	---
Transgender	0.2%	0.4%	---
Other/prefer to self-describe	0.1%	0.2%	0.3%
TOTAL	100.0%	100.0%	100.0%

*In 2023, the option for 'Female' was changed to 'Female/Woman/Trans woman/Girl', and 'Male' was updated to 'Male/Man/Trans man/Boy'

The distribution of employment was fairly consistent between survey years (Table 28). Unemployed persons are slightly over-represented in the survey relative to ACS percentages.

This is consistent with the over-representation of older adults in the survey, who are more likely to be retired than younger adults.

TABLE 28. EMPLOYMENT OF SURVEYED ADULTS (UNWEIGHTED).

EMPLOYMENT	2019	2021	2023	2015-2019 ACS
Employed	65.8%	67.0%	64.6%	91.8%
Unemployed	34.2%	33.0%	35.4%	8.2%
TOTAL	**100.0%**	**100.0%**	**100.0%**	**100.0%**

8.3 TRIP RATES

Consistent with declines in travel during the COVID-19 pandemic era, the overall trip rate in the region decreased by 32%, from 4.17 trips per person per day in 2019 to 2.82 trips per person per day in 2021 (Table 29). Trip rates rebounded somewhat in 2023, to 3.48 trips per person per day, but this was still about 17% below the pre-pandemic trip rate.

TABLE 29. UNWEIGHTED AND WEIGHTED TRIP RATE BY YEAR.

			UNWEIGHTED			**WEIGHTED**
Survey Year	Trips	Days	Trip Rate	Trips	Days	Trip Rate
2019	329,021	79,556	4.18 ± 0.02	15,348,964	3,682,918	4.17 ± 0.04
2021	176,093	49,567	3.48 ± 0.02	10,670,349	3,748,414	2.85 ± 0.04
2023	104,611	28,838	3.68 ± 0.03	13,038,925	3,750,006	3.48 ± 0.05
Error margins are +/- 1 standard error of the mean.						

Relative to low-income households, higher-income households saw the largest declines in trip rates from 2019 (Figure 10). This finding is consistent with the adoption of trip replacement behaviors like teleworking and online shopping that became more widespread during the pandemic, especially among high-income households with higher-salary office jobs. Before the pandemic, the highest-income households had the highest trip rates by a wide margin; after the pandemic, the margin had closed.

FIGURE 10. TRIP RATE BY HOUSEHOLD INCOME AND SURVEY YEAR.

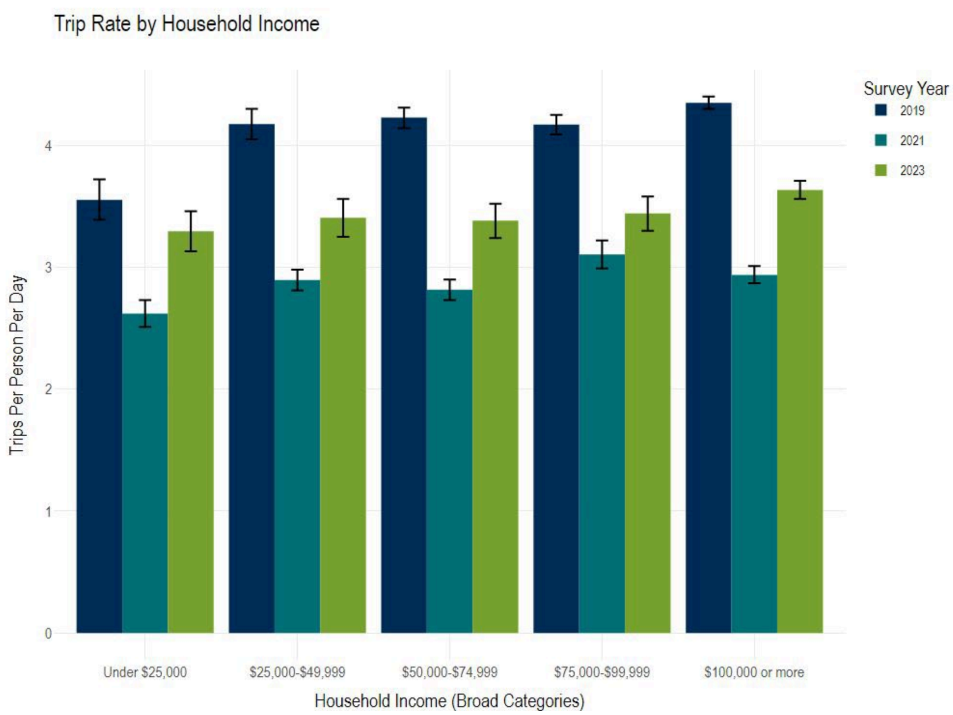


Figure 11 displays the average trip rate per person per day by trip departure hour, segmented into six time periods. The highest trip rates occurred in the evening peak hours (4:00 p.m.- 6:00 p.m.), with 2019 showing the peak at more than 1 trips per person per day. Trip rates were generally lower in the overnight hours (10:00 p.m.- 6:00 a.m.) and evening (7:00 p.m.- 9:00 p.m.).

Trip rates decreased in every time period from 2019 to 2021, but the decrease was most substantial for the evening peak hour (4:00 p.m.- 6:00 p.m.). From 2021 to 2023, trip rates rebounded in every time period.

FIGURE 11. TRIP RATE BY TIME OF DAY AND SURVEY YEAR.

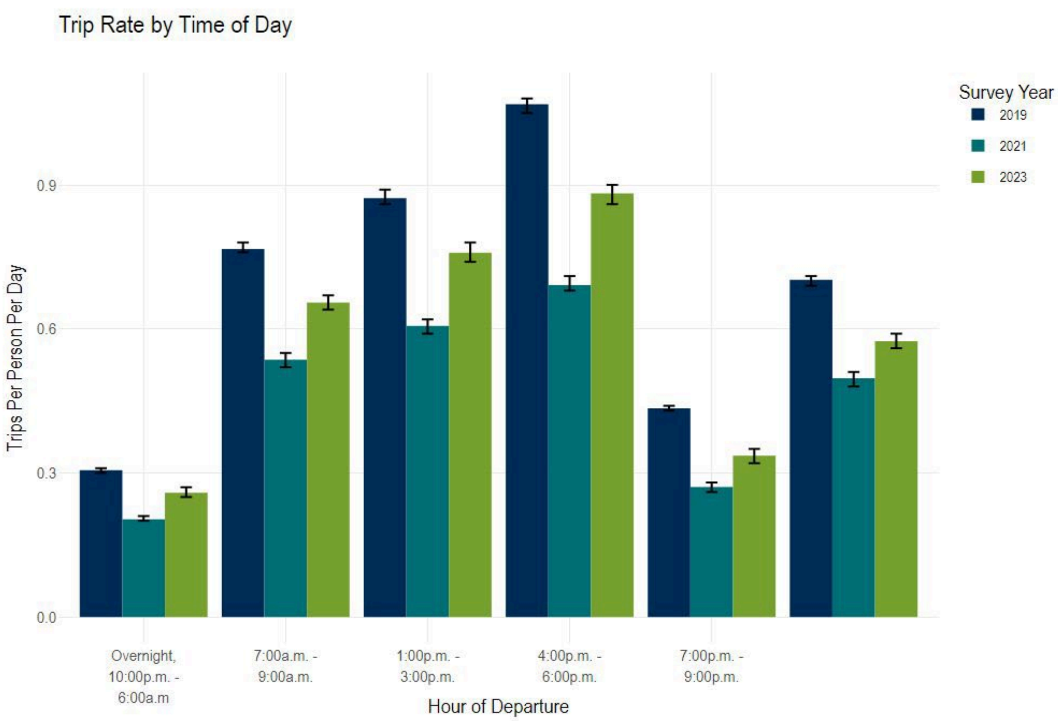
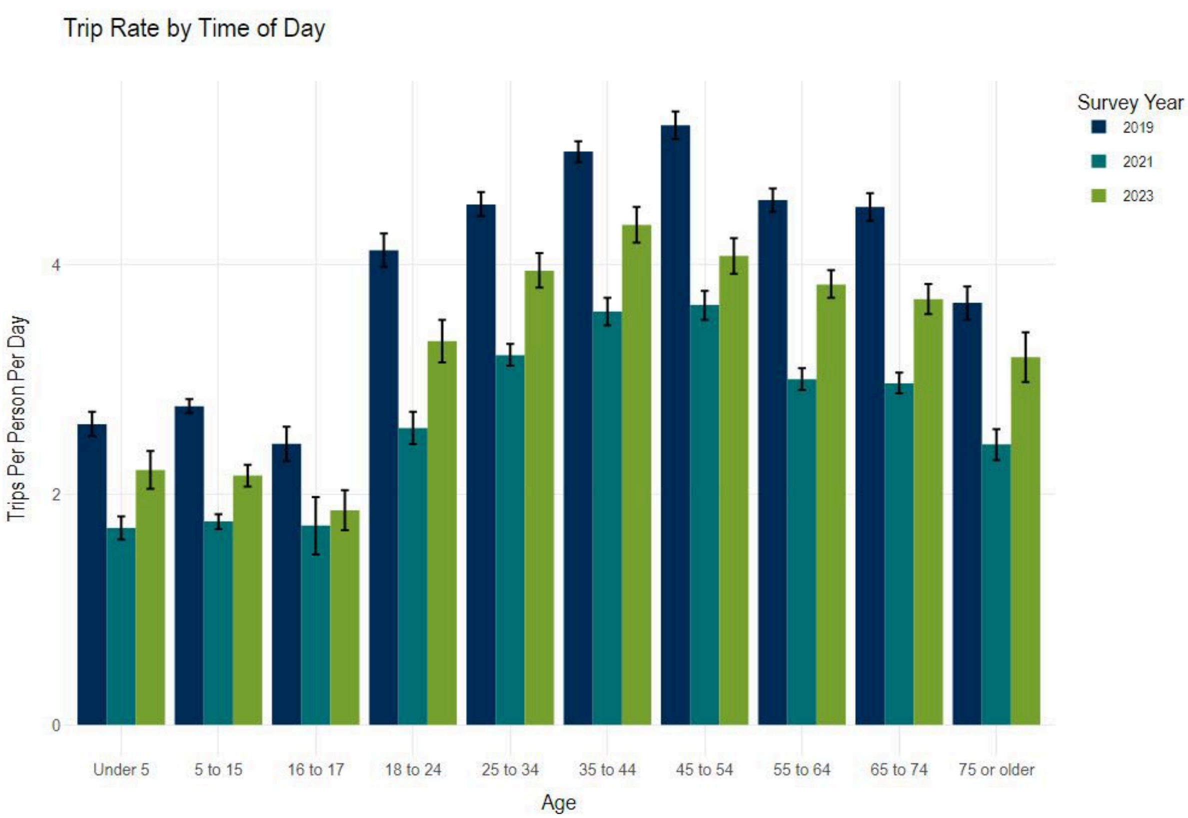


Figure 12 shows trip rates by the traveler's age and survey year. Across all three waves of the survey, middle-aged adults had the highest trip rates, consistent with a greater propensity for this group to make shopping, escort and errand trips. Children had the lowest trip rates, but these results should be interpreted with caution: even with advancements in proxy reporting methods and data cleaning, childrens' trip rates are historically under-reported and difficult to capture.

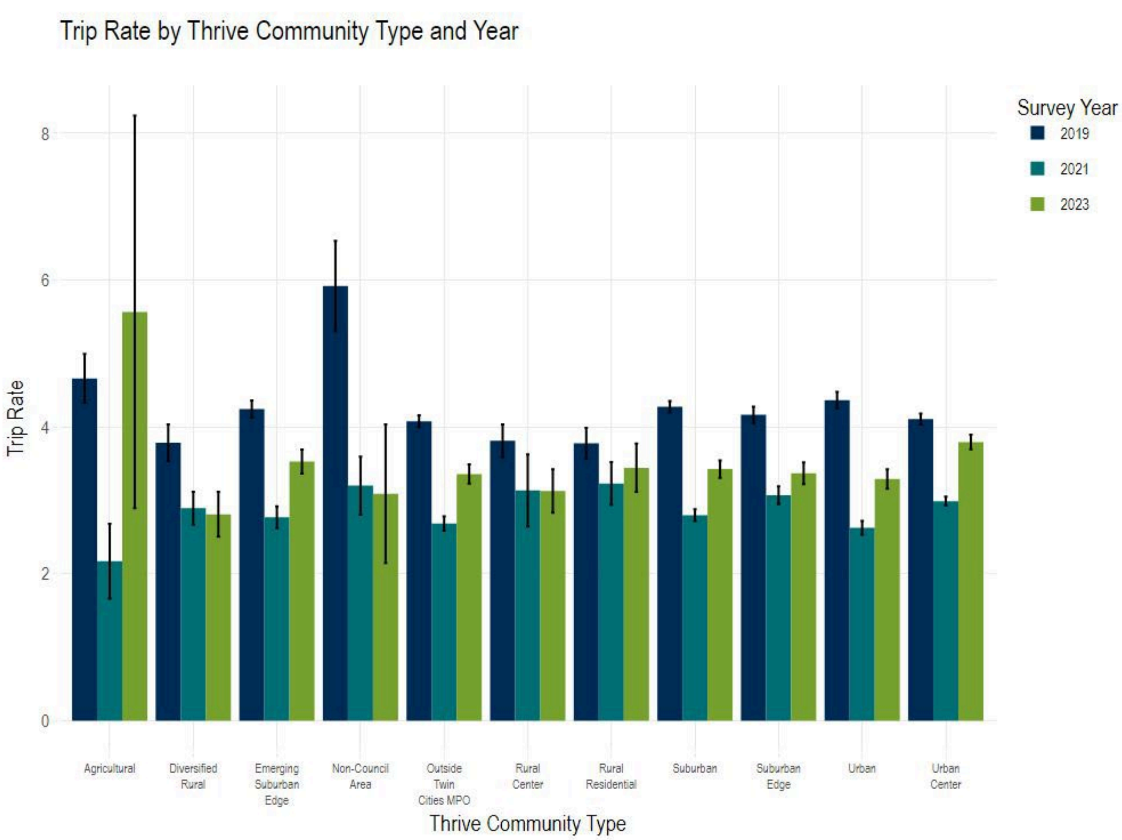
Trip rates declined for all age groups from 2019 to 2021 with the onset of the COVID-19 pandemic. From 2021 to 2023, trip rates rebounded for all age groups.

FIGURE 12. TRIP RATE BY AGE GROUP AND SURVEY YEAR.



Declines and rebounds in trip rates also varied by geography (Figure 13). A rebound in trip rates to near pre-pandemic levels was observed in Urban Center and Agricultural geographies, but Suburban, Suburban Edge and Urban geographies all exemplified consistent declines or a lack of a rebound.

FIGURE 13. TRIP RATE BY THRIVE COMMUNITY TYPE AND SURVEY YEAR.



8.4 MODE SHARE: HOW WE GET AROUND

Driving remains the predominant mode of travel in the region, representing 86% of trips in 2023 (Figure 14). Transit mode share declined from 2019 to 2021, then slightly rebounded from 2021 to 2023.

FIGURE 14. TRIP MODE TYPE BY SURVEY YEAR (WEIGHTED)

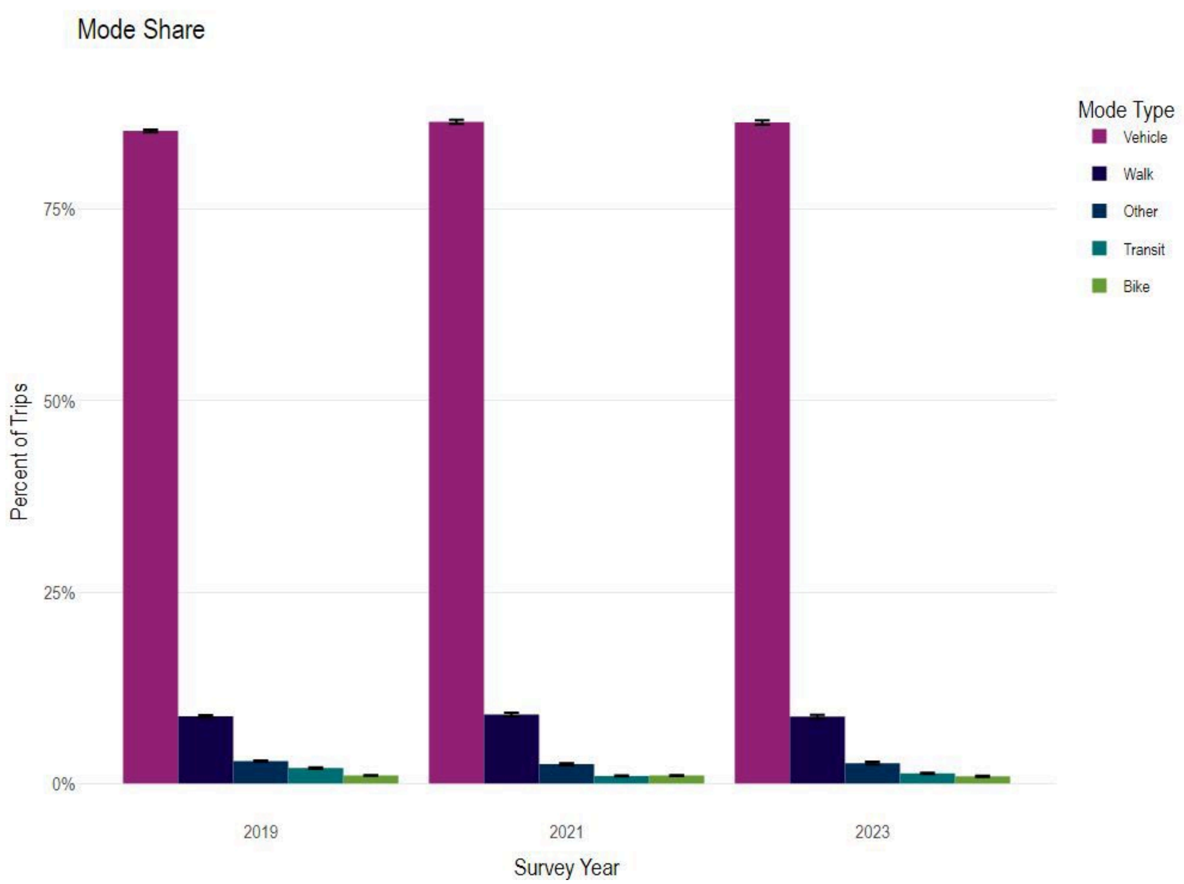


Figure 15 shows the share of people who drive, walk, take transit, bike, or use another mode on a typical weekday. This participation rate accounts for changes in the overall composition of travel (mode share) but also accounts for people who make no trips on any particular day, which has increased over time.

The share of people who drive on any given day has declined from 76% in 2019 to 70% in 2023. In 2023, roughly 1 in 10 people made a trip by walking, 1 in 40 made a trip using transit, and 1 in 100 made a trip by bicycle.

FIGURE 15. MODE PARTICIPATION RATES (WEIGHTED)

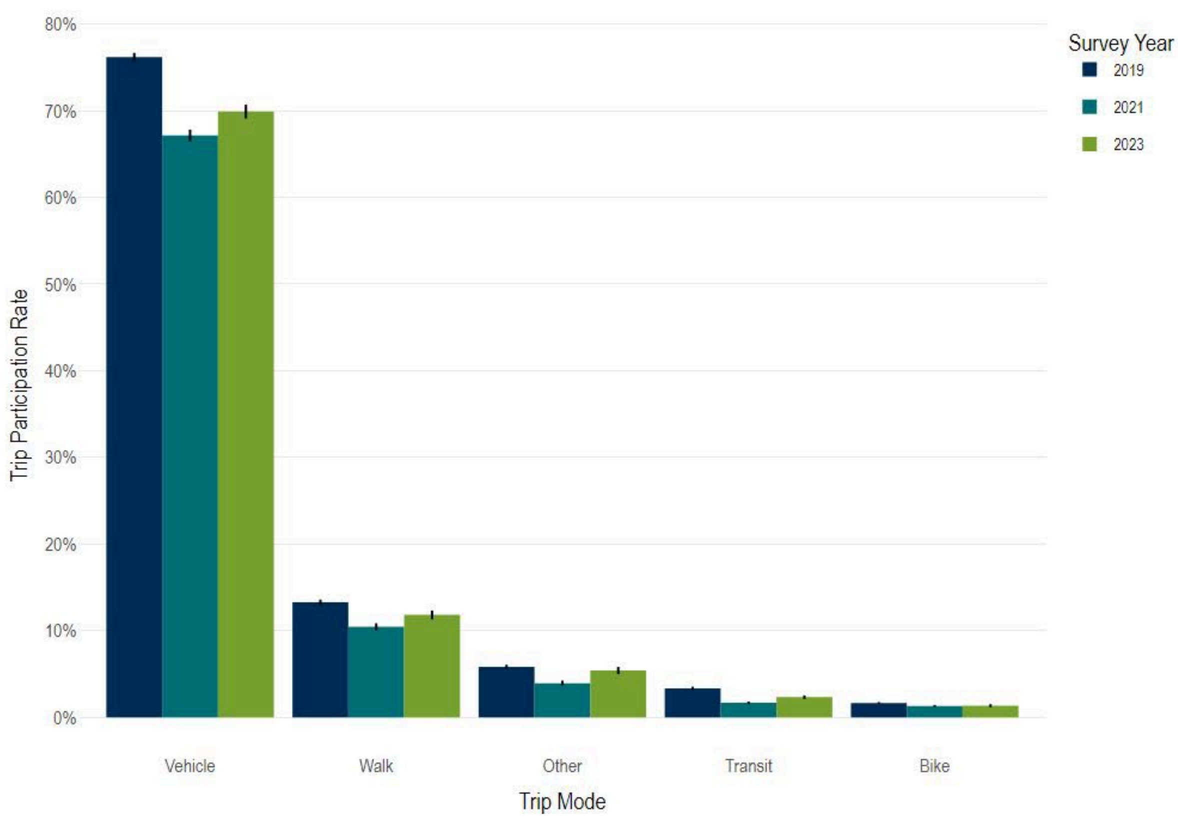


Figure 16 illustrates the mode share of trips by household income. The mode types include Vehicle, Walk, Transit, Micromobility, and Other. Across all income groups and years, the majority of trips are made by vehicle, followed by walking and transit. Micromobility and Other modes constitute a smaller proportion of trips. Vehicle use dominates in these groups throughout all years, with minimal changes in the proportion of trips made by walking and other modes. The share of trips by transit fell from 2019 to 2021 in most income groups, and then increased slightly in 2023 for most of the groups.

FIGURE 16. TRAVEL MODE BY HOUSEHOLD INCOME (WEIGHTED)

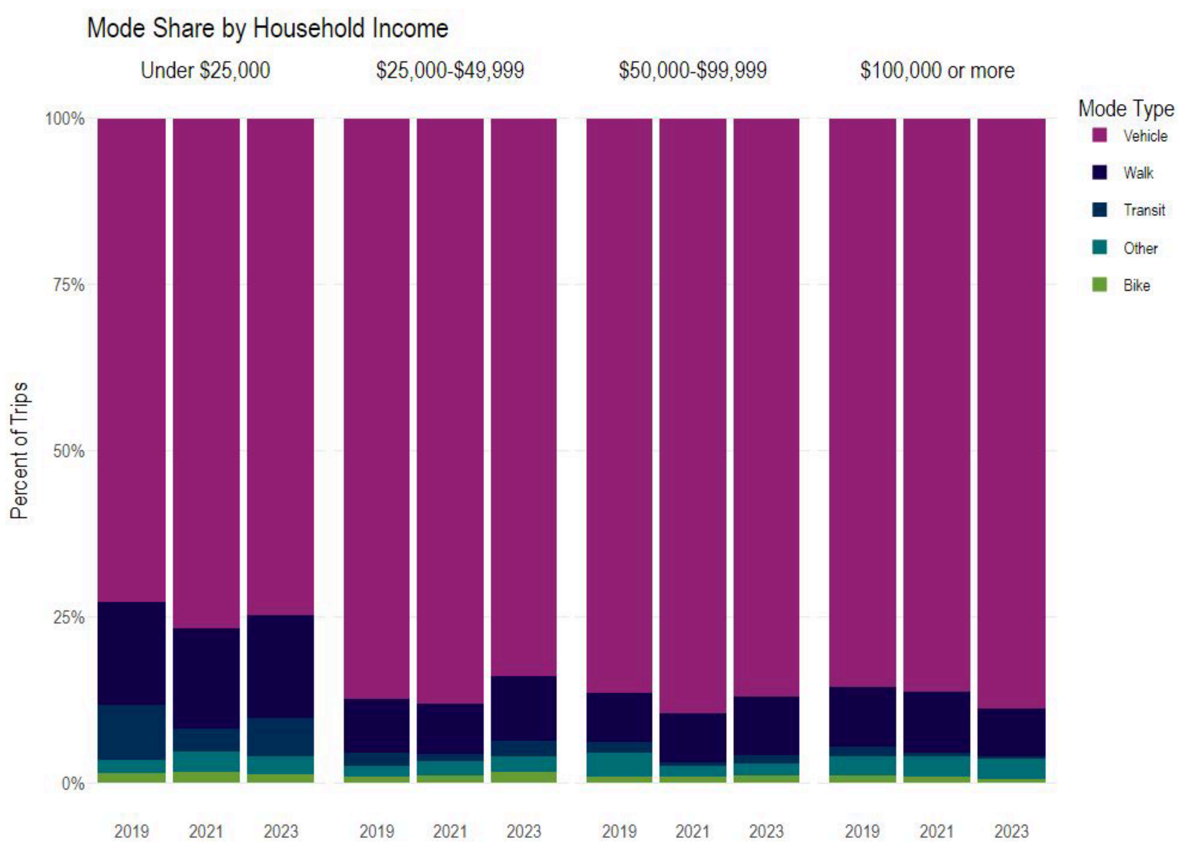
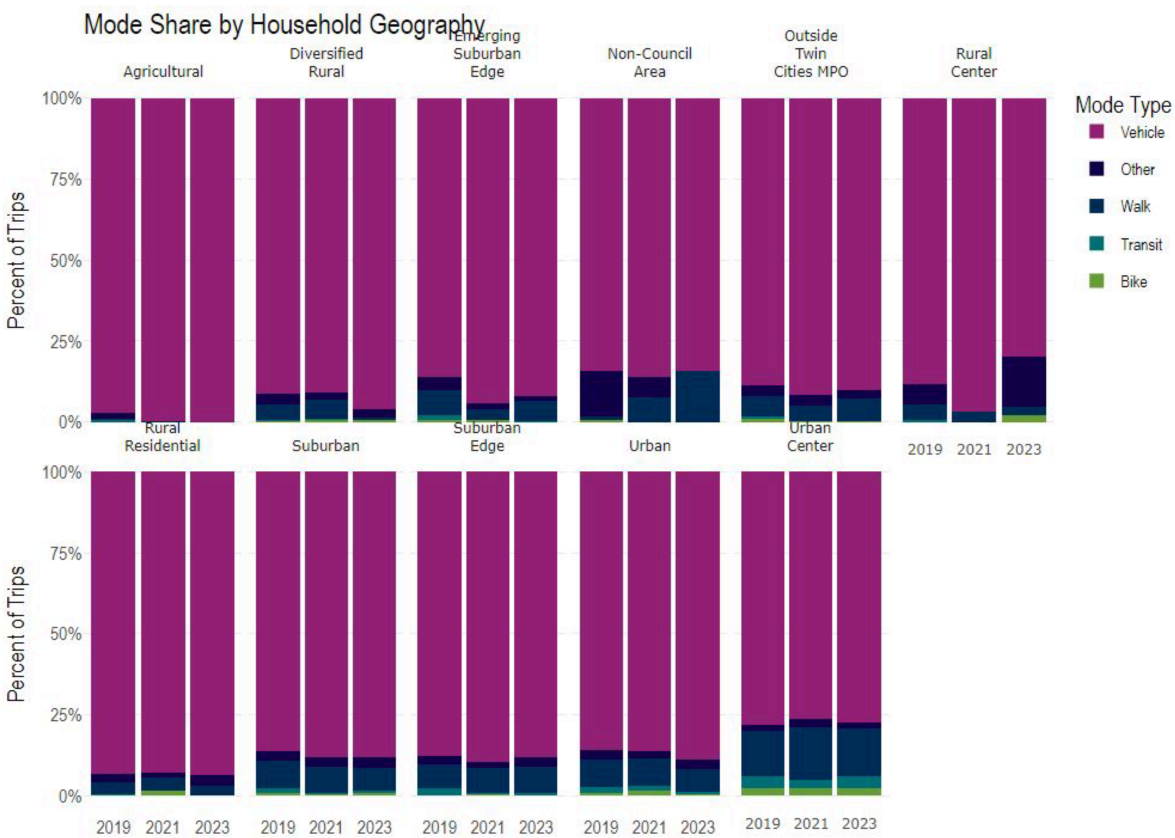


Figure 17 shows how mode choice varies across land use types in the Twin Cities region. Urban center households conduct a greater share of their travel using non-auto modes compared to rural and suburban households.

FIGURE 17. TRAVEL MODE BY THRIVE COMMUNITY TYPE (WEIGHTED)

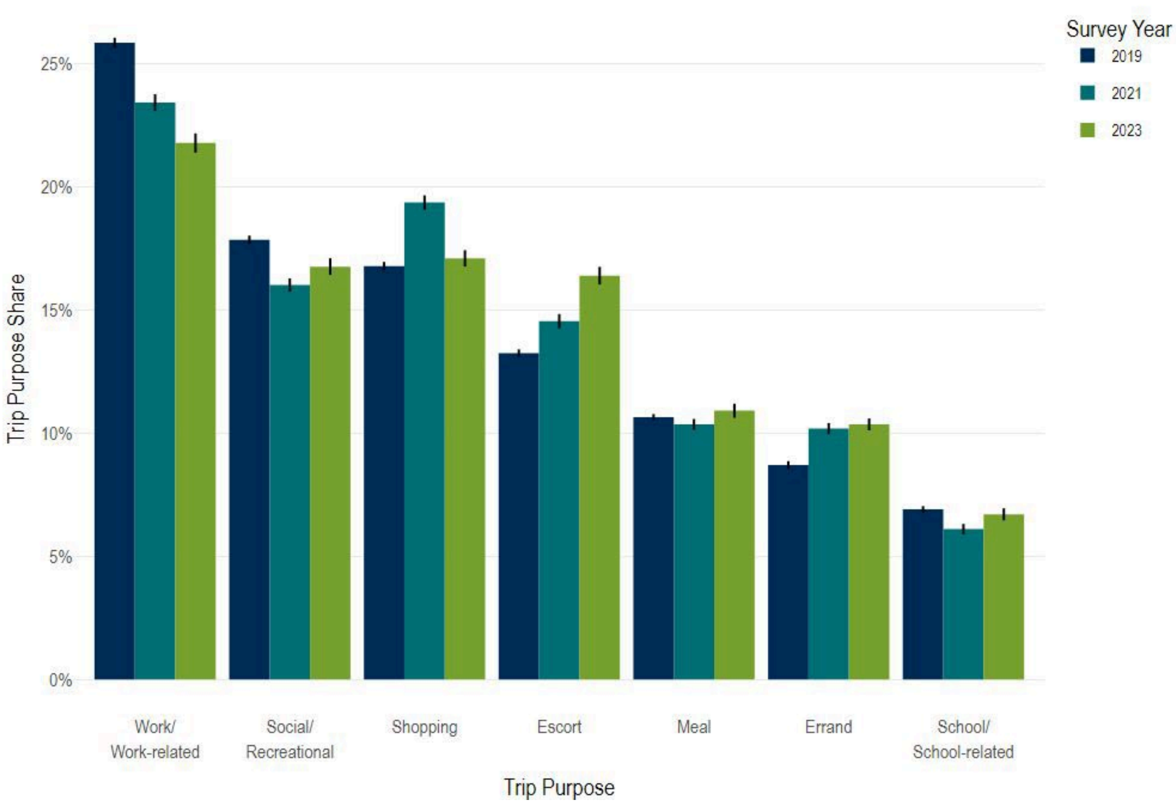


8.5 TRIP PURPOSE

Despite the extraordinary effects of the COVID-19 pandemic on daily life and work culture, the share of trips made by purpose stayed mostly consistent from 2019 to 2023 (Figure 18). Across all three years, the most common trip purpose was work or work-related, followed by social/recreational trips and shopping trips.

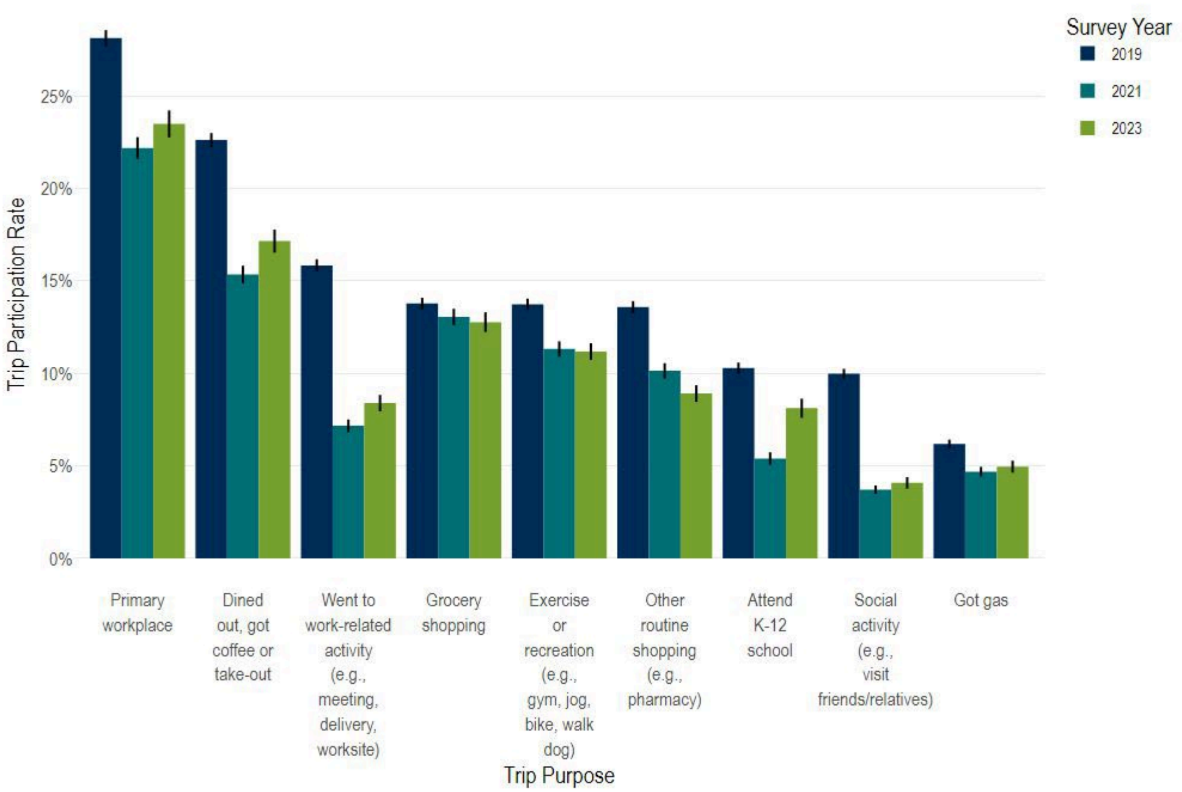
From 2019 to 2021, the share of trips made for work and work-related purposes declined from 26% of trips to 23% of trips, then declined again slightly to 22% of trips in 2023. Conversely, the share of trips made for shopping and errands increased from 2019 to 2021, then declined from 2021 to 2023.

FIGURE 18. SHARE OF TRIPS BY PURPOSE AND SURVEY YEAR (WEIGHTED)



Comparisons of the share of trip purpose by year must be made carefully, because overall trip rates declined during the COVID-19 pandemic and did not fully recover (Table 4). An alternative measure, the participation rate, can be used instead: this measure represents the share of people who make any given type of trip on a typical day (Figure 19), and as such accounts for both changes in the types of trips people make and the changes in the amount of travel people make.

FIGURE 19. TRIP PURPOSE PARTICIPATION RATE BY YEAR (SELECTED PURPOSES, DETAILED CATEGORIES; WEIGHTED)



The trends in Figure 19 suggest a number of underlying processes at work.

The effects of telework on travel behavior are easy to spot: the share of people making work and work-related trips declined from 2019 to 2021, then slightly rebounded from 2021 to 2023. Work-related trips (to meetings, deliveries, worksites) took a greater hit than trips to and from primary workplaces. From 2019 to 2023, the share of people making trips to work-related locations fell from 16% to 8%. Teleconferencing could be replacing trips that used to be made in person.

The fingerprint of e-commerce is also visible in the data. For example, the percent of people making a routine shopping trip on a typical day fell from 14% in 2019 to 10% in 2021, and then to 9% in 2023.

Finally, the lingering effects of the pandemic on social life are also apparent. The share of people making a social visit trip more than halved from 2019 to 2021, then rebounded only slightly in 2023. Similarly, the share of people making trips to exercise or walk outside declined from 2019 to 2021 and again from 2021 to 2023.

8.6 VEHICLE OWNERSHIP AND USE

A key measure of vehicle use is vehicle miles traveled, or VMT. VMT is calculated by dividing the number of miles traveled in a car trip by the number of occupants in the car. Table 30shows that VMT decreased substantially from 2019 to 2021, then rebounded from 2021 to 2023. In 2023, individuals drove 17 miles per day. Households accrued 42 miles of vehicle travel per day on average in 2023.

TABLE 30. VEHICLE MILES TRAVELED BY YEAR (WEIGHTED).

	HOUSEHOLD-LEVEL VMT		PERSON-LEVEL VMT	
Survey Year	**Average**	**Standard Error**	**Average**	**Standard Error**
2019	54.9	0.9	22.2	0.3
2021	37.2	1.6	15.1	0.5
2023	42.0	1.5	17.0	0.5

Figure 21shows Vehicle Miles Traveled (VMT) per person per day, segmented by gender, age group, and survey year (2019, 2021, and 2023). For all age groups, males generally accrue more VMT per day than females. The 35-64 age group shows the highest VMT per day for both genders across all survey years. The VMT per day decreased across both genders from 2019 to 2021. The VMT per day then increased slightly for each gender and age group from 2021 to 2023, with the exception of the 18-34 age group who saw little change in VMT per day from 2021 to 2023.

These trends hint at a host of socioeconomic, social, cultural, and behavioral factors at play. For example, men may be more likely to work in jobs that require longer commutes or have jobs that require travel for work (driver, sales); women may be less likely to have access to the household vehicle in households led by heterosexual couples; and men may have higher incomes than women, allowing for more discretionary travel.

FIGURE 20. VEHICLE MILES TRAVELED BY AGE, GENDER AND SURVEY YEAR (WEIGHTED)

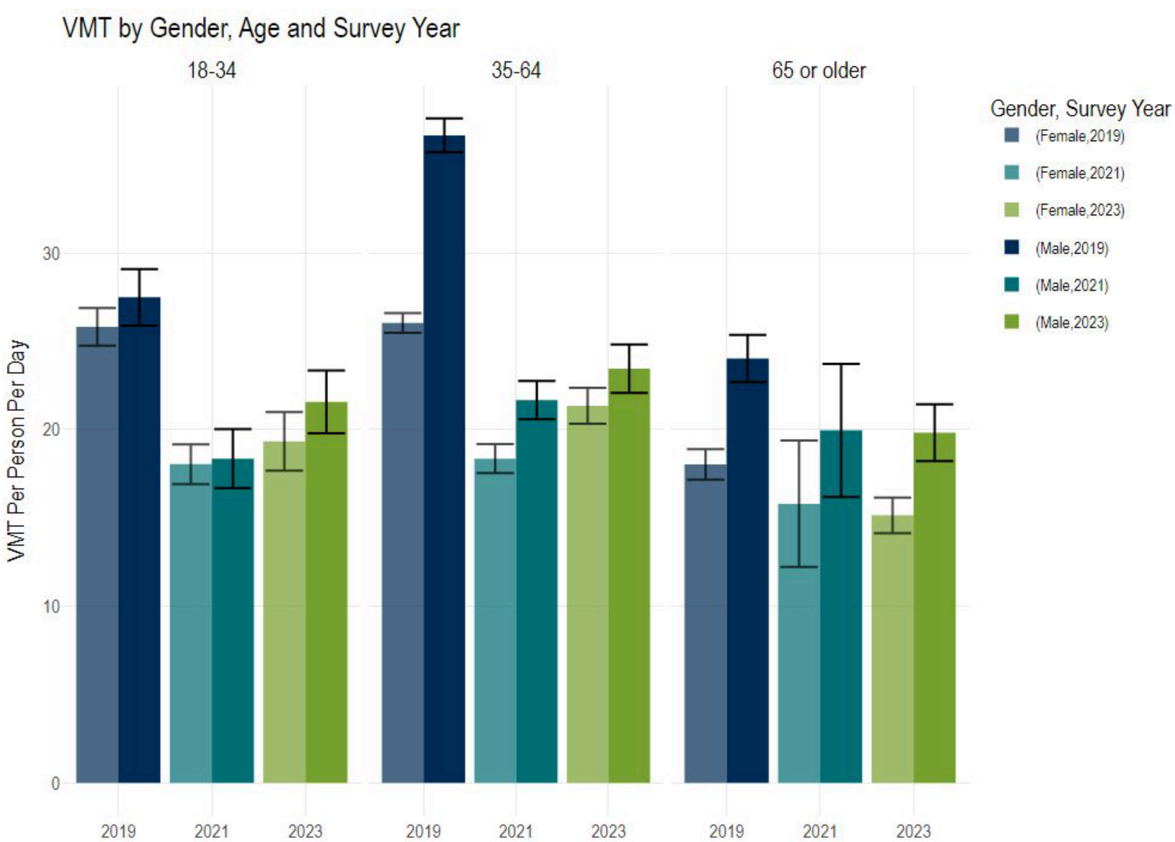


Figure 22 shows how household-level VMT varies according to household income. Across all years, higher-income households accrue more VMT than people in lower-income households. Households earning $100,000 or more saw the greatest declines in VMT from 2019 to 2023, likely owing to the propensity of this income group to adopt working from home during the pandemic.

FIGURE 21. VEHICLE MILES TRAVELED BY INCOME (WEIGHTED)

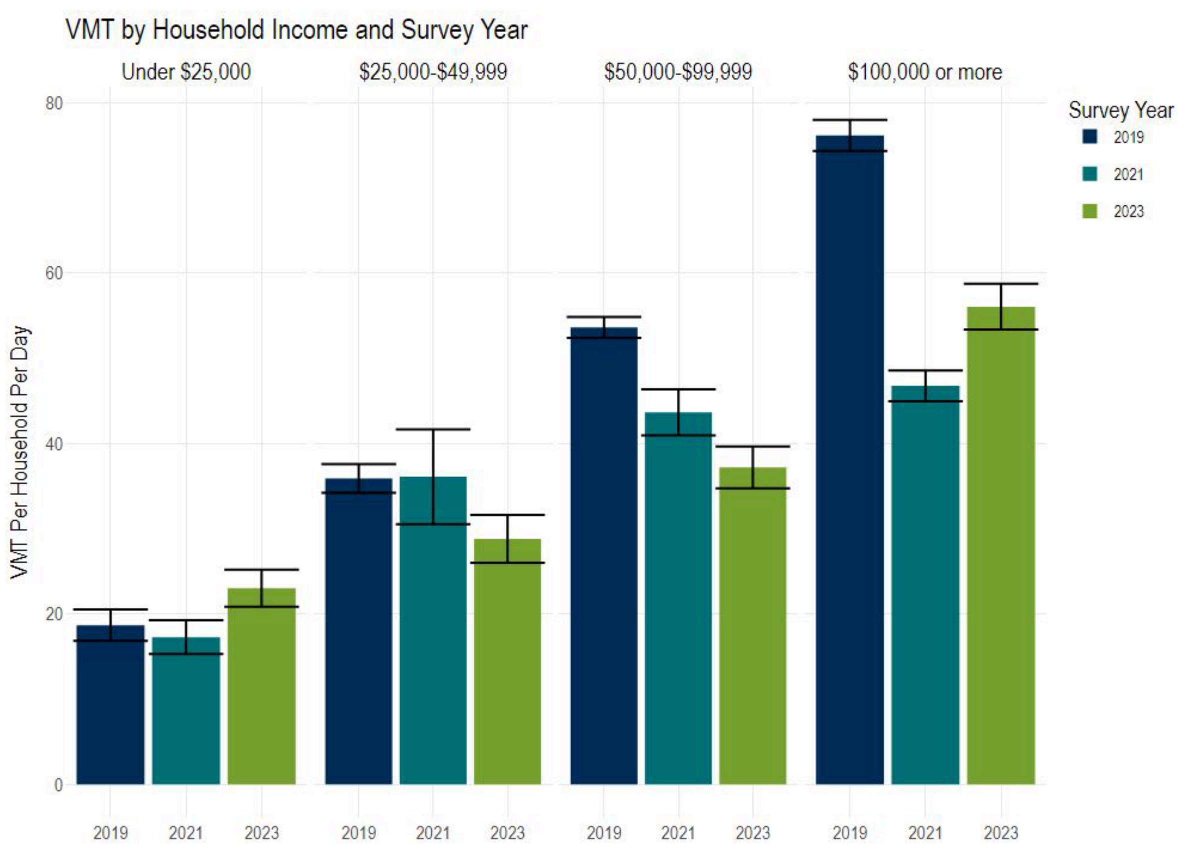


Figure 23 shows how VMT varies by job type. Those who drive for work or have a job with multiple job sites accrue the greatest VMT, and those who work only from home accrue the least. In 2023, hybrid workers had similar VMT (23 miles traveled) as those who work only one location outside of the home (24 miles traveled).

FIGURE 22. VEHICLE MILES TRAVELED BY JOB TYPE (WEIGHTED)

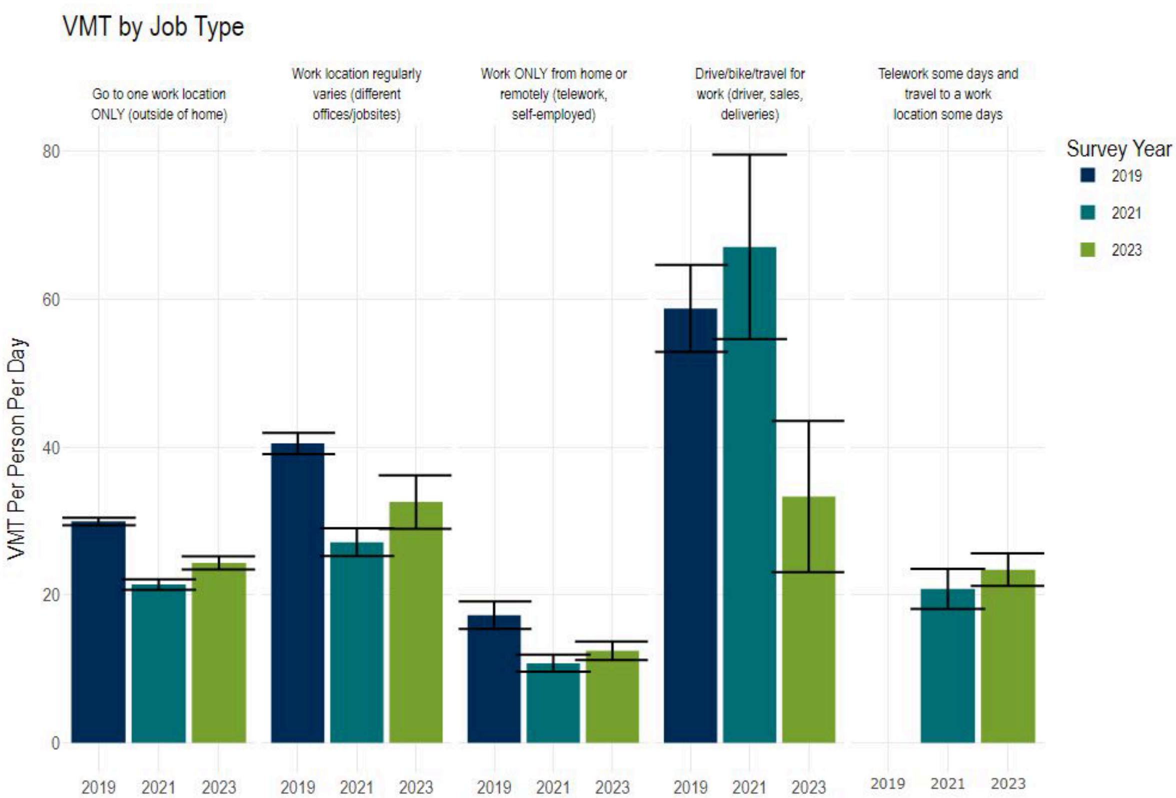
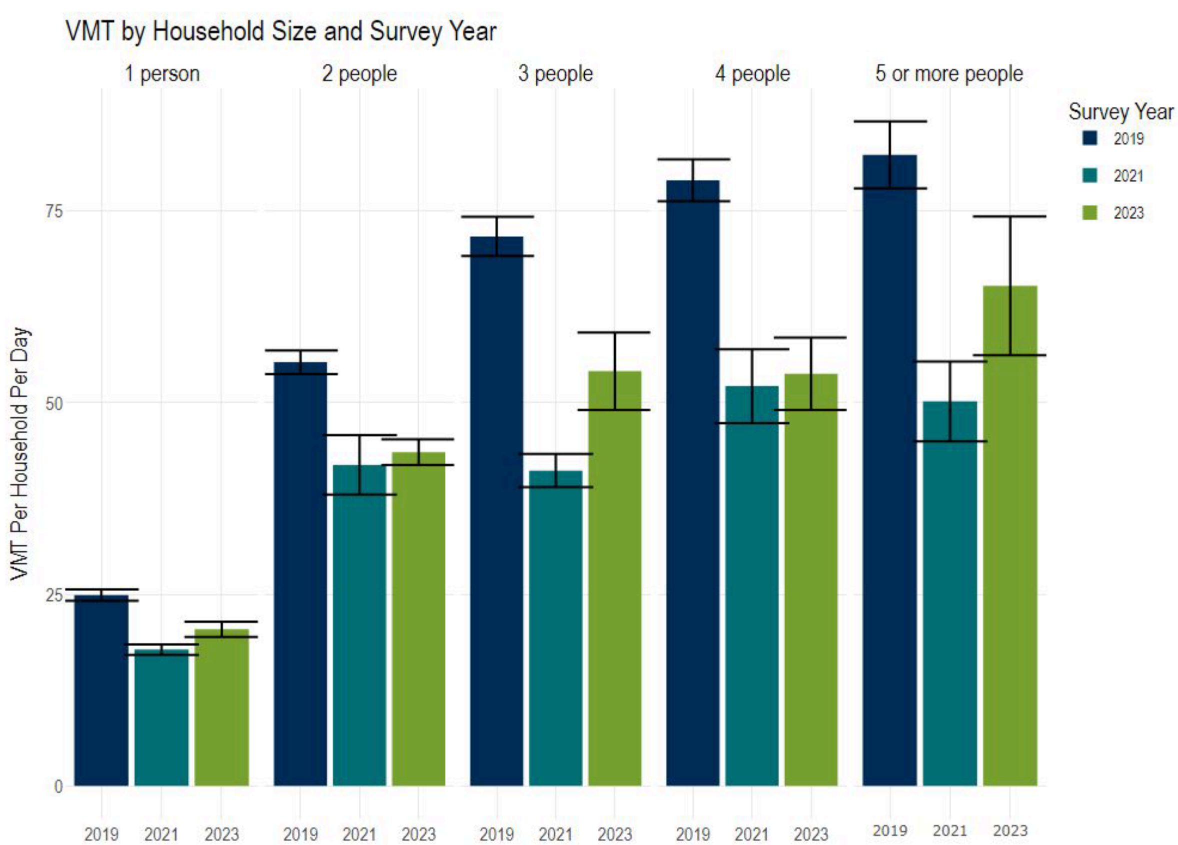


Figure 24 shows how VMT at the household level varies by household size. Instead of doubling for every person added, VMT increases only slightly as household size increases past 3 or more people.

Larger households saw the greatest decreases in VMT after the pandemic compared to smaller households (1 or 2 persons).

FIGURE 23. VEHICLE MILES TRAVELED BY HOUSEHOLD SIZE (WEIGHTED)



In all 3 waves, about half of vehicle trips ended at a parking lot, ramp, or garage (Table 31). However, the share of trips ending at a parking lot decreased from 2019 to 2021, consistent with decreases in commutes to work; this rebounded slightly from 2021 to 2023. Meanwhile, a greater share of trips ended at home or were drive-through trips.

TABLE 31. VEHICLE PARK LOCATION BY YEAR (WEIGHTED)

VEHICLE PARK LOCATION	2019	2021	2023
Home driveway/garage (yours or someone else's)	26%	33%	34%
Someone else's driveway	4%	NA	NA
Parking lot/ramp/garage	53%	46%	48%
On-street parking	10%	12%	9%
Park & Ride lot	0%	0%	0%
Didn't park (waited, drop-off, drive-thru, gas)	6%	9%	8%
Other	1%	1%	1%
TOTAL	**100%**	**100%**	**100%**

8.7 MICROMOBILITY

Participation in shared mobility services rose dramatically from 2019 to 2023 (Table 32). Each shared mobility service option besides vanpool was used more in 2023 than in prior waves. This was especially true for Uber, Lyft and other TNCs: in 2019, about 29% of residents said they used a TNC service on a regular basis; by 2023, 36% of residents said they used TNC services regularly. Participation in scooter-shares and carshares, while small overall, also rose dramatically over time.

TABLE 32. SHARED MOBILITY SERVICE USE BY YEAR (WEIGHTED).

	PERCENT OF PEOPLE USING SERVICE (WEIGHTED)		
Shared Mobility Service	**2019**	**2021**	**2023**
Bikeshare or bike rental service	1.7%	1.2%	1.9%
Carshare (e.g., Zipcar)	0.4%	0.2%	0.9%
Electric vehicle carshare (e.g., Evie)	--	--	0.6%
Moped share (e.g., Scoot)	--	0.1%	0.1%
None of the above	69.8%	79.2%	62.7%
Peer-to-peer car rental (e.g., Turo)	0.1%	0.6%	1.4%
Scooter share (e.g., Bird, Lime)	--	2.1%	3.5%
SouthWest Prime or MVTA Connect	--	0.5%	1.2%
Uber, Lyft, or other smartphone-app ride service	29.2%	19.4%	35.8%
Vanpool	0.3%	0.2%	0.2%

A multiple-response variable; participants could select more than one option. Totals add to more than 100%.

The majority of households own at least one bicycle, with roughly half owning two or more bicycles. The share of households with no bicycles increased slightly from 2021 to 2023 (Table 33).

TABLE 33. NUMBER OF BICYCLES OWNED BY HOUSEHOLD (WEIGHTED).

NUMBER OF BICYCLES OWNED	2021	2023
0	28.1%	29.2%
1	20.0%	19.7%
2	21.4%	21.0%
3	11.0%	10.4%
4	9.8%	10.4%
5	4.9%	4.2%
6	2.3%	2.9%
7	1.2%	0.6%
8 or more	1.4%	1.7%
TOTAL	**100.0%**	**100.0%**

Of those who owned bicycles, the share of those who owned an electric bicycle doubled from 2021 to 2023 (Table 34). In 2023, roughly one in every twelve bike owners was also an e-bike owner.

TABLE 34. BICYCLE OWNERSHIP BY BIKE TYPE AND SURVEY YEAR (WEIGHTED).

TYPE OF BICYCLE OWNED	2021	2023
Electric bicycle	3.7%	7.2%
Other	2.3%	3.0%
Standard bicycle	98.6%	97.7%
A multiple-response variable; participants could select more than one option. Totals add to more than 100%.		

The share of bicycle trips that ended at non-home locations decreased from 2019 to 2023 (Table 35). The dropoff was especially steep for trips ending at a secured bike room during the height of the COVID-19 pandemic in 2021.

TABLE 35. BICYCLE PARK LOCATION SURVEY YEAR (WEIGHTED).

BICYCLE PARK LOCATION	2019	2021	2023
Inside house/apartment (includes garage, porch, storage area)	24.8%	41.3%	41.6%
Bike rack	31.4%	23.7%	23.9%
Bike locker	1.6%	1.0%	0.3%
Secured bike room	7.5%	1.7%	4.3%
Locked to other object (e.g., post, tree)	7.6%	11.7%	7.4%
Bike-share designated docking station	1.4%	1.2%	1.9%
Unlocked on-street	6.0%	2.9%	3.2%
In a parking garage/ramp/lot	NA	1.9%	3.3%
Carried it with me	NA	7.5%	10.9%
Other	19.8%	7.2%	3.3%
TOTAL	**100.0%**	**100.0%**	**100.0%**

Table 36 shows the places people stored their bicycles. Most people reported storing their bicycles inside their home or garage, with about 6% (in 2023) reporting that they stored at a bike rack. Only 2.9% of bike owners stored their bike in a designated indoor facility, such as a bike room or bike locker.

TABLE 36. BICYCLE STORAGE LOCATION BY SURVEY YEAR (WEIGHTED)

BICYCLE STORAGE LOCATION	2021	2023
Inside house/apartment (includes garage, porch, storage area)	91.1%	92.7%
Bike rack	4.5%	6.0%
Bike locker	0.6%	0.3%
Secured bike room	2.1%	2.6%
Locked to other object (e.g., post, tree)	2.3%	3.5%
In a parking garage/ramp/lot	5.1%	4.7%
Unlocked on-street	0.3%	0.2%
Other	2.3%	2.5%

A multiple-response variable; participants could select more than one option. Totals add to more than 100%.

The percent of households who owned other micro-mobility devices, such as scooters and mopeds, stayed consistent from 2021 to 2023 (Table 37). In 2023, 11.8% of households owned scooters, with 9-12.9% owning skateboards or roller blades.

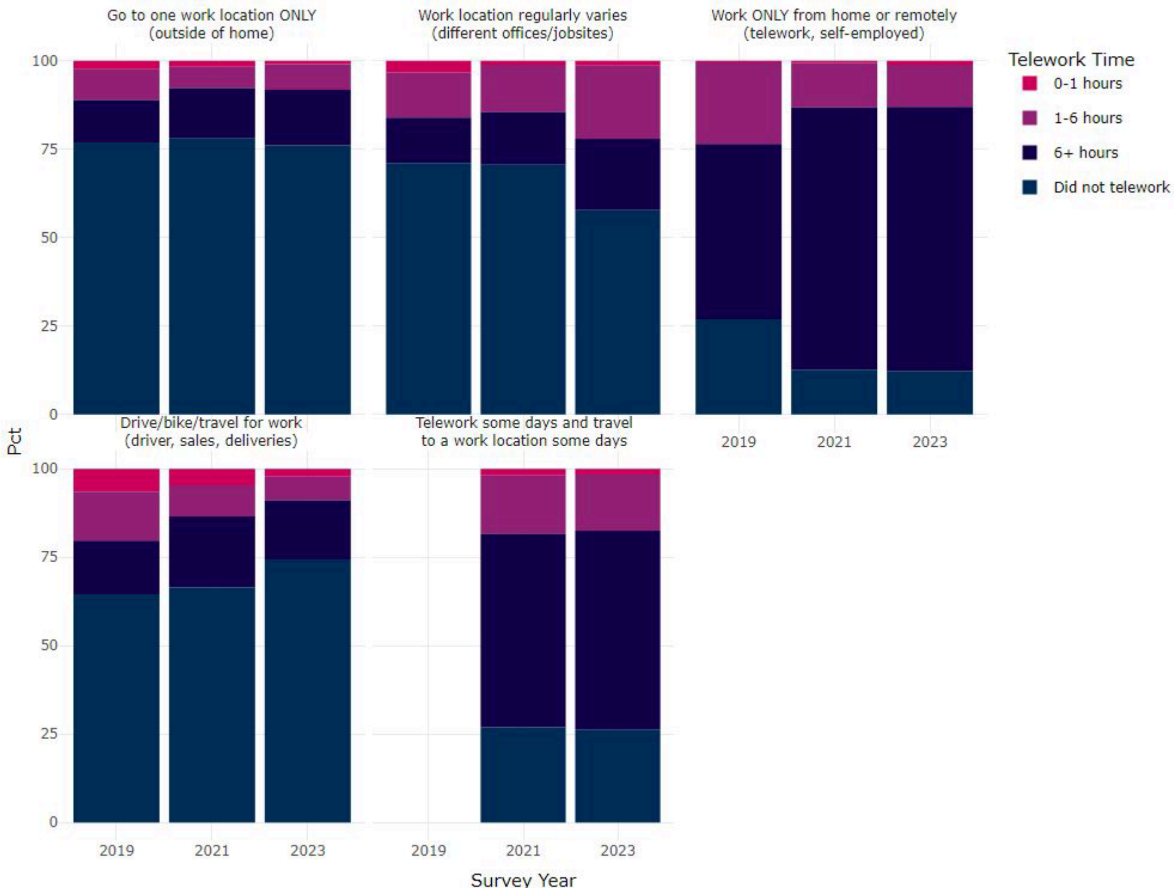
TABLE 37. MICROMOBILITY DEVICE OWNERSHIP BY SURVEY YEAR (WEIGHTED)

	PERCENT OF HOUSEHOLDS OWNING DEVICE (WEIGHTED)	
Micro-Mobility Device	2021	2023
Moped	0.7%	1.2%
None	83.5%	86.0%
Other	4.2%	1.3%
Scooter	11.4%	11.8%
Segway	0.6%	0.3%
Skateboard or rollerblades	12.5%	12.9%
A multiple-response variable; participants could select more than one option. Totals add to more than 100%.		

8.8 TELEWORK AND TELECOMMUTING

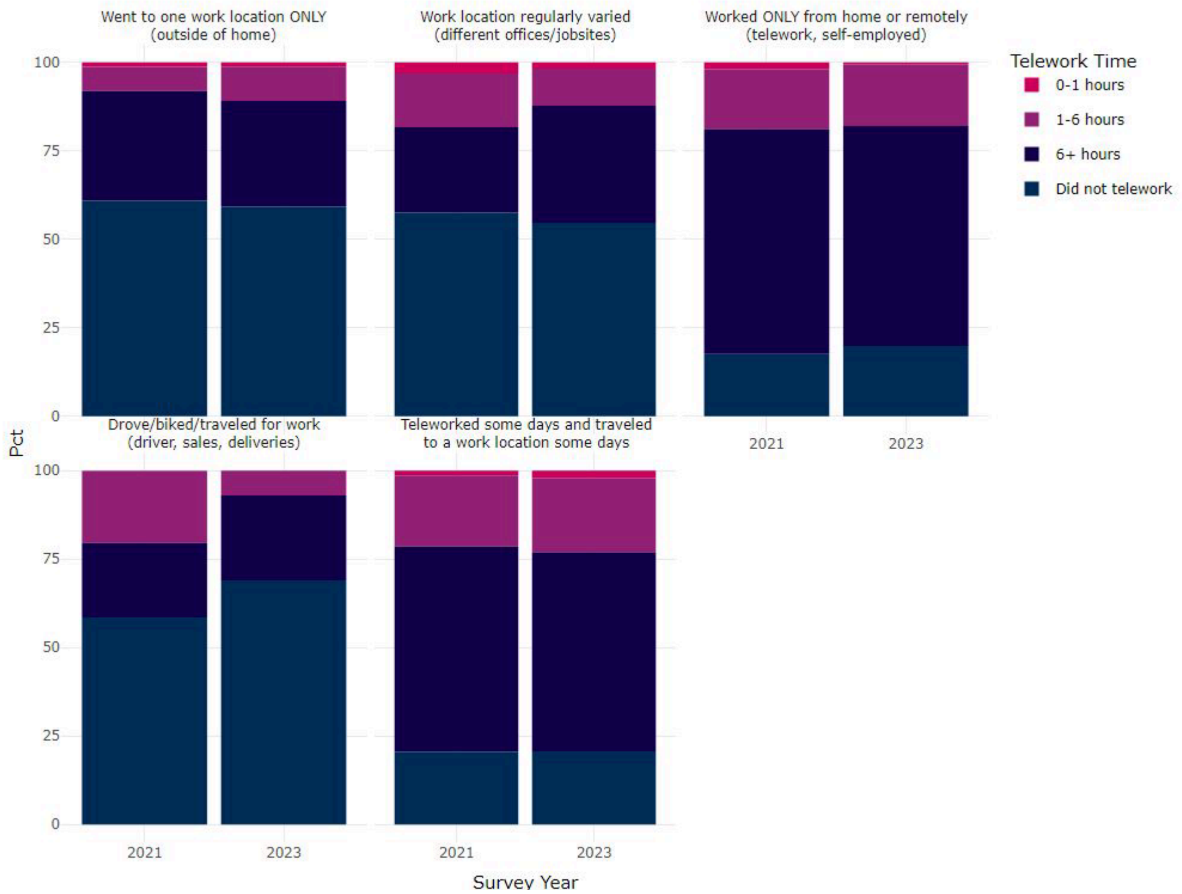
Participants who said they teleworked some days and traveled to work other days teleworked 6+ hours over half of their travel days in both 2021 and 2023 (Figure 24).

FIGURE 24. SUMMARY OF TELEWORK TIME, BY JOB TYPE (EMPLOYED ADULTS, WEIGHTED).



The distribution in telework time was fairly consistent across job types pre-COVID (Figure 25).

FIGURE 25. TELEWORK TIME BY JOB TYPE, PRE-COVID (EMPLOYED ADULTS, WEIGHTED).



The percent of people receiving a delivery on a typical weekday peaked in 2021 (Table 40). Roughly one in five residents receives a package delivery on a typical weekday.

TABLE 38. DELIVERY FREQUENCY (WEIGHTED).

HOME DELIVERIES	2019	2021	2023
Take-out/prepared food delivered to home	0.9%	1.9%	1.6%
Someone came to do work at home (e.g., babysitter, housecleaning, lawn)	1.3%	1.4%	0.9%
Groceries delivered to home	NA	1.0%	0.6%
Received packages at home (e.g., USPS, FedEx, UPS)	15.4%	22.8%	23.0%
Received personal packages at work	1.4%	0.3%	0.2%
Received packages at another location (e.g., Amazon Locker, package pick-up point)	0.2%	0.6%	0.8%
Other item delivered to home (e.g., appliance)	0.2%	0.2%	0.1%
None of the above	79.7%	68.2%	69.6%
Deliveries and home services -- Multiple deliveries	0.9%	3.4%	3.1%
TOTAL	**100.0%**	**100.0%**	**100.0%**

The percent of people who had free/discount transit fare available to them as a commute subsidy decreased in each wave (Table 41).

TABLE 39. COMMUTE SUBSIDY AVAILABILITY (WEIGHTED)

COMMUTE SUBSIDY	2019	2021	2023
Free/discount transit fare	12.5%	9.4%	8.5%
Stipend for working at home (e.g., internet, equipment)	--	--	3.5%
Free/discount vanpool	2.7%	0.9%	0.9%
Free/discount Uber, Lyft, or other smartphone-app ride service	1.4%	0.3%	0.6%
Cash or incentives for carpooling, walking, or biking to work	3.5%	1.3%	1.0%
Free/discount bikeshare membership	1.2%	0.4%	0.2%
Free/discount bicycle tune-up/maintenance	1.0%	0.3%	0.2%
Free/discount carshare membership/use (e.g., Zipcar)	0.6%	0.2%	0.0%
Free/discount shuttle service	2.2%	1.0%	1.0%
Free/subsidized parking pass	--	7.8%	--
None of the above	84.0%	73.6%	74.3%
Don't know	--	11.2%	12.5%
A multiple-response variable; participants could select more than one option. Totals add to more than 100%.			

APPENDIX A. DATA USER'S GUIDE (R)

Getting Started

This data user's guide relies on the combined, upcoded 2019-2023 dataset. In the code chunk below, the raw data are labeled using a convenience function, factorize_df, from the open-source R package travelSurveyTools (link here). The function takes as input the value labels from the combined 2019-2023 codebook. See section *About the Survey: Dataset Preparation and QAQC* for more information.

Like the remainder of this Synthesis report, this user's guide also includes data for only weighted households, to exclude those households that did not meet the criteria for weighting during the 2019-2021 re-weighting process, and also those 2023 outreach households that did not meet initial completion criteria for weighting (See Section *About the Survey: Expansion and Weighting* for more information).

To filter unweighted households from all tables in the Travel Behavior Inventory dataset, users can run the following commands in R:

```
weighted_hhs <- tbi$hh[!is.na(hh_weight) & hh_weight > 0, hh_id]

tbi <-
  lapply(
    tbi,
    function(dt) {
      dt <- dt[hh_id %in% weighted_hhs]
    }
  )
```

The resulting dataset, tbi, is a list of seven tables in data.table format.

```
nrows <- lapply(tbi, nrow) %>%
  stack(.) %>%
  rename(., "Rows" = "values") %>%
  rename(., "Table" = "ind")

nrows %>%
  select(Table, Rows) %>%
```

```
  gt() %>%
  gt::fmt_number(columns = c(2), decimals = 0)
```

TABLE	ROWS
hh	19,172
person	38,693
day	157,961
trip	623,926
vehicle	30,240
trip_purpose	844,728
linked_trip	609,725

Basic Operations

Summarizing SRCVs: Single Response Categorical Variables

A single-response categorical variable is a column in the dataset containing answers for a question for which the survey respondent could select only one option from a list of pre-populated answers. For this example, use income_broad.

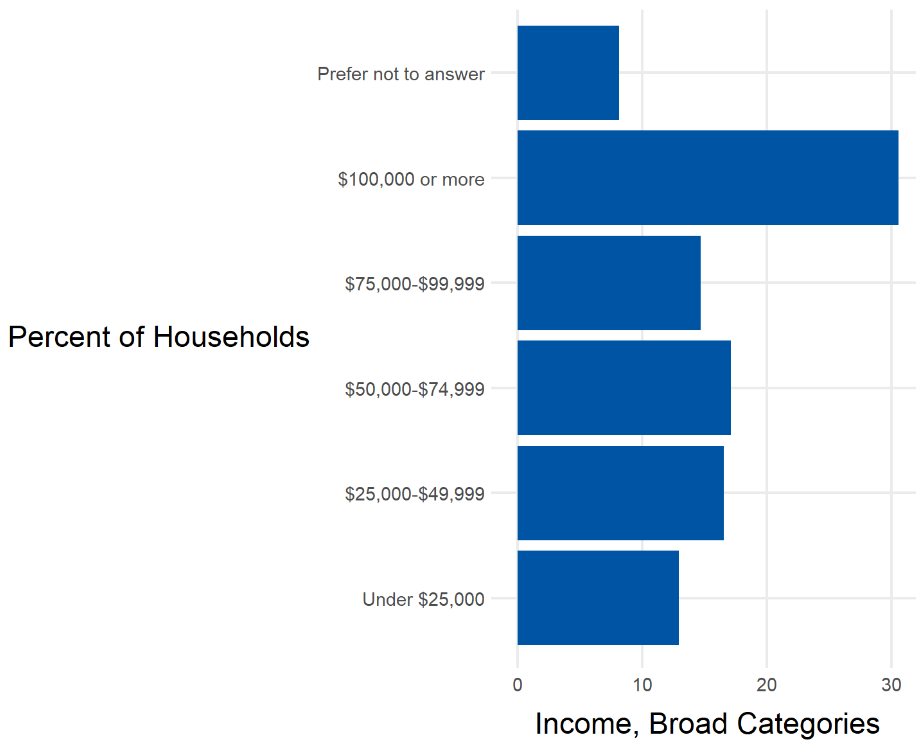
```
tbl <- tbi$hh %>%
  count(income_broad) %>%
  mutate(Percent = n / sum(n) * 100)

tbl %>%
  gt()
```

income_broad	n	Percent
Under $25,000	2480	12.935531
$25,000-$49,999	3169	16.529314
$50,000-$74,999	3279	17.103067
$75,000-$99,999	2816	14.688087
$100,000 or more	5861	30.570624
Prefer not to answer	1567	8.173378

```
tbl %>%
  ggplot(aes(x = as.factor(income_broad), y = Percent)) +
```

```
geom_col(fill = councilR::colors$councilBlue) +
labs(x = "Percent of Households", y = "Income, Broad Categories") +
coord_flip() +
councilR::theme_council()
```



Be careful with count variables! Can't simply take the mean:

```
tbi$hh %>%
  group_by(survey_year) %>%
  count(num_bicycles) %>%
  ungroup() %>%
  pivot_wider(names_from = survey_year, values_from = n) %>%
  gt::gt() %>%
  gt::sub_missing(columns = c(2:4))
```

num_bicycles	2019	2021	2023
NA	7516	---	---
0 (No bicycles)	---	2468	1217
1 bicycle	---	1944	893
2 bicycles	---	1693	775
3 bicycles	---	657	344
4 bicycles	---	522	258
5 bicycles	---	275	121
6 bicycles	---	126	65
7 bicycles	---	58	16
8 or more	---	101	44
Missing	---	63	16

Summarizing MRCVs: Multiple Response Categorical Variables

Some questions in the survey allow respondents to check multiple options. Common examples include questions about race and ethnicity, or reasons for not traveling. These "checkbox" variables, also known as multiple response categorical variables (MRCVs), are represented differently in the dataset from SRCVs (Single Response Categorical Variables) and have to be handled differently.

Let's use no_travel as an example.

MRCV Names and Descriptions

What does the data actually look like for this question?

```
tbi$day %>%
  select(person_id, day_id, starts_with("no_travel")) %>%
  tail() %>%
  gt::gt()
```

person_id	day_id	no_travel_other_specify	no_travel_8	no_travel_9	no_travel_4	no_travel_5	no_travel_6	no_travel_2	no_travel_99	no_travel_7	no_travel_3	no_travel_11	no_travel_1	no_travel_12
2314127201	231412720103	NA	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing
2311924101	231192410104	NA	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Selected	Not selected	Not selected	Not selected	Not selected	Not selected
2302769301	230276930104	NA	Not selected	Not selected	Not selected	Not selected	Not selected	Selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
2303498001	230349800104	NA	Not selected	Not selected	Selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
2312417001	231241700104	NA	Not selected	Not selected	Selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
2315044102	231504410205	NA	Not selected	Not selected	Selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected

```
tbi$day %>%
  left_join(tbi$hh) %>%
  group_by(survey_year) %>%
  count(no_travel_11) %>%
  ungroup() %>%
  gt::gt()
```

Joining with `by = join_by(hh_id)`

survey_year	no_travel_11	n
2019	Not selected	6387
2019	Selected	889
2019	Missing	72280
2021	Not selected	7924
2021	Selected	458
2021	Missing	41185
2023	Not selected	4236
2023	Selected	355
2023	Missing	24247

Get the labels for these columns

```
nt_desc <- filter(variable_list, str_detect(variable, "no_travel")) %>%
  select(variable, description)

nt_desc %>%
  gt::gt()
```

variable	description
no_travel_1	Reason for no travel on date: I did make trips on date
no_travel_11	Reason for no travel on date: Weather conditions (e.g., snowstorm)
no_travel_12	Reason for no travel on date: person may have made trips, but I don't know when or where
no_travel_2	Reason for no travel on date: Not scheduled to work/took day off
no_travel_3	Reason for no travel on date: Worked at home for pay (e.g., telework, self-employed)
no_travel_4	Reason for no travel on date: Hung out around home
no_travel_5	Reason for no travel on date: Scheduled school closure (e.g., holiday)
no_travel_6	Reason for no travel on date: No available transportation (e.g., no car, no bus)
no_travel_7	Reason for no travel on date: Sick or quarantining (self or others)
no_travel_8	Reason for no travel on date: Waited for visitor/delivery (e.g., plumber)
no_travel_9	Reason for no travel on date: person did online/remote/home school
no_travel_99	Reason for no travel on date: Other reason
no_travel_other_specify	Other reason for no travel on date open text answer

Clean up labels by removing the common part.

```
nt_desc2 <- nt_desc %>%
  mutate(description = str_replace(
    description,
    "Reason for no travel on date: ",
    ""
  ))

nt_desc2 %>%
  gt::gt()
```

variable	description
no_travel_1	I did make trips on date
no_travel_11	Weather conditions (e.g., snowstorm)
no_travel_12	person may have made trips, but I don't know when or where
no_travel_2	Not scheduled to work/took day off
no_travel_3	Worked at home for pay (e.g., telework, self-employed)
no_travel_4	Hung out around home
no_travel_5	Scheduled school closure (e.g., holiday)
no_travel_6	No available transportation (e.g., no car, no bus)
no_travel_7	Sick or quarantining (self or others)
no_travel_8	Waited for visitor/delivery (e.g., plumber)
no_travel_9	person did online/remote/home school
no_travel_99	Other reason
no_travel_other_specify	Other reason for no travel on date open text answer

Remove parentheticals at the end.

```
nt_desc3 <- nt_desc2 %>%
  mutate(description = str_replace(
    description,
    " \\(.*\\)$",
    ""
  ))

nt_desc3 %>%
  gt::gt()
```

variable	description
no_travel_1	I did make trips on date
no_travel_11	Weather conditions
no_travel_12	person may have made trips, but I don't know when or where
no_travel_2	Not scheduled to work/took day off
no_travel_3	Worked at home for pay
no_travel_4	Hung out around home
no_travel_5	Scheduled school closure
no_travel_6	No available transportation
no_travel_7	Sick or quarantining
no_travel_8	Waited for visitor/delivery
no_travel_9	person did online/remote/home school
no_travel_99	Other reason
no_travel_other_specify	Other reason for no travel on date open text answer

Summarize MRCV Data

To summarize the data follow these steps:

- Select the columns
- Filter out all the days that are missing responses (i.e., they did travel)
- Filter out all the days that aren't weighted (e.g. day not complete)
- Pivot data so we don't have to work across columns
- Calculate the fraction of respondents that checked each option
- Label responses

Select the columns:

```
missing_value <- "Missing"

nt_summary <-
  tbi$day %>%
  left_join(tbi$hh) %>%
  # select all the columns:
  select(survey_year, person_id, day_id, day_weight, starts_with("no_travel")) %>%
  # Get rid of open-ended text answer:
  select(-contains("other_specify"))
```

```
nt_summary %>%
  head(10) %>%
  gt()
```

survey_year	person_id	day_id	day_weight	no_travel_8	no_travel_9	no_travel_4	no_travel_5	no_travel_6	no_travel_2	no_travel_99	no_travel_7	no_travel_3	no_travel_11	no_travel_1	no_travel_12
2019	1822516701	182251670101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1830175901	183017590101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1831965201	183196520101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1834127101	183412710101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1834438301	183443830101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1835758001	183575800101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1836448001	183644800101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1846241501	184624150101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1849549501	184954950101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA
2019	1850835001	185083500101	0	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing	NA	NA

Filter and pivot the data:

```
nt_summary <-
  nt_summary %>%
  # Filter out all the rows that aren't weighted (e.g. day not complete)
  filter(
    !is.na(day_weight) & day_weight > 0
  ) %>%
  filter(
    !if_all(starts_with("no_travel"), ~ . %in% c("Missing", NA))
  ) %>%
  pivot_longer(cols = starts_with("no_travel"), names_to = "variable")


nt_summary %>%
  head(10) %>%
  gt()
```

survey_year	person_id	day_id	day_weight	variable	value
2019	1981190301	198119030102	29.49463	no_travel_8	Not selected
2019	1981190301	198119030102	29.49463	no_travel_9	Not selected
2019	1981190301	198119030102	29.49463	no_travel_4	Selected
2019	1981190301	198119030102	29.49463	no_travel_5	Not selected
2019	1981190301	198119030102	29.49463	no_travel_6	Not selected
2019	1981190301	198119030102	29.49463	no_travel_2	Not selected
2019	1981190301	198119030102	29.49463	no_travel_99	Not selected
2019	1981190301	198119030102	29.49463	no_travel_7	Not selected
2019	1981190301	198119030102	29.49463	no_travel_3	Not selected
2019	1981190301	198119030102	29.49463	no_travel_11	Not selected

Calculate the fraction of respondents that checked each option:

```
nt_summary <-
  nt_summary %>%
  # Remove missing responses
  filter(value != "Missing") %>%
  filter(!is.na(value)) %>%
  group_by(survey_year, variable, value) %>%
  summarize(
```

```
    wtd_num = sum(day_weight)
  ) %>%
  ungroup() %>%
  group_by(survey_year, variable) %>%
  mutate(wtd_pct = wtd_num / sum(wtd_num)) %>%
  ungroup()

nt_summary %>%
  head(10) %>%
  gt()
```

survey_year	variable	value	wtd_num	wtd_pct
2019	no_travel_1	NA	461644.29	1.0000000
2019	no_travel_11	Not selected	379270.90	0.8215652
2019	no_travel_11	Selected	82373.39	0.1784348
2019	no_travel_12	NA	461644.29	1.0000000
2019	no_travel_2	Not selected	182081.96	0.8066498
2019	no_travel_2	Selected	43644.20	0.1933502
2019	no_travel_3	Not selected	185662.81	0.8225135
2019	no_travel_3	Selected	40063.35	0.1774865
2019	no_travel_4	Not selected	295412.61	0.6399139
2019	no_travel_4	Selected	166231.68	0.3600861

Label responses and format table:

```
# Label responses
nt_summary <-
  nt_summary %>%
  # Label Responses
  left_join(nt_desc3, by = "variable")

# Put survey year in separate columns, format;
nt_summary %>%
  # Filter to the percent that "Selected" the response option:
  filter(value == "Selected") %>%
  filter(!is.na(value)) %>%
  select(survey_year, variable, description, wtd_pct) %>%
  pivot_wider(names_from = "survey_year", values_from = "wtd_pct") %>%
  gt() %>%
```

```
  gt::fmt_percent(columns = c(3:5), decimals = 1) %>%
  gt::sub_missing(columns = c(3:5), missing_text = "--")
```

variable	description	2019	2021	2023
no_travel_11	Weather conditions	17.8%	4.1%	8.9%
no_travel_2	Not scheduled to work/took day off	19.3%	12.5%	10.9%
no_travel_3	Worked at home for pay	17.7%	22.0%	22.3%
no_travel_4	Hung out around home	36.0%	52.3%	47.0%
no_travel_5	Scheduled school closure	7.7%	1.6%	2.8%
no_travel_6	No available transportation	2.6%	1.2%	1.5%
no_travel_7	Sick or quarantining	12.0%	6.2%	6.3%
no_travel_8	Waited for visitor/delivery	1.4%	1.6%	1.1%
no_travel_9	person did online/remote/home school	0.9%	2.8%	3.2%
no_travel_99	Other reason	14.1%	11.3%	11.4%
no_travel_1	I did make trips on date	--	1.0%	1.1%
no_travel_12	person may have made trips, but I don't know when or where	--	0.7%	1.2%

Working with Weights

Weighted proportions are calculated here in two different ways. First, they are calculated manually to illustrate a simple approach.

Second, the srvyr package (basically a wrapper around the survey package) is demonstrated to calculate standard errors and confidence intervals as well as the proportions themselves.

Weighted Proportions

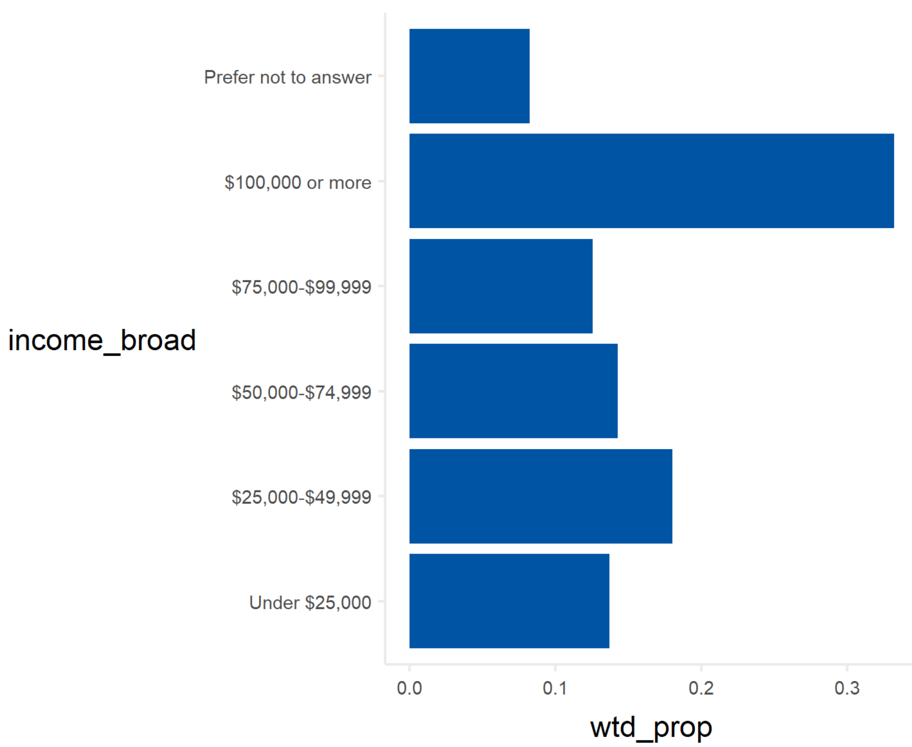
To calculate weighted proportions, or weighted counts, we simply sum the weights instead of counting the number of rows.

```
tbl <- tbi$hh %>%
  group_by(income_broad) %>%
  summarize(N = n(), wtd_N = sum(hh_weight)) %>%
  mutate(prop = N / sum(N), wtd_prop = wtd_N / sum(wtd_N)) %>%
  ungroup()

tbl %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 0) %>%
  gt::fmt_percent(columns = c(4:5), decimals = 1)
```

income_broad	N	wtd_N	prop	wtd_prop
Under $25,000	2,480	620,099	12.9%	13.7%
$25,000-$49,999	3,169	813,693	16.5%	18.0%
$50,000-$74,999	3,279	644,296	17.1%	14.3%
$75,000-$99,999	2,816	567,726	14.7%	12.6%
$100,000 or more	5,861	1,501,144	30.6%	33.2%
Prefer not to answer	1,567	372,129	8.2%	8.2%

```
tbl %>%
  ggplot(aes(x = income_broad, y = wtd_prop)) +
  geom_col(fill = councilR::colors$councilBlue) +
  councilR::theme_council_open() +
  coord_flip()
```



Using the srvyr Package

Here the srvyr package is used to calculate the standard error as well.

The srvyr package uses a design object instead of the data directly. This contains information about the survey design.

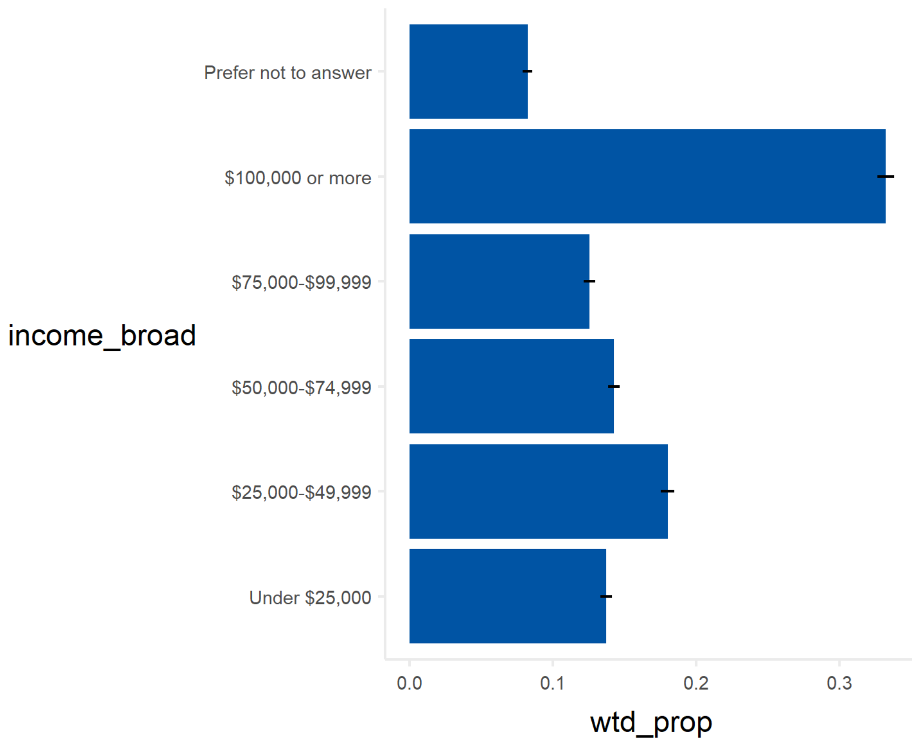
```
# Define the design
hh_design <- tbi$hh %>%
  filter(!is.na(hh_weight)) %>%
  as_survey_design(
    ids = hh_id,
    weights = hh_weight,
    strata = sample_segment
  )
tbl <- hh_design %>%
  group_by(income_broad) %>%
  summarize(
    N = n(),
    wtd_est = survey_total(vartype = "se"),
    wtd_prop = survey_prop(vartype = "se", proportion = TRUE)
  ) %>%
  ungroup()


tbl %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:4), decimals = 0) %>%
  gt::fmt_percent(columns = c(5:6), decimals = 1)
```

income_broad	N	wtd_est	wtd_est_se	wtd_prop	wtd_prop_se
Under $25,000	2,480	620,099	20,079	13.7%	0.4%
$25,000-$49,999	3,169	813,693	22,047	18.0%	0.5%
$50,000-$74,999	3,279	644,296	19,284	14.3%	0.4%
$75,000-$99,999	2,816	567,726	18,379	12.6%	0.4%
$100,000 or more	5,861	1,501,144	31,250	33.2%	0.6%
Prefer not to answer	1,567	372,129	15,522	8.2%	0.3%

```
tbl %>%
  ggplot(aes(x = income_broad, y = wtd_prop)) +
```

```
geom_col(fill = councilR::colors$councilBlue) +
geom_linerange(
  aes(ymin = wtd_prop - wtd_prop_se, ymax = wtd_prop + wtd_prop_se)
) +
coord_flip() +
councilR::theme_council_open()
```



A Note on Sample "Strata"

In the code above, the survey design specifies "strata", corresponding to the sample segments from the sample plan. Specifying strata improves the precision of confidence intervals and standard errors, but does not change the point estimates.

In survey data analysis, strata are used to ensure that the sample is representative of key subgroups in the population or when survey sampling rates differed across subpopulations or subgeographies. In the Travel Behavior Inventory, Core Urban sample segments were oversampled, and hence strata specification (using sample segments as strata) is warranted.

Failing to specify strata, while not exactly correct, rarely has more than a miniscule effect on error estimates. In the example below, we show how confidence intervals are affected by specifying strata in the survey design object.

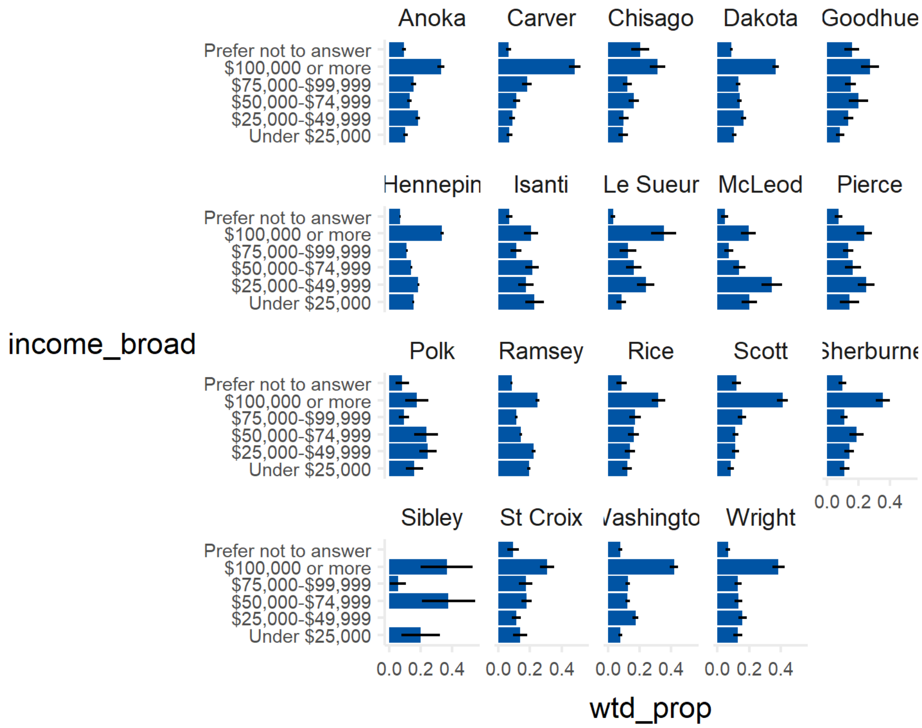
Weighted Proportion by Another Variable

Calculating proportions grouped by another variable requires some thought as to the weights applied.

Variables in the Same Table

Here we calculate income by home county. Both variables are household-level variables, so household weights are applied.

```
tbl %>%
  mutate(home_county = gsub(" County, MN| County, WI", "", home_county)) %>%
  ggplot(aes(x = income_broad, y = wtd_prop)) +
  geom_col(fill = councilR::colors$councilBlue) +
  geom_linerange(
    aes(ymin = wtd_prop - wtd_prop_se, ymax = wtd_prop + wtd_prop_se)
  ) +
  coord_flip() +
  facet_wrap(~home_county) +
  councilR::theme_council_open()
```



Variables in Different Tables

When the crosstab variables are in different tables, the general rule of thumb is to use the weight from the lowest level of the weighting hierarchy: trip weights, day weights, person weights, and finally household weights.

	TABLE 2					
	hh	**vehicle**	**person**	**day**	**trip**	**trip_purpose**
TABLE 1						
hh	hh_weight					
vehicle	hh_weight	hh_weight				
person	person_weight	person_weight	person_weight			
day	day_weight	day_weight	day_weight	day_weight		
trip	trip_weight	trip_weight	trip_weight	trip_weight	trip_weight	
trip_purpose*	trip_purpose_weight	trip_purpose_weight	trip_purpose_weight	trip_purpose_weight	**	trip_purpose_weight
*This table is provided solely for analyses of overall trip purpose. Do not join without first aggregating trip measures to linked trip IDs.						

Example 1: Telework by County

For example, to find the proportion of people who teleworked part-time before the pandemic (a person-level variable) by home county (a household-level variable), the data user joins the household and person tables, then summarizes the person weight for a weighted estimate.

```
tab <- tbi$person %>%
  left_join(tbi$hh, by = "hh_id") %>%
  filter(survey_year == 2023) %>%
  group_by(home_county, job_type_pre_covid) %>%
  summarize(wtd_population = sum(person_weight)) %>%
  ungroup() %>%
  group_by(home_county) %>%
  mutate(wtd_prop_population = wtd_population / sum(wtd_population)) %>%
  ungroup()


tab %>%
  head(12) %>%
  gt::gt() %>%
  gt::fmt_percent(columns = 4) %>%
  gt::fmt_number(columns = 3, decimals = 0)
```

home_county	job_type_pre_covid	wtd_population	wtd_prop_population
Anoka County, MN	Went to one work location ONLY (outside of home)	142,719	42.30%
Anoka County, MN	Work location regularly varied (different offices/jobsites)	10,191	3.02%
Anoka County, MN	Worked ONLY from home or remotely (telework, self-employed)	10,790	3.20%
Anoka County, MN	Drove/biked/traveled for work (driver, sales, deliveries)	2,366	0.70%
Anoka County, MN	Teleworked some days and traveled to a work location some days	10,012	2.97%
Anoka County, MN	Missing	161,314	47.81%
Carver County, MN	Went to one work location ONLY (outside of home)	39,615	35.90%
Carver County, MN	Work location regularly varied (different offices/jobsites)	4,480	4.06%
Carver County, MN	Worked ONLY from home or remotely (telework, self-employed)	6,967	6.31%
Carver County, MN	Teleworked some days and traveled to a work location some days	5,048	4.58%
Carver County, MN	Missing	54,226	49.15%
Chisago County, MN	Went to one work location ONLY (outside of home)	26,314	53.18%

Example 2: Trip Distance by Mode

In this second example, we find the percent of trips that are under 1, 5, or 10 miles for each mode. This kind of analysis might be used to answer the question “what percent of drive trips are short enough that they might be shifted to [walking/biking]?” In this example, we use srvyr to get both the weighted proportions and their standard errors.

The results below show that, in 2023, 15.6% of drive drips were less than 1 mile long, compared to 84.5% of walk trips.

```
tab <- tbi$trip %>%
  left_join(select(tbi$hh, "hh_id", "sample_segment", "survey_year")) %>%
  filter(!is.na(distance_miles)) %>%
  filter(survey_year == 2023) %>%
  mutate(distance_category = case_when(distance_miles < 1 ~ "Less than 1 mile",
    distance_miles < 5 ~ "1-5 miles",
    distance_miles < 10 ~ "5-10 miles",
```

```
    distance_miles >= 10 ~ "10 miles or more",
    .default = "Missing"
  )) %>%
  as_survey_design(
    ids = trip_id,
    weights = trip_weight,
    strata = sample_segment
  ) %>%
  group_by(survey_year, mode_type, distance_category) %>%
  summarize(prop = survey_prop(proportion = FALSE)) %>%
  ungroup()


tab %>%
  mutate(distance_category = factor(distance_category,
    levels = c(
      "Less than 1 mile",
      "1-5 miles",
      "5-10 miles",
      "10 miles or more"
    )
  )) %>%
  arrange(mode_type, distance_category) %>%
  filter(mode_type %in% c("Walk", "Household Vehicle")) %>%
  head(12) %>%
  gt::gt() %>%
  gt::fmt_percent(columns = 4:5, decimals = 1)
```

survey_year	mode_type	distance_category	prop	prop_se
2023	Household Vehicle	Less than 1 mile	15.6%	0.4%
2023	Household Vehicle	1-5 miles	43.0%	0.5%
2023	Household Vehicle	5-10 miles	18.7%	0.4%
2023	Household Vehicle	10 miles or more	22.7%	0.5%
2023	Walk	Less than 1 mile	84.5%	1.0%
2023	Walk	1-5 miles	12.6%	0.9%
2023	Walk	5-10 miles	1.3%	0.4%
2023	Walk	10 miles or more	1.5%	0.4%

Sample Analyses

Weighted Trip Rates

To calculate a weighted trip rate (the number of trips per day), the number of weighted trips is divided by the number of weighted travel days.

For example:

- If there are 300,000 weighted person-trips across 75,000 person-days, then the average person-trip rate is 4.0 per day.
- If there are 225,000 person-trips by car across 75,000 person-days, then the person-trip rate for car trips is 3.0.

Note: Travel days with no trips are not represented in the trip table! They must be obtained by performing a join on the day table.

Weighted trip rates can either be calculated using the unlinked trip table ("trip"), or the linked trip table ("linked_trip"). The unlinked trip rate represents the number of trips made inclusive of multi-part journeys, like walking to a bus stop. The linked trip table consolidates these multi-trip journeys into a single trip record. As a result, linked trip rates are always lower than unlinked trip rates.

```
trip_counts <- tbi$trip %>%
  filter(!is.na(trip_weight) & trip_weight > 0) %>%
  group_by(day_id) %>%
  summarize(wtd_num_trips = sum(trip_weight)) %>%
  ungroup()

trip_counts %>%
  tail(10) %>%
  gt::gt()
```

day_id	wtd_num_trips
231735820302	1439.4171
231735820402	1439.4171
231735820502	1439.4171
231737120101	1335.3016
231737120201	1573.3379
231737120301	511.6536
231737120401	255.8268
231737990101	989.7791
231737990102	395.9116
231737990103	791.8233

```
# Join to day table

trip_counts <- tbi$day %>%
  filter(day_weight > 0) %>%
  select(hh_id, day_id, day_weight) %>%
  left_join(trip_counts, by = "day_id") %>%
  left_join(select(tbi$hh, hh_id, sample_segment, income_broad), by = "hh_id")



trip_counts %>%
  tail(10) %>%
  gt::gt()
```

hh_id	day_id	day_weight	wtd_num_trips	sample_segment	income_broad
23034980	230349800107	62.45773	NA	Core Urban BGs	$100,000 or more
23050326	230503260101	734.23731	NA	Core Urban BGs	Prefer not to answer
23088895	230888950101	50.31765	123.4231	Core Urban BGs	$50,000-$74,999
23049602	230496020301	2573.17431	NA	Core Urban BGs	$100,000 or more
23033885	230338850301	512.73826	1061.1328	Core Urban BGs	$100,000 or more
23150441	231504410202	115.50532	NA	Core Rural	$75,000-$99,999
23125965	231259650203	99.00411	198.0082	Core Urban BGs	$100,000 or more
23003220	230032200103	34.53968	NA	Core Urban BGs	Under $25,000
23092788	230927880205	99.07492	693.5245	Core Urban BGs	$100,000 or more
23089592	230895920105	49.98151	NA	Core Urban BGs	$100,000 or more

Replace the NAs with 0s.

```
trip_counts <- trip_counts %>%
  mutate(wtd_num_trips = replace_na(wtd_num_trips, 0))

trip_counts %>%
  tail(10) %>%
  gt::gt()
```

hh_id	day_id	day_weight	wtd_num_trips	sample_segment	income_broad
23034980	230349800107	62.45773	0.0000	Core Urban BGs	$100,000 or more
23050326	230503260101	734.23731	0.0000	Core Urban BGs	Prefer not to answer
23088895	230888950101	50.31765	123.4231	Core Urban BGs	$50,000-$74,999
23049602	230496020301	2573.17431	0.0000	Core Urban BGs	$100,000 or more
23033885	230338850301	512.73826	1061.1328	Core Urban BGs	$100,000 or more
23150441	231504410202	115.50532	0.0000	Core Rural	$75,000-$99,999
23125965	231259650203	99.00411	198.0082	Core Urban BGs	$100,000 or more
23003220	230032200103	34.53968	0.0000	Core Urban BGs	Under $25,000
23092788	230927880205	99.07492	693.5245	Core Urban BGs	$100,000 or more
23089592	230895920105	49.98151	0.0000	Core Urban BGs	$100,000 or more

Finally, calculate the mean. Note that it is not appropriate to calculate the mean trip rate with the mean function.

```r
trip_counts %>%
  summarize(
    N = n(),
    correct = sum(wtd_num_trips) / sum(day_weight),
    INCORRECT = mean(wtd_num_trips / day_weight)
  ) %>%
  gt::gt()
```

N	correct	INCORRECT
64015	3.544836	4.031447

As a simple demonstration of why there is a difference:

```r
foo <- data.frame(trips = c(2, 10), days = c(2, 5))
foo %>%
  gt::gt()
```

trips	days
2	2
10	5

```r
foo %>%
  summarize(
    correct = sum(trips) / sum(days),
    INCORRECT = mean(trips / days)
  ) %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(1:2), decimals = 2)
```

correct	INCORRECT
1.71	1.50

To use srvyr, simply specify the calculation in the call to survey_mean.

```
tc_design <- trip_counts %>%
  as_survey_design(
    ids = day_id,
    weights = day_weight,
    strata = sample_segment
  )

tc_design %>%
  summarize(
    N = n(),
    wtd_trip_rate = survey_mean(wtd_num_trips / day_weight)
  ) %>%
  gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 2)
```

N	wtd_trip_rate	wtd_trip_rate_se
64015	3.54	0.02

Calculate by group as well.

```
tc_design %>%
  group_by(income_broad) %>%
  summarize(
    N = n(),
    wtd_trip_rate = survey_mean(wtd_num_trips / day_weight)
  ) %>%
  ungroup() %>%
  gt() %>%
  gt::fmt_number(columns = c(2), decimals = 0) %>%
  gt::fmt_number(columns = c(3:4), decimals = 2)
```

income_broad	N	wtd_trip_rate	wtd_trip_rate_se
Under $25,000	5,315	3.24	0.09
$25,000-$49,999	8,182	3.45	0.07
$50,000-$74,999	10,042	3.49	0.06
$75,000-$99,999	10,481	3.66	0.07
$100,000 or more	25,721	3.72	0.04
Prefer not to answer	4,274	3.09	0.08

To recap, when calculating a simple aggregate trip rate, joins are made to the day table to account for *days without travel.* The user:

- Gets the weighted number of trips for each day from the trip table.
- Joins to the day table.
- Fills in the weighted number of trips with zeros for those days without travel.

Summary statistics can then be created from the weighted number of trips divided by the day weight. The code below consolidates the example walk through from above:

```
trip_counts <- tbi$trip %>%
  filter(!is.na(trip_weight) & trip_weight > 0) %>%
  group_by(day_id) %>%
  summarize(wtd_num_trips = sum(trip_weight)) %>%
  ungroup()

trip_counts <- tbi$day %>%
  filter(day_weight > 0) %>%
  select(hh_id, day_id, day_weight) %>%
  left_join(trip_counts, by = "day_id") %>%
  left_join(select(tbi$hh, hh_id, sample_segment, survey_year),
    by = "hh_id"
  )

trip_counts <- trip_counts %>%
  mutate(wtd_num_trips = replace_na(wtd_num_trips, 0))

trip_counts %>%
  as_survey_design(
    ids = day_id,
    weights = day_weight,
```

```
  strata = sample_segment
) %>%
group_by(survey_year) %>%
summarize(
  N = n(),
  wtd_trip_rate = survey_mean(wtd_num_trips / day_weight)
) %>%
ungroup() %>%
gt() %>%
gt::fmt_number(columns = c(2), decimals = 0) %>%
gt::fmt_number(columns = c(3:4), decimals = 2)
```

survey_year	N	wtd_trip_rate	wtd_trip_rate_se
2019	32,150	4.19	0.04
2021	20,780	2.87	0.04
2023	11,085	3.58	0.05

Special Case: Trip Rates by a Trip Variable

When calculating trip rates by a trip attribute (i.e., a variable in the trip table) the user must not only account for days without travel, but also *days without trips that match that attribute.* In the case of mode type, this requires an expansion of the dataset to encompass every day, and every mode.

For example, to find trip rates by mode, the user:

- Calculates the weighted number of trips by mode and day ID.
- Expands the trip count data frame to include days without travel by each mode. In the code below, this is done by creating an empty data frame containing every combination of day_id and mode_type.
- Fills in missing values for the weighted number of trips with zeros, representing days without travel by that mode type.

```
# Weighted number of trips by mode type and day:
trip_counts <- tbi$trip %>%
  filter(!is.na(trip_weight) & trip_weight > 0) %>%
  group_by(day_id, mode_type) %>%
  summarize(wtd_num_trips = sum(trip_weight)) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'day_id'. You can override using the
`.groups` argument.
```

```
# An empty data frame to hold every combination of day id and
empty_grid <- expand.grid(
  day_id = tbi$day$day_id,
  mode_type = levels(tbi$trip$mode_type)
)

empty_grid$mode_type <- factor(empty_grid$mode_type,
  levels = levels(tbi$trip$mode_type),
  ordered = TRUE
)

all_levels <- union(levels(trip_counts$mode_type), levels(empty_grid$mode_type))

# Align levels
trip_counts$mode_type <- factor(trip_counts$mode_type, levels = all_levels, ordered = TRUE)
empty_grid$mode_type <- factor(empty_grid$mode_type, levels = all_levels, ordered = TRUE)

# Join to our empty grid of modes and days:
trip_counts <- empty_grid %>%
  left_join(trip_counts, by = c("day_id", "mode_type")) # can you fill NAs with zeros here?

# Fill in zeros:
trip_counts <- trip_counts %>%
  mutate(wtd_num_trips = nafill(wtd_num_trips, fill = 0))

trip_counts %>%
  head(10) %>%
  gt()
```

day_id	mode_type	wtd_num_trips
182251670101	Rail	0
183017590101	Rail	0
183196520101	Rail	0
183412710101	Rail	0
183443830101	Rail	0
183575800101	Rail	0
183644800101	Rail	0
184624150101	Rail	0
184954950101	Rail	0
185083500101	Rail	0

Once created, the expanded table of trip counts can be joined to other, non-trip tables and summarized in the same manner as other trip rate crosstabs. For example, to calculate trip rate by home county and mode type:

```
trip_counts %>%
  left_join(tbi$day) %>%
  left_join(tbi$hh) %>%
  group_by(survey_year, home_county, mode_type) %>%
  summarize(trip_rate = sum(wtd_num_trips) / sum(day_weight)) %>%
  ungroup() %>%
  head(10) %>%
  gt()
```

survey_year	home_county	mode_type	trip_rate
2019	Anoka County, MN	Rail	0.003992510
2019	Anoka County, MN	School Bus	0.078010777
2019	Anoka County, MN	Public Bus	0.033112483
2019	Anoka County, MN	Other Bus	0.018334470
2019	Anoka County, MN	Long distance passenger mode	0.002931250
2019	Anoka County, MN	Smartphone ridehailing service	0.001922443
2019	Anoka County, MN	For-Hire Vehicle	0.001066327
2019	Anoka County, MN	Household Vehicle	3.434414204
2019	Anoka County, MN	Other Vehicle	0.180770027
2019	Anoka County, MN	Micromobility	0.023927973

To extend this process to multiple trip variables, such as trip mode and trip departure hour, the empty data frame simply needs to be expanded to include the second variable before joining and filling in days without travel.

```
empty_grid <- expand.grid(
  day_id = tbi$day$day_id,
  mode_type = value_labels_upcoded[variable == "mode_type", value],
  depart_hour = 0:23
)

empty_grid %>%
  head(10) %>%
  gt()
```

day_id	mode_type	depart_hour
182251670101	1	0
183017590101	1	0
183196520101	1	0
183412710101	1	0
183443830101	1	0
183575800101	1	0
183644800101	1	0
184624150101	1	0
184954950101	1	0
185083500101	1	0

Weighted Mode Share

Mode share represents the proportion of trips taken by walking, driving, biking, transit, and other modes. It is calculated from the linked trip table.

Prepare the design object for srvyr:

```
trips1 <- tbi$linked_trip %>%
  left_join(select(tbi$hh, hh_id, sample_segment), by = "hh_id") %>%
  filter(!is.na(linked_trip_weight) & linked_trip_weight > 0)

trip_design <- trips1 %>%
  as_survey_design(
    ids = linked_trip_id,
    weights = linked_trip_weight,
    strata = sample_segment
  )
```

Calculate and label:

```
tbl <- trip_design %>%
  group_by(mode_type) %>%
  summarize(
    N = n(),
    wtd_N = sum(linked_trip_weight),
    wtd_prop = survey_prop(vartype = "se", proportion = TRUE)
```

```
  ) %>%
  ungroup()


tbl %>%
  select(mode_type, N, wtd_N, wtd_prop, wtd_prop_se) %>%
  arrange(desc(wtd_prop)) %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 0) %>%
  gt::fmt_percent(columns = c(4:5), decimals = 1) %>%
  gt::sub_small_vals(threshold = 0.001, small_pattern = "<0.1%")
```

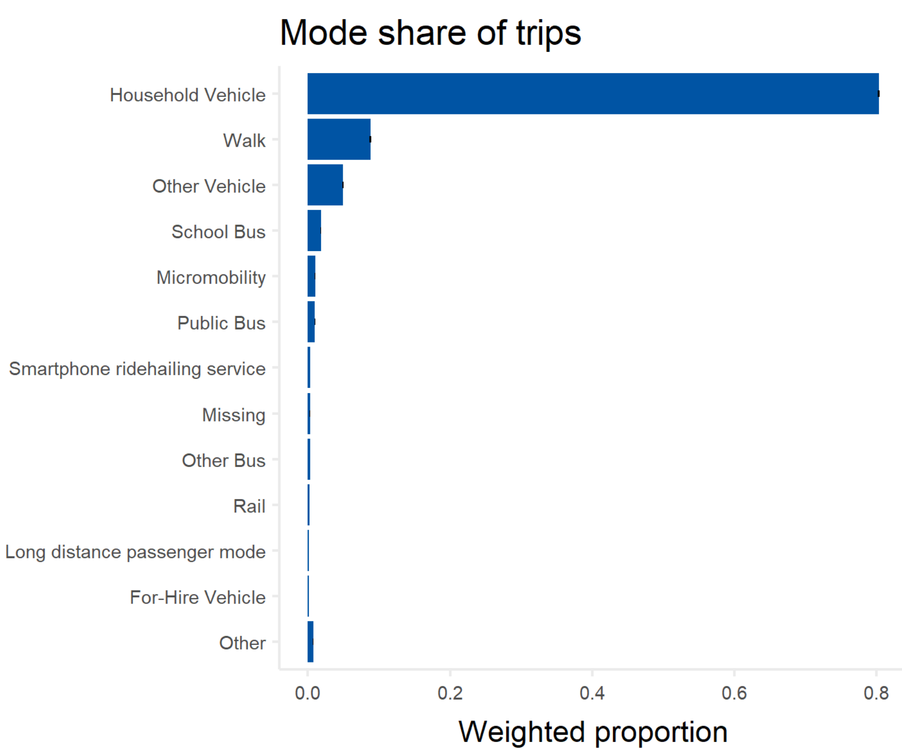
mode_type	N	wtd_N	wtd_prop	wtd_prop_se
Household Vehicle	180,512	31,374,063	80.3%	0.2%
Walk	29,051	3,437,077	8.8%	0.1%
Other Vehicle	13,277	1,913,337	4.9%	<0.1%
School Bus	2,876	716,939	1.8%	<0.1%
Micromobility	3,734	391,142	1.0%	<0.1%
Public Bus	4,596	376,769	1.0%	<0.1%
Other	1,856	286,898	0.7%	<0.1%
Smartphone ridehailing service	901	129,296	0.3%	<0.1%
Missing	476	118,171	0.3%	<0.1%
Other Bus	806	116,212	0.3%	<0.1%
Rail	1,841	91,470	0.2%	<0.1%
Long distance passenger mode	400	57,953	0.1%	<0.1%
For-Hire Vehicle	219	48,910	0.1%	<0.1%

Plot:

```
tbl <- tbl %>%
  mutate(mode_type = mode_type %>%
    fct_reorder(wtd_prop) %>%
    fct_relevel("Other"))

ggplot(tbl, aes(x = mode_type, y = wtd_prop)) +
  geom_col(fill = councilR::colors$councilBlue) +
  geom_linerange(
    aes(ymin = wtd_prop - wtd_prop_se, ymax = wtd_prop + wtd_prop_se),
```

```
  linewidth = 1.2
) +
coord_flip() +
labs(
  title = "Mode share of trips",
  x = "",
  y = "Weighted proportion"
) +
councilR::theme_council_open()
```



Weighted Purpose Share

Assigning the "purpose" to a trip can be done in a few ways. In previous analyses of the Travel Behavior Inventory data, trip purpose has been aggregated in one of two ways:

1. The user could use the destination place type (d_purpose, d_purpose_category)
2. The user can account for the origin and destination at the same time, by grouping trips into the "tour" type used during modeling (e.g., categorize trips as home-based work,

home-based other, or non-home-based trips). This is typically the approach used in modeling.

RSG has provided a third option with the trip_purpose table. This table allows users to calculate aggregate, purpose-related statistics that accounts for the origin *and* destination place types of trips, in a manner more in keeping with the way a layperson might describe the “purpose” of their travel.

The trip_purpose table:

1. Re-links transit trips to minimize “change mode” purposes in the final trip purpose dataset. Some “change mode” purposes will still be present, for long-distance trips.
2. Divides linked trips into two categories: those with a home end, and those without.
3. For trips with a home end, the purpose of the trip is the destination place type of the non-home end. For example, a trip from home to work receives the purpose of work. The weight for the trip purpose record is equal to the original weight of the trip.
4. For trips without a home end, the purpose of the trip is shared equally between the origin end and destination end. For example, a trip from grocery shopping to school has the purpose of both shopping and school. The single trip is divided into two trip_purpose records, one for the grocery shopping end and a second for the school end. Half of the original trip weight is assigned to the grocery shop trip, with the other half to the school trip.

The end result is a table with *fewer* original trips than the trip table, owing to transit-trip re-linking; and *more* rows than the original trip table, owing to the creation of duplicate records for non-home-based trips.

Trip purpose by the destination place type

Note that “home” is the predominant “purpose” (destination place type).

```
tbl <-
  tbi$trip %>%
  left_join(tbi$hh) %>%
  filter(survey_year == 2023) %>%
  as_survey_design(
    weights = "trip_weight",
    ids = "trip_id",
    strata = "sample_segment"
  ) %>%
```

```
  group_by(d_purpose_category) %>%
  summarize(prop = survey_prop(proportion = TRUE, se = TRUE)) %>%
  ungroup()


tbl %>%
  arrange(-prop) %>%
  gt::gt() %>%
  gt::fmt_percent(columns = c(2:3))
```

d_purpose_category	prop	prop_se
Home	32.35%	0.44%
Shopping	10.79%	0.28%
Social/Recreation	9.73%	0.27%
Escort	9.58%	0.29%
Work	8.17%	0.27%
Meal	6.93%	0.25%
Errand	5.45%	0.19%
Work related	5.18%	0.19%
School	3.49%	0.18%
Change mode	3.37%	0.14%
Overnight	2.50%	0.13%
Other	1.93%	0.12%
School related	0.48%	0.06%
Not imputable	0.04%	0.01%

Trip purpose according to the trip purpose table

Note that some “home” trips remain; these are loop trips that start and end at home. some “not imputable” trips also remain, where the trip purpose could not be inferred for both ends. These trips can be removed from aggregate analyses.

```
tbl <-
  tbi$trip_purpose %>%
  left_join(tbi$hh) %>%
  filter(survey_year == 2023) %>%
  as_survey_design(
    weights = "trip_purpose_weight",
    ids = "trip_purpose_id",
```

```
    strata = "sample_segment"
  ) %>%
  group_by(purpose_category) %>%
  summarize(prop = survey_prop(proportion = TRUE, se = TRUE)) %>%
  ungroup()


tbl %>%
  arrange(-prop) %>%
  gt::gt() %>%
  gt::fmt_percent(columns = c(2:3))
```

purpose_category	prop	prop_se
Shopping	15.12%	0.29%
Social/Recreation	14.82%	0.30%
Escort	14.50%	0.32%
Work	12.46%	0.31%
Meal	9.65%	0.26%
Errand	9.16%	0.21%
Missing	7.15%	0.21%
Work related	6.81%	0.19%
School	5.28%	0.21%
Overnight	3.12%	0.13%
Other	1.15%	0.07%
School related	0.65%	0.07%
Home	0.08%	0.02%
Not imputable	0.05%	0.01%

Weighted VMT

Vehicle miles traveled at a trip-level is the miles traveled for any auto trip, divided by auto occupancy (num_travelers on the vehicle trip, inclusive of the driver).

```
auto_modes <- c(
  "Household Vehicle",
  "Other Vehicle",
  "For-Hire Vehicle",
  "Smartphone ridehailing service"
)
```

```
vmt <- tbi$trip %>%
  mutate(vmt = ifelse(mode_type %in% auto_modes, distance_miles / as.numeric(num_travelers
), 0))

vmt %>%
  select(trip_id, mode_type, distance_miles, num_travelers, vmt, trip_weight) %>%
  head(15) %>%
  gt()
```

trip_id	mode_type	distance_miles	num_travelers	vmt	trip_weight
1948330701004	Household Vehicle	0.23488	1 traveler	0.23488	0.00000
1948330701003	Household Vehicle	0.23488	1 traveler	0.23488	0.00000
1979608101002	Micromobility	0.06474	1 traveler	0.00000	39.22129
1979608101009	Micromobility	0.06564	1 traveler	0.00000	0.00000
1852656801004	Micromobility	0.14230	5+ travelers	0.00000	80.41805
1873245201038	Rail	0.41876	1 traveler	0.00000	0.00000
1917694203001	Other Vehicle	0.78029	1 traveler	0.78029	0.00000
1917694203025	Micromobility	0.70227	1 traveler	0.00000	471.42069
1917694203015	Micromobility	0.69517	1 traveler	0.00000	471.42069
1917694203016	Micromobility	0.23870	1 traveler	0.00000	471.42069
1917694203024	Micromobility	0.40851	1 traveler	0.00000	471.42069
1965884901032	Micromobility	0.37715	2 travelers	0.00000	0.00000
1928817801004	Micromobility	0.57222	1 traveler	0.00000	0.00000
1928817801010	Micromobility	0.74898	1 traveler	0.00000	178.52850
1976150402010	Micromobility	0.41479	1 traveler	0.00000	110.79780

VMT statistics are very sensitive to outliers. Removing vehicle trips that are “flagged” as having excessively high speeds (over 100 mph) is recommended. These are likely trips where the user incorrectly classified a trip in their survey.

```
vmt <- vmt %>%
  mutate(vmt = ifelse(speed_mph >= 100, 0, vmt))
```

To scale up individual VMT records to the region, we multiply VMT for each trip by its corresponding trip weight. This is weighted VMT.

```
weighted_vmt <- vmt %>%
  mutate(weighted_vmt = vmt * trip_weight)

weighted_vmt %>%
  select(trip_id, mode_type, distance_miles, num_travelers, vmt, trip_weight) %>%
  head(15) %>%
  gt()
```

trip_id	mode_type	distance_miles	num_travelers	vmt	trip_weight
1948330701004	Household Vehicle	0.23488	1 traveler	NA	0.00000
1948330701003	Household Vehicle	0.23488	1 traveler	NA	0.00000
1979608101002	Micromobility	0.06474	1 traveler	0.00000	39.22129
1979608101009	Micromobility	0.06564	1 traveler	0.00000	0.00000
1852656801004	Micromobility	0.14230	5+ travelers	0.00000	80.41805
1873245201038	Rail	0.41876	1 traveler	0.00000	0.00000
1917694203001	Other Vehicle	0.78029	1 traveler	0.78029	0.00000
1917694203025	Micromobility	0.70227	1 traveler	0.00000	471.42069
1917694203015	Micromobility	0.69517	1 traveler	0.00000	471.42069
1917694203016	Micromobility	0.23870	1 traveler	0.00000	471.42069
1917694203024	Micromobility	0.40851	1 traveler	0.00000	471.42069
1965884901032	Micromobility	0.37715	2 travelers	0.00000	0.00000
1928817801004	Micromobility	0.57222	1 traveler	0.00000	0.00000
1928817801010	Micromobility	0.74898	1 traveler	0.00000	178.52850
1976150402010	Micromobility	0.41479	1 traveler	0.00000	110.79780

The total weighted VMT represents the total VMT by residents of the region.

Just as with trip rates, this may include travel outside of the region. Data users may wish to filter trips that start or end outside of their regional boundaries.

```
total_vmt <-
  weighted_vmt %>%
  left_join(tbi$hh, by = "hh_id") %>%
  group_by(survey_year) %>%
  summarize(total_daily_vmt = sum(weighted_vmt, na.rm = T)) %>%
  ungroup()

total_vmt %>%
```

```
  gt::gt() %>%
  gt::fmt_number(columns = c(2), decimals = 0)
```

survey_year	total_daily_vmt
2019	81,703,466
2021	56,441,417
2023	63,620,090

Household-Level VMT

To calculate average VMT per household per day, we divide the total weighted VMT by the total of the household weights, i.e., the total number of households in the region. This comes to about 40-50 miles per household, per day.

```
total_households <-
  tbi$hh %>%
  group_by(survey_year) %>%
  summarize(total_hh = sum(hh_weight)) %>%
  ungroup()


total_households %>%
  left_join(total_vmt) %>%
  mutate(vmt_per_hh_per_day = total_daily_vmt / total_hh) %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 0) %>%
  gt::fmt_number(columns = 4, decimals = 1)
```

survey_year	total_hh	total_daily_vmt	vmt_per_hh_per_day
2019	1,487,294	81,703,466	54.9
2021	1,517,145	56,441,417	37.2
2023	1,514,648	63,620,090	42.0

Person-Level VMT

Similarly, to calculate the average VMT per person per day, we divide the total weighted VMT by the total of the person weights, i.e., the total number of households in the region.

```
total_persons <-
  tbi$person %>%
  left_join(tbi$hh, by = "hh_id") %>%
  group_by(survey_year) %>%
  summarize(total_pop = sum(person_weight)) %>%
  ungroup()

total_persons %>%
  left_join(total_vmt) %>%
  mutate(vmt_per_person_per_day = total_daily_vmt / total_pop) %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 0) %>%
  gt::fmt_number(columns = 4, decimals = 1)
```

survey_year	total_pop	total_daily_vmt	vmt_per_person_per_day
2019	3,682,918	81,703,466	22.2
2021	3,748,414	56,441,417	15.1
2023	3,750,006	63,620,090	17.0

Person-Level VMT with Standard Errors

If variance estimates or more granular summaries are required, a similar process to calculating trip rates is used to join trip-level weighted VMT to day records.

First, we total the weighted VMT by day ID:

```
person_day_vmt <- weighted_vmt %>%
  group_by(day_id) %>%
  summarize(weighted_vmt = sum(weighted_vmt)) %>%
  ungroup()
```

Just as with trip rates, travel days with no VMT - whether because the person did not travel by car, or because the person did not travel at all - are not represented in the trip table.

For this reason, weighted VMT estimates must be obtained by doing a join on the day table with zeros filled in for those days without auto travel.

```
person_day_vmt <- tbi$day %>%
  left_join(person_day_vmt, by = "day_id") %>%
  mutate(weighted_vmt = replace_na(weighted_vmt, 0))
```

Next, we:

- filter to weighted days;
- create a survey design object, specifying weights and strata;
- and use srvyr to calculate a weighted average of the weighted VMT divided by the day weight.

```
person_day_vmt %>%
  filter(day_weight > 0) %>%
  left_join(tbi$hh, by = "hh_id") %>%
  srvyr::as_survey_design(
    ids = day_id,
    weights = day_weight,
    strata = sample_segment
  ) %>%
  group_by(survey_year) %>%
  summarize(avg_vmt = srvyr::survey_mean(weighted_vmt / day_weight)) %>%
  ungroup() %>%
  gt::gt() %>%
  gt::fmt_number(columns = c(2:3), decimals = 1)
```

survey_year	avg_vmt	avg_vmt_se
2019	22.2	0.3
2021	15.0	0.5
2023	16.9	0.5

Additional Resources

These are just a handful of example analyses that can be done with the Travel Behavior Inventory dataset. Many interesting applications of the data await discovery by intrepid data explorers.

For more information about survey statistics and applications in R, consult *Sampling Design and Analysis* by Sharon Lohr (link).

Full documentation for the srvyr package, including example analyses, is available online at http://gdfe.co/srvyr/.

APPENDIX B. DOCUMENT LIBRARY

Dataset and Codebook

- FIXME - Link to dataset when published
- FIXME - Link to codebook when published
- Combined Codebook

Questionnaires

- Questionnaire 2021
- Questionnaire 2023
- Questionnaire Online 2019
- Questionnaire Smartphone 2019

Survey Administration

- Survey Management Plan 2021
- Study Design Experiment Results 2019
- Privacy Policy 2019
- Privacy Policy 2021

Invitation Materials

- Invitation 2021
- Invitation Envelope 2019
- Invitation Envelope 2023
- Invitation Language Insert 2019
- Invitation Letter 2019 Invitation Letter 2023
- Invitation Letter RAIO 2019
- Invitation Postcard 2019
- Invitation Postcard 2023
- Invitation Postcard RAIO 2019

Technical Documentation

- Survey Methodology Report 2019
- Survey Methodology Report 2021
- QAQC Plan 2021
- Technical Report 2021
- Re-Weighting Memo 2019 & 2021
- Weighting Memo 2023

Outreach Summaries

- Outreach Feedback Questionnaire 2021
- Outreach Focus Group Discussion Guide 2021
- Outreach New Publica Summary 2021

Workshops and Presentations

- Workshop Executive Summary 2019
- Workshop For Data Users 2019
- Workshop Agenda 2021
- Workshop Executive Summary 2021
- Workshop For Data Users 2021