

**GEORGIA DOT RESEARCH PROJECT 18-08**

**FINAL REPORT**

**Improvement of the Georgia Statewide Travel Demand  
Model (GSTDM) – Phase 2**



**OFFICE OF PERFORMANCE-BASED  
MANAGEMENT AND RESEARCH**

**600 WEST PEACHTREE STREET NW  
ATLANTA, GA 30308**

**TECHNICAL REPORT DOCUMENTATION PAGE**

<b>1. Report No.</b> FHWA-GA-21-1808		<b>2. Government Accession No.</b> N/A		<b>3. Recipient's Catalog No.</b> N/A	
<b>4. Title and Subtitle</b> Improvement of the Georgia Statewide Travel Demand Model (GSTDM) – Phase 2			<b>5. Report Date</b> September 2021		
			<b>6. Performing Organization Code</b> N/A		
<b>7. Authors</b> Giovanni Circella, Ph.D.; Sungtaek Choi, Ph.D.; Ali Etezady, Ph.D.; Alyas Widita, Ph.D.; Kara Todd, MSc.			<b>8. Performing Organization Report No.</b> 18-08		
<b>9. Performing Organization Name and Address</b> School of Civil and Environmental Engineering Georgia Institute of Technology 790 Atlantic Dr NW Atlanta, GA 30322			<b>10. Work Unit No.</b> N/A		
			<b>11. Contract or Grant No.</b> PI# 0016341		
<b>12. Sponsoring Agency Name and Address</b> Georgia Department of Transportation Office of Performance-based Management and Research 600 West Peachtree Street Northwest Atlanta, GA 30308			<b>13. Type of Report and Period Covered</b> Final Report (May 2019–September 2021)		
			<b>14. Sponsoring Agency Code</b> N/A		
<b>15. Supplementary Notes</b> Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.					
<b>16. Abstract</b> This report details a number of proposed improvements in the Georgia Statewide Travel Demand Model (GSTDM) using the 2017 National Household Travel Survey (NHTS) and its Georgia add-on portion. These improvements include (1) the development of a vehicle ownership model and a time-of-day segmentation, (2) estimating and evaluating a destination choice model, and (3) investigating a mode choice model for the GSTDM. Considering the importance of these model improvements, the research team conducted extensive reviews of the state of research and practice on the mentioned topics, augmented the 2017 NHTS data with other relevant data sources, developed appropriate methodologies, and presented the results and discussed their application in the GSTDM in this report.					
<b>17. Key Words</b> Georgia Statewide Travel Demand Model, GSTDM, National Household Travel Survey, NHTS, Travel Demand, Mode Choice, Vehicle Ownership, Destination Choice			<b>18. Distribution Statement</b> No restrictions.		
<b>19. Security Classification (of this report)</b> Unclassified		<b>20. Security Classification (of this page)</b> Unclassified		<b>21. No. of Pages</b> 173	<b>22. Price</b> Free

GDOT Research Project No. 18-08

Final Report

IMPROVEMENT OF THE GEORGIA STATEWIDE TRAVEL  
DEMAND MODEL (GSTDM) – PHASE 2

By

Giovanni Circella, Ph.D.

Sungtaek Choi, Ph.D.

Alyas Widita, Ph.D.

Ali Etezady, Ph.D.

Kara Todd, MSc.

Georgia Institute of Technology

Contract with  
Georgia Department of Transportation

In cooperation with  
U.S. Department of Transportation Federal Highway Administration

September 2021

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Georgia Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

## SI\* (MODERN METRIC) CONVERSION FACTORS

### APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
<b>LENGTH</b>				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
<b>AREA</b>				
in <sup>2</sup>	square inches	645.2	square millimeters	mm <sup>2</sup>
ft <sup>2</sup>	square feet	0.093	square meters	m <sup>2</sup>
yd <sup>2</sup>	square yard	0.836	square meters	m <sup>2</sup>
ac	acres	0.405	hectares	ha
mi <sup>2</sup>	square miles	2.59	square kilometers	km <sup>2</sup>
<b>VOLUME</b>				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft <sup>3</sup>	cubic feet	0.028	cubic meters	m <sup>3</sup>
yd <sup>3</sup>	cubic yards	0.765	cubic meters	m <sup>3</sup>
NOTE: volumes greater than 1000 L shall be shown in m <sup>3</sup>				
<b>MASS</b>				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
<b>TEMPERATURE (exact degrees)</b>				
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
<b>ILLUMINATION</b>				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m <sup>2</sup>	cd/m <sup>2</sup>
<b>FORCE and PRESSURE or STRESS</b>				
lbf	poundforce	4.45	newtons	N
lbf/in <sup>2</sup>	poundforce per square inch	6.89	kilopascals	kPa

### APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
<b>LENGTH</b>				
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
<b>AREA</b>				
mm <sup>2</sup>	square millimeters	0.0016	square inches	in <sup>2</sup>
m <sup>2</sup>	square meters	10.764	square feet	ft <sup>2</sup>
m <sup>2</sup>	square meters	1.195	square yards	yd <sup>2</sup>
ha	hectares	2.47	acres	ac
km <sup>2</sup>	square kilometers	0.386	square miles	mi <sup>2</sup>
<b>VOLUME</b>				
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m <sup>3</sup>	cubic meters	35.314	cubic feet	ft <sup>3</sup>
m <sup>3</sup>	cubic meters	1.307	cubic yards	yd <sup>3</sup>
<b>MASS</b>				
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
<b>TEMPERATURE (exact degrees)</b>				
°C	Celsius	1.8C+32	Fahrenheit	°F
<b>ILLUMINATION</b>				
lx	lux	0.0929	foot-candles	fc
cd/m <sup>2</sup>	candela/m <sup>2</sup>	0.2919	foot-Lamberts	fl
<b>FORCE and PRESSURE or STRESS</b>				
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in <sup>2</sup>

## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	1
CHAPTER 1. INTRODUCTION .....	7
Overview of the Research .....	9
Document Structure .....	11
CHAPTER 2. DATA EXPLORATION .....	12
NHTS Household Data .....	12
NHTS Person and Trip Data .....	14
CHAPTER 3. INCORPORATING VEHICLE OWNERSHIP MODELS IN THE GSTDM....	18
Literature Review .....	19
Data .....	20
Methodology .....	22
Behavioral Modeling Approach .....	22
Data-driven Modeling Approach .....	25
Results .....	26
Disaggregate Models .....	26
Aggregate Model .....	35
CHAPTER 4. INTRODUCING TIME OF DAY (TOD) IN THE GSTDM .....	37
Literature Review .....	37
TOD in Statewide Travel Demand Models .....	37
Classification of TOD Methods .....	40
Methodology .....	43
Overview .....	43
Benchmark Time for TOD .....	44
Trip Purpose .....	45
Results .....	46
Number of TOD Periods .....	46
Short-distance Trips .....	49
Spatial Distribution of Trips by Time Period .....	52
Long-distance Trips .....	63
Spatial Distribution of Trips by Time Period .....	68
Through Trips (External to External Trips) .....	71
TOD Implementation After Trip Generation .....	76

CHAPTER 5. EVALUATING THE INCLUSION OF A DESTINATION CHOICE MODEL.....	78
Current Status of Trip Distribution in the GSTDM.....	78
Review of Other Statewide Models.....	79
Data Preparation .....	80
Choice Set Formation .....	80
Dataset Augmentation.....	82
Method.....	84
Results .....	85
Home-based Work Models .....	85
Home-based Other Models .....	91
Nonhome-based Trip Models .....	95
Suggestions for Model Improvement .....	100
CHAPTER 6. DEVELOPMENT OF A MODE CHOICE MODEL.....	102
Composition of the Mode Choice Model .....	102
Short-distance Trips.....	103
Data Assembly Process.....	103
Model Specification.....	115
Explanatory Variables.....	117
Estimation Results .....	120
Long-distance Trips.....	124
Exploratory Analysis of 2017 Georgia NHTS Long-distance Trips .....	125
Augmenting Data with Comparable States.....	127
CHAPTER 7. ADDITIONAL IMPROVEMENT TO THE GSTDM: PROPOSING A TOUR-BASED APPROACH .....	132
Definition of Tours .....	133
Data Composition.....	135
Tour-related Variables .....	135
Logic of Defining a Primary Trip .....	137
Profiles of Tour-based Trips.....	137
Model Specifications.....	139
Overall Structure.....	140
Tour Model .....	140
Trip Model .....	141
Explanatory Variables.....	142
Combination Rules for Tour and Trip Modes .....	144

Estimation Result.....	145
Tour Model .....	146
Trip Model .....	147
Limitations on Application of the Tour-based Approach to the GSTDM.....	151
CHAPTER 8. SUMMARY AND CONCLUSIONS.....	153
APPENDIX.....	157
ACKNOWLEDGMENTS .....	158
REFERENCES .....	159

## LIST OF FIGURES

Figure 1. Bar graphs. Distribution of household vehicle ownership in 2017 NHTS Georgia portion (N=8,610).....	13
Figure 2. Bar graph. Age category in the short-distance trips dataset (2017 NHTS) compared to ACS 2018 and Decennial Census 2010.....	15
Figure 3. Bar graph. Race category in the short-distance trips dataset (2017 NHTS) compared to ACS 2018 and Decennial Census 2010.....	15
Figure 4. Area graph. Median income in the short-distance trips dataset (2017 NHTS) compared to ACS 2018. ....	16
Figure 5. Area graph. Average household size in the short-distance trips dataset (2017 NHTS) compared to ACS 2018.....	17
Figure 6. Diagram. Schematic of the latent class of this study.....	29
Figure 7. Bar graph. Feature importance plot of the random forest model. ....	33
Figure 8. Map. States with TDM's that account for time of day.....	38
Figure 9. Diagram. Comparison of trips in motion approach to traditional time period classification.....	45
Figure 10. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, short-distance trips.....	47
Figure 11. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach, short-distance trips.....	47
Figure 12. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, all trips. ....	48
Figure 13. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach, all trips. ....	48
Figure 14. Pie graph. Shares of total short-distance trips by purpose.....	49
Figure 15. Bar graphs. Shares of total short-distance trips by mode and purpose.....	50
Figure 16. Pie graph. Shares of total short-distance trips by time period (all purposes).....	51
Figure 17. Bar graph. Share of short-distance trips by trip purpose and directionality.....	53
Figure 18. Maps. Spatial distribution of generated HBW short-distance trips at the census tract TAZ level during the AM peak period.....	55
Figure 19. Maps. Spatial distribution of generated HBW short-distance trips at the census tract TAZ level during the PM peak period.....	56



Figure 20. Maps. Spatial distribution of generated HBO short-distance trips at the census tract TAZ level during the AM peak period.....	58
Figure 21. Maps. Spatial distribution of generated HBO short-distance trips at the census tract TAZ level during the PM peak period.....	59
Figure 22. Bar graph. TOD factors, short-distance trips (weighted). .....	62
Figure 23. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, long-distance trips.....	64
Figure 24. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach for, long-distance trips. ....	65
Figure 25. Pie graph. Share of total long-distance trips by purpose. ....	66
Figure 26. Bar charts. Share of total long-distance trips by mode and purpose. ....	67
Figure 27. Bar charts. Share of total long-distance trips by directionality. ....	69
Figure 28. Bar charts. TOD factors, long-distance trips (weighted).....	70
Figure 29. Map. Location of traffic count stations in Georgia. ....	75
Figure 30. Screenshot. Data format of hourly averages report from Georgia state border in 2017. ....	75
Figure 31. Flowchart. TOD implementation in the current GSTDM. ....	76
Figure 32. Flowchart. Proposed method of the TOD implementation in GSTDM. ....	77
Figure 33. Bar graph. Distribution of the U.S. statewide travel demand models based on their trip distribution model.....	79
Figure 34. Stacked bar graphs. Share of mode choices for short- and long-distance trips by trip purpose.....	103
Figure 35. Classification chart. Linking mode classification between NHTS and GSTDM.....	105
Figure 36. Scatter plot. AllTransit score for the origin and destination combination by mode (auto and transit), short-distance trips. Larger values denote greater transit presence.....	108
Figure 37. Area graph. Density distribution of travel time between NHTS and Google API. ....	108
Figure 38. Area graphs. Density distribution of travel time from NHTS add-on data, short-distance trips.....	109
Figure 39. Area graphs. Density distribution of travel time derived from Google API, short-distance trips.....	110
Figure 40. Histogram. Estimated travel cost for auto and transit. ....	113
Figure 41. Pie graphs. Share of mode choices for short-distance trips by trip purpose.....	115

Figure 42. Model diagram. Mode choice set with nested structure. ....	116
Figure 43 Stacked bar graph. Mode share distribution of long-distance trips based on location in GA. ....	125
Figure 44. Stacked bar graph. Mode share for all long-distance trips by trip purpose in GA. ....	126
Figure 45. Diagrams. Trip flows by rural-urban continuum area classification for select states that have NHTS add-on data component. ....	128
Figure 46. Stacked bar graph. Mode share for all long-distance trips by trip purpose in SC. ....	130
Figure 47. Diagram. Example of a home-based tour. ....	134
Figure 48. Diagram. Composition of a home-based tour. ....	135
Figure 49. Screenshot. <i>Hometrip</i> variable in the dataset (example from R Studio). ....	136
Figure 50. Screenshot. Data format of the tour-based model (example from R Studio). ....	136
Figure 51. Bar graph. Shares of tour trips with a single trip leg by primary mode. ....	138
Figure 52. Histograms. Distributions of the number of trip legs by trip purpose. ....	139
Figure 52. Model diagram. Mode choice set with a nested structure (tour model). ....	141
Figure 53. Model diagram. Mode choice set with a nested structure (trip model). ....	141

## LIST OF TABLES

Table 1. Summary statistics of the sample households derived from the 2017 Georgia NHTS (n = 8,611). .....	14
Table 2. Multinomial logit regression (N=6205). .....	28
Table 3. Membership model parameters and descriptive statistics.....	30
Table 4. Latent class multinomial regression (N=6205).....	32
Table 5. Prediction accuracy across models. ....	34
Table 6. Linear regression predicting household vehicle count (N=8307).....	36
Table 7. A comparison of predicted vehicle counts and observations from the ACS. ....	36
Table 8. Summary of TOD implementation in statewide travel demand models.....	42
Table 9. Trip shares by time period and purpose (short-distance trips).....	52
Table 10. Number of generated HBW trips by directionality, short-distance trips. ....	57
Table 11. Number of generated HBO trips by directionality, short-distance trips. ....	60
Table 12. Number of generated NHB trips, short-distance trips. ....	60
Table 13. TOD factors, short-distance trips (weighted). ....	63
Table 14. Trip shares by time period and purpose (long-distance trips). ....	68
Table 15. TOD factors, long-distance trips (weighted). ....	71
Table 16. 2016 Daily factors by road hierarchy in Georgia. ....	73
Table 17. 2016 monthly factors by road hierarchy in Georgia. ....	73
Table 18. List of traffic count stations for external trips. ....	74
Table 19. Final TOD factors for through trips.....	76
Table 20. Destination choice model for the HBW, AM peak trips.....	87
Table 21. Destination choice model for the HBW, PM peak trips. ....	89
Table 22. Destination choice model for the HBW, Mid-day period trips.....	90
Table 23. Destination choice model for the HBW, Night period trips. ....	91
Table 24. Destination choice model for the HBO, AM peak trips. ....	92
Table 25. Destination choice models for the HBO, PM peak trips.....	93
Table 26. Destination choice model for the HBO, Mid-day period trips. ....	94

Table 27. Destination choice model for the HBO, Night period trips. ....	95
Table 28. Destination choice model for the NHB, AM peak trips. ....	97
Table 29. Destination choice model for the NHB, PM peak trips. ....	98
Table 30. Destination choice model for the NHB, Mid-day period trips. ....	99
Table 31. Destination choice model for the NHB, Night period trips. ....	100
Table 32. Transit fare in Georgia. ....	113
Table 33. Mode attribute variables of the mode choice model. ....	118
Table 34. Socioeconomic variables of the mode choice model. ....	119
Table 35. Accessibility variables of the mode choice model. ....	120
Table 36. Mode choice model for all purposes (nested logit form). ....	122
Table 37. Comparison of VOTTS by model. ....	124
Table 38. Distribution of long-distance trip distances by mode of travel in GA. ....	127
Table 39. Long-distance trip distance distribution by mode of travel in SC. ....	131
Table 40. Mode attribute variables of the tour-based mode choice model. ....	143
Table 41. Socioeconomic variables of the tour-based mode choice model. ....	144
Table 42. Accessibility variables of the mode choice model. ....	144
Table 43. Combination rules for tour and trip modes. ....	145
Table 44. Tour-based mode choice model (tour model). ....	147
Table 45. Tour-based mode choice model (trip model). ....	150

## EXECUTIVE SUMMARY

This report discusses a number of areas for proposed improvements in the Georgia Statewide Travel Demand Model (GSTDM) based on the analysis of the 2017 National Household Travel Survey (NHTS) and its Georgia add-on portion. These improvements include: (1) development of a vehicle ownership model, (2) development of a time-of-day segmentation, (3) estimation and evaluation of a destination choice model, and (4) investigation of approaches for the inclusion of a travel mode choice model in the GSTDM. Considering the importance of these areas for model improvement, the research team conducted extensive reviews of the state of research and practice on the mentioned topics, augmented the 2017 NHTS data with other relevant data sources, developed appropriate methodologies, and presented the results and discussed their applicability in the GSTDM in this report.

In the first task of this project, the research team developed a set of vehicle ownership models based on analysis of the 2017 NHTS Georgia add-on data. Considering that the 2017 NHTS records vehicle ownership at the household level and taking into consideration the behavioral importance of the factors influencing households' vehicle-ownership decisions, the research team first developed a set of disaggregate models using both discrete choice (behavioral) modeling and data-driven modeling approaches. We identified the number of drivers/workers in a household in addition to income level, race, household composition, and built environment to be among the most influential factors influencing households' vehicle-ownership decisions. Considering the aggregate structure of GSTDM, moreover, the team estimated a linear regression model whose results could be aggregated to the traffic analysis zone (TAZ) used in the GSTDM, which was informed by the more detailed disaggregate models. The output of this aggregate

model is the average vehicle ownership per each TAZ in the GSTDM, i.e., an outcome measure that can be readily incorporated in the other steps of the model (trip generation, distribution, and mode choice) to help improve model accuracy and sensitivity.

In the second task of this project, the research team investigated methods to introduce time of day in the GSTDM. The latest version of the GSTDM does not include a time-of-day classification of trips, and only applies time of day as a postprocessing step following the trip assignment. After reviewing the different time-of-day methodologies in the literature, the research team selected a “trips-in-motion” approach where trips that span more than one time-of-day period are accounted for more properly, as the most appropriate method to consider the impact of the time of trips on traffic congestion conditions during the various times of the day. Analyzing the temporal distribution of the 2017 NHTS data, we proposed four time-of-day periods, namely the AM peak (6 AM–10 AM), Midday (10 AM–3 PM), PM peak (3 PM–7 PM), and Night (7 PM–6 AM) periods. For each of these time periods, then, we computed the shares of trips, or time-of-day factors, by each period and trip purpose. Incorporating the proposed time-of-day classification in the GSTDM will help generate a more realistic temporal representation of trips in the modeling process and will increase the overall modeling accuracy and sensitivity.

In the third task of this project, the research team aimed to improve the trip distribution step in the GSTDM. This is a particularly important task as trip distribution is one of the largest sources of error in travel demand modeling. A gravity model is currently used to distribute trips in the GSTDM, and the research team evaluated the inclusion and performance of a full destination choice model in the GSTDM. We used multiple data sources to complement the 2017 NHTS dataset, and estimated 12 destination choice models, one for each time of day (AM peak, Midday, PM peak, Night) and trip purpose (HBW, HBO, NHB). We found that variables such as

distance, income, vehicle ownership, TAZ size variables (employment and population), and unique geographical features (such as parks and trails, presence of airports, colleges, and military bases) are influential factors in the destination choice. The models, in addition, showed promising results, and provide more flexibility in including socioeconomic characteristics in trip distribution. We further provided guidance on improving the models' performances in the future with more detailed data.

In the fourth task of this project, the research team evaluated the estimation of a mode choice model for the GSTDM. Considering the different nature of long- vs. short-distance trips, we segmented the data based on an already-defined criterion in the GSTDM: trips longer than 50 miles are categorized as long-distance trips, and those shorter than 50 miles are categorized as short-distance trips. For short-distance trips, we constructed a mode choice dataset by extracting data for the possible alternative modes for each trip using information obtained from the Google API, and further augmented the dataset using other sources such as AllTransit data. We tested multiple model structures to get the best short-distance mode choice model, including multinomial logit (MNL) and nested logit models, and found that trip-specific variables such as travel time (in-vehicle and out-of-vehicle) and travel cost, socioeconomic variables such as vehicle availability, and transit accessibility influence mode choice decisions. For long-distance trips, however, the team could not estimate a satisfactory model because of the small number of long-distance trips in the dataset. We, however, provided exploratory insights into long-distance trip patterns, and discussed recommendations on remedying the lack of long-distance data such as merging data from other U.S. states. We discuss similarities of long-distance travel patterns to those in the state of Georgia in a few other U.S. states, and suggest a list of candidate states for

the analysis of long-distance trip patterns among the states that also participated to the NHTS add-on program.

In the final task of the report, the research team also analyzed travel mode choice decisions at the tour level. The reason for this investigation is that individual trips are usually made as part of larger tours. Accordingly, analyzing trip mode choice at the trip level might lead to misleading results as the decision on what travel mode to use for a specific trip is usually conditional on the characteristics of the larger tour. Hence, the research team investigated the mode choice decision in the 2017 NHTS data using a tour-based modeling approach. In order to do that, we conducted a literature review on how to define a tour, and presented the best ways to categorize tours, define primary tour purposes, and structure the tour and trip mode choice models. While the implementation of a tour-based mode choice model is currently not feasible in an aggregate travel demand model such as the GSTDM, the tour-based mode choice results highlighted several important implications. One of these relates to the more reasonable value of travel time that is obtained from the tour-based mode choice model, compared to the trip-based model, which would encourage exploring further modeling improvement in the GSTDM toward more disaggregate approaches that could allow harvesting these behavioral details and would improve the ability to properly model travel demand in the state.

As a result of the work conducted in this project, the research team presents four important recommendations and take-aways from this research:

- The research team recommends that GDOT incorporates a vehicle ownership model in the GSTDM, replacing the current simplified approach that is included in the existing model. The vehicle ownership analyses presented in this report, carried out at the



disaggregate and aggregate (TAZ-level) levels, can inform this task. In particular, the TAZ-level vehicle ownership model estimated in this study is ready to be implemented in the GSTDM framework, to better support the trip generation, trip distribution, and mode choice steps of the model.

- The research team recommends that GDOT incorporates the time-of-day segmentation in the GSTDM. Based on the results from this study, implanting the time-of-day segmentation following the trip-generation step seems an appropriate approach. The four time-of-day periods developed in this study can greatly help in more realistically modeling the trip distribution, mode choice, and assignment models, where time of day often significantly impacts the travel patterns. Further, time of day is fundamental when evaluation traffic congestion on the road network during the various times of the day.
- The research team recommends that GDOT further explores the inclusion of a proper destination choice model for short-distance trips in the GSTDM. This modification appears justified by the large number of short-distance trips in the state, whereas keeping a gravity model for the long-distance trips appears appropriate. This report provides a detailed description on how to develop a destination choice model for short-distance trips in the GSTDM, and the models (by trip purpose and time of day) estimated in this study showed promising results. These models may also be further enhanced based on some of the recommendations included in this report.
- The research team recommends that GDOT further explores the inclusion of an improved mode choice model in the GSTDM. While this study evaluated both the estimation of a trip-based and a tour-based mode choice model for short-distance trips in the state of

Georgia, the results could be used to inform the development of an aggregate mode split component for the GSTDM.

- Future improvements in the GSTDM could explore the possibility to upgrade the modeling framework from a trip-based to a tour-based or activity-based modeling approach, similar to what has been done by other U.S. states (e.g., California). This model upgrade could enable the development of more detailed and disaggregate model components, which could more carefully capture the nuanced nature of travel demand-related decisions and their impacts on traffic patterns and investment decisions.

## CHAPTER 1. INTRODUCTION

The Georgia Department of Transportation (GDOT), in collaboration with its consultants, has developed a statewide travel demand model to assist with the formulation of statewide transportation plans. The Georgia Statewide Travel Demand Model (GSTDM) incorporates both freight and passenger travel demand forecasting components, and serves a variety of purposes, including, but not limited to, the estimation of intercity passenger and truck travel volumes, interstate and state highway corridor volumes, changes in travel flows on major corridors due to changes in land use or economic policies, etc. The model is quite comprehensive and serves as an effective planning tool for the state (Peevy and Kassa 2012). The GSTDM is maintained and updated by the technical staff from the GDOT Office of Planning in cooperation with a team of consultants using updated information about transportation patterns, sociodemographic data, and observed traffic flows available from multiple sources. These sources include other state and federal agencies and local metropolitan planning organizations (MPOs). The current version of the model covers the entire 48 continental U.S. states and includes 3,770 traffic analysis zones (TAZs), of which 3,243 are in Georgia. The highway network includes a total of 80,400 miles, of which 18,600 are in the state of Georgia. With the recent updates introduced in the GSTDM, the base year has been updated to 2015 as part of the maintenance program carried out by a team of GDOT consultants.

The current maintenance and updates to the GSTDM, however, do not encompass a variety of modern solutions that have been developed in statewide models to more accurately predict travel demand. As an example, the Travel Forecasting Resource (TFR) online repository<sup>1</sup> provides an

---

<sup>1</sup> [https://tfresource.org/topics/Statewide\\_models.html](https://tfresource.org/topics/Statewide_models.html)

overview of the state of the practice in statewide modeling and provides a map depicting statewide model development efforts across the nation. Many of these modern developments and solutions use realistic travel behavior assumptions and up-to-date data sources that help model a variety of transportation services and options with solutions that are cost effective for a statewide model.

The availability of the 2017 National Household Travel Survey (NHTS) data and the exclusive Georgia add-on that was funded by the Georgia Department of Transportation, specifically, provides a prime opportunity for GDOT to employ the up-to-date datasets in conjunction with more sophisticated approaches to upgrade the current GSTDM and improve the model specifications to better forecast travel demand and traffic patterns in the state.

The principal investigator of this project, along with two of the research team members, worked at a previous “Phase 1” study (GDOT Research Project 16-12, PI: Dr. Circella) that helped the GDOT Office of Planning integrate the GSTDM with the regional models in the state, making the statewide model consistent with the regional models used by the 14 MPOs in Georgia whose models are directly maintained by GDOT.

Phase 2 of the study, undertaken in this project, and summarized in this report, explores the development of several improvements in other components of the GSTDM, including improving the understanding of how vehicle ownership varies by sociodemographic characteristics and geographic location, adding an improved temporal resolution of the GSTDM (introducing the “time of day” in which trips are modeled, instead of the 24-hour average travel forecasts produced by the current model), harvesting the opportunities offered by recently collected data sources, including the new NHTS, to improve the representation of transportation infrastructure

and services, and improving the trip distribution and mode choice components of the model. There is a critical need to integrate these improvements in the newer version of the statewide travel demand model to produce better travel forecasts and inform transportation investment decisions in the state. This report discusses some opportunities and provides recommendations for such model improvements.

## **OVERVIEW OF THE RESEARCH**

Under the activities of this project, the research team worked closely with the GDOT Office of Planning to explore several areas for the GSTDM improvement. We designed five main tasks to achieve this goal:

**Task 1.** Investigate improved approaches to account for vehicle availability in various geographic regions in the state, among different socioeconomic and demographic groups, and in the presence of various land use/neighborhood types in the GSTDM model. As part of this task, the project team reviewed the approaches used in other statewide and regional travel demand models to account for vehicle ownership, focusing in particular on applications to four-step models that can be more similar to the GSTDM modeling framework. The research team will build on the preliminary findings from the GDOT Research Project, “Analysis of the Georgia Add-on to the 2016–2017 National Household Travel Survey” and further analyze the 2017 NHTS add-on data for Georgia, with the aim of estimating a vehicle ownership model that can account for variation in household vehicle ownership by geographic region, neighborhood type, and SE characteristics.

**Task 2.** Investigate and introduce time of day in the GSTDM. In this task, the project team reviewed the modeling approaches adopted in other regional and statewide travel demand

forecasting models, and proposed a modeling solution that will model travel demand in the GSTDM framework separately for four main time periods: AM peak, Midday, PM peak, and Late evening/Off-peak. The existing GSTDM only produces 24-hour daily travel forecasts, and GDOT needs postprocessing methods to achieve specific time-of-day outputs. The revised approach of this task, therefore, presents a major improvement to the GSTDM, and will align it with other state-of-the-art statewide modeling frameworks.

**Task 3.** Investigate and evaluate the inclusion of a destination choice model to replace the current Gravity trip distribution model in the GSTDM. In this task, we reviewed the existing modeling approaches used in other statewide and regional four-step travel demand models. We analyzed the 2017 NHTS add-on data for Georgia and other available data sources, with the aim of analyzing trip destination patterns by time of day and trip purpose, and proposed ways to implement such improvements in the trip destination modeling processes in the GSTDM.

**Task 4.** Evaluate the inclusion of a mode choice component in the GSTDM framework. The current version of the GSTDM does not properly account for travelers' choices on the travel modes that are used for various trips. In this task, the research team reviewed the existing literature and the modeling approaches used in other statewide and regional travel demand models to account for travel mode choice. We analyzed data from the 2017 NHTS add-on data for Georgia, complemented the NHTS add-on data with additional information for commuting and noncommuting trips (e.g., computing travel time and distances for unchosen alternatives) and estimated a mode choice model that accounts for the impacts of socioeconomic variables, trip characteristics, geographic location and land use variables on the travel mode choice.

**Task 5.** Investigate additional improvements to the GSTDM. Based on the results of task 4, this task explored the estimation of a tour-based mode choice model as opposed to the trip-based model of the previous task. The logic and benefits of defining tours and incorporating them in the mode choice model are discussed, and model development steps, results, and insights are presented.

## **DOCUMENT STRUCTURE**

This report is structured around each of the tasks discussed in Overview of the Research above. Before we begin discussion of the tasks, we provide a brief exploration of the main dataset used in this project in chapter 2. In each of the subsequent chapters, then, we detail the steps taken to achieve the goals of one of the tasks described above and present the results and possible implementation guidelines for the GSTDM. We end the report with our conclusions in chapter 8, and provide a summary of the developed analyses and associated recommendations for further improvement of the GSTDM.

## CHAPTER 2. DATA EXPLORATION

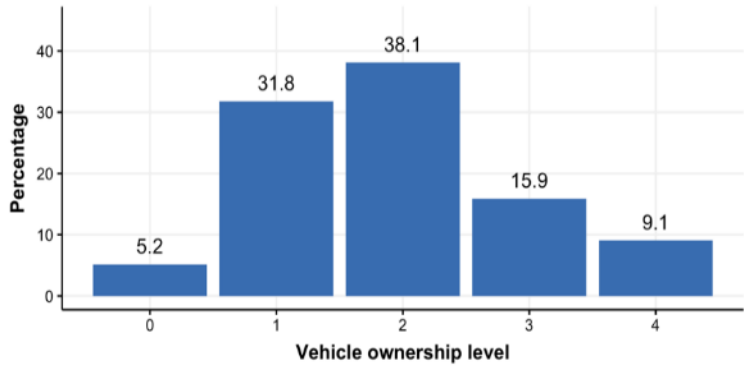
Although GDOT Report 18-24 (Kash, Mokhtarian, and Circella 2021) provides an extensive exploration and descriptive statistics for the 2017 NHTS Georgia add-on data, this chapter provides some specific data exploration more related to the goals of this report. For a complete descriptive statistic of the 2017 NHTS Georgia add-on data, therefore, we refer interested readers to Kash, Mokhtarian, and Circella (2021).

The 2017 NHTS *household-level* dataset is the primary data for estimating the vehicle ownership models in this project (task 1), while the *person-* and *trip-level* data were used in the other tasks. Consequently, in this chapter we first present the characteristics of the household-level data, and then discuss the person- and trip-level data. Each chapter then contains additional descriptions of more specific data sources and considerations related to the analysis done in that chapter.

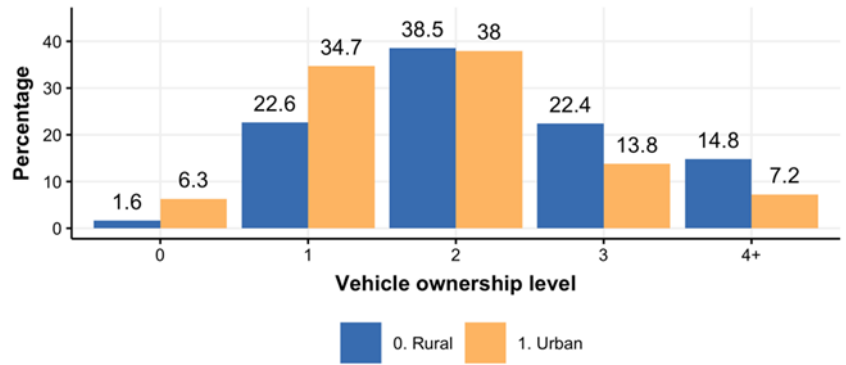
### NHTS HOUSEHOLD DATA

Figure 1(a) shows a distribution of this variable in the Georgia portion of the 2017 NHTS. The share of households with no vehicles is the smallest at 6.9 percent, while households with 1 to 2 vehicles constitute the majority of the sample households. The overall average vehicle ownership (VO) in the state of Georgia is 1.92 vehicles per household.





(a)



(b)

**Figure 1. Bar graphs. Distribution of household vehicle ownership in 2017 NHTS Georgia portion (N=8,610).**

Figure 1(b) investigates how VO levels differ based on built environment characteristics. As expected, we can see that the urban Georgia households tend to own fewer vehicles than their rural counterparts. The shares of households with 0 or 1 vehicle in their household are considerably lower in rural areas, while the shares of 3- and 4+-vehicle households in such areas are considerably higher. Overall, and on average, the average VO in urban areas is approximately 1.80 vehicles, while the average VO in rural areas approximately equals 2.35.

Furthermore, table 1 presents the summary statistics of the dataset derived from the NHTS. As shown in the table, it appears around a third of the sample households can be considered as low income; whereas, approximately 22 percent of the sample households fall in the high-income category. For the variables representing household composition, it appears that 19 percent of the sample is a single-person household and another 20.8 percent is a two or more persons household with no children. To account for the adoption of emerging travel mode, we incorporate a count indicator of the frequency of using taxi or ride hailing in the last week. The mean of this variable is 0.42, indicating that only a fraction of the sample is a frequent patron of such services.

**Table 1. Summary statistics of the sample households derived from the 2017 Georgia NHTS (n = 8,611).**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Low HH* Income (<\$35k), binary	0.325	0.468	0	1
High HH Income (>\$100k), binary	0.221	0.415	0	1
1 Person HH – No Children, binary	0.190	0.392	0	1
2+ Person HH – No Children, binary	0.208	0.406	0	1
No. of HH Workers	0.971	0.866	0	5
No. of Drivers	1.660	0.765	0	7
Frequency Taxi/Ridehailing	0.423	1.883	0	20

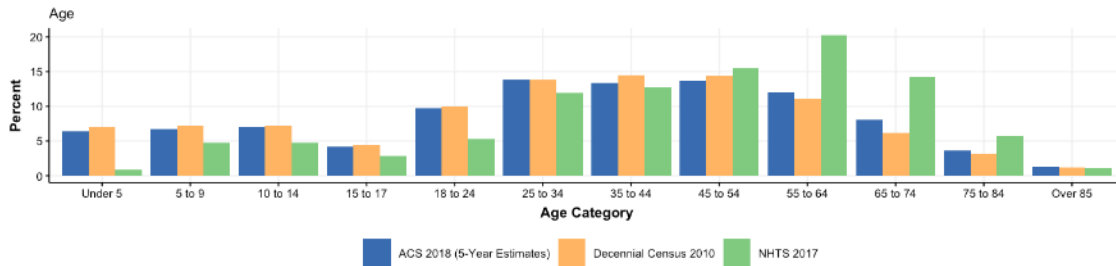
\* HH = Household

## **NHTS PERSON AND TRIP DATA**

The investigations in chapter 4, chapter 5, and chapter 6 are primarily of the short-distance trips portion of the NHTS; accordingly, this section provides a number of descriptive statistics on the short-distance data. The research team selected a number of indicators that could be evaluated

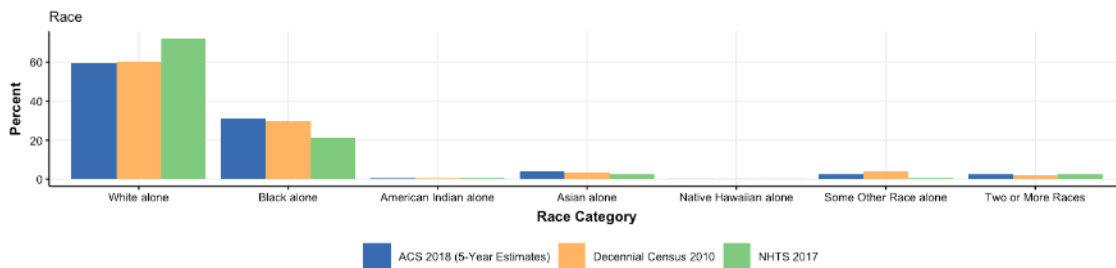
among the 2017 NHTS, 2014–2018 American Community Survey (2018 ACS, 5-year estimates), and the Decennial Census 2010 data, in order to provide a more complete picture of the dataset.

One of the indicators the research team investigated was age category. As shown in figure 2, the NHTS appears to have a higher share of elderly population but lower share of the younger population. This might not be surprising since the NHTS targets travelers capable of transporting themselves and, thus, the portion of younger individuals is substantially lower than the population.



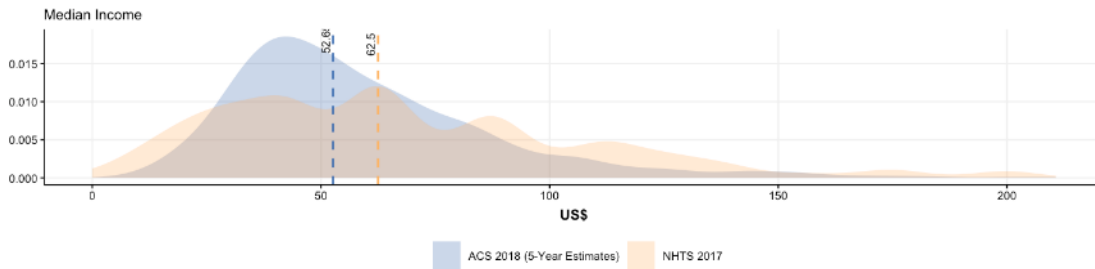
**Figure 2. Bar graph. Age category in the short-distance trips dataset (2017 NHTS) compared to ACS 2018 and Decennial Census 2010.**

Another indicator that the research team investigated was the race category. As shown in figure 3, the NHTS has a higher share of White individuals and lower share of African American individuals than the Census.



**Figure 3. Bar graph. Race category in the short-distance trips dataset (2017 NHTS) compared to ACS 2018 and Decennial Census 2010.**

Figure 4, moreover, depicts the distribution of median income between the 2017 NHTS and the ACS 2018.<sup>2</sup> As shown in figure 4, the samples in the short-distance trips dataset tend to have higher median income than the population estimates based on the ACS 2018. Specifically, the median income in the short-distance trips dataset was \$62,500 while the median income according to population estimates was \$52,600.

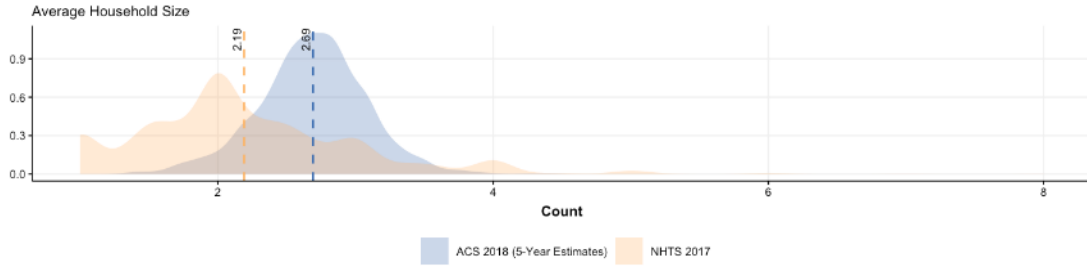


**Figure 4. Area graph. Median income in the short-distance trips dataset (2017 NHTS) compared to ACS 2018.**

Figure 5 shows an additional dataset assessment indicating that the samples in the short-distance trips dataset tend to have lower average household size than the population estimates derived from the ACS 2018.

---

<sup>2</sup> The median income indicator is not reported in the Decennial Census.



**Figure 5. Area graph. Average household size in the short-distance trips dataset (2017 NHTS) compared to ACS 2018.**

### **CHAPTER 3. INCORPORATING VEHICLE OWNERSHIP MODELS IN THE GSTDM**

Understanding factors associated with vehicle ownership is of critical importance to travel demand models (TDMs), where trip generation, distribution, and mode choice can be directly affected by this household-level variable. Considering the importance of vehicle ownership models, and the current absence of such a model in the GSTDM, this task aims to investigate and develop a set of VO models based on the 2017 NHTS for the state of Georgia. To this aim, we develop and estimate two categories of models: one based on the disaggregate, household-level data available in the 2017 NHTS, and one based on aggregate TAZ-level units that may be directly used in the GSTDM. Disaggregate models are able to use the full potential of the available data and present a more detailed and behaviorally explainable choice of vehicle-ownership levels among the Georgia households. The downside of these models, however, is their incompatibility with the GSTDM model structure, which is based on aggregate-level TAZ units. We, therefore, first estimate disaggregate models (behavioral and data-driven) to gain a better insight into how the Georgian households make decisions regarding VO, and then use the disaggregate models and translate their structure and results to be compatible with the GSTDM TAZ-based structure.

An initial investigation of vehicle ownership in the 2017 NHTS Georgia add-on data is presented in GDOT Report 18-24 (Kash, Mokhtarian, and Circella 2021). The findings in chapter 1 of Report 18-24, which aided the modeling work presented in this chapter, shed light on vehicle availability, usage, and fleet characteristics among Georgian households, and provide a complementary read to the results and discussions presented here.

## LITERATURE REVIEW

As one of the most researched topics in transportation literature, a wealth of studies exist on vehicle ownership. In this literature review, we will focus on the current studies that we consider most relevant to the purpose of this work.

Several studies on vehicle ownership in recent years have used behavioral class models. For instance, a 2013 study exhibits the application of a latent class multinomial logit (MNL) model to estimate factors associated with vehicle type ownership (Beck et al. 2013). Using data from Sydney, Australia, which was collected through an interviewer-assisted online survey platform, their study seeks to predict a given respondent choice of vehicle type, i.e., petrol, diesel, or hybrid. In doing so, the authors use a latent class approach to deal with the “preference heterogeneity across classes” as an attempt to account for the unobserved factor. With emphasis on assessing the influence of attitudinal factors, the results suggest the relative importance of these factors and, therefore, policies aimed to promote increasing adoption of environmentally friendly vehicles need to account for inducing potential attitudinal changes.

The application of the latent class approach was also the highlight of a 2014 study analyzing car ownership in Quebec City, Canada (Anowar et al. 2014). Using a family of behavior class models in the form of latent segmentation-based ordered logit (LSOL) and latent segmentation-based multinomial logit (LSMNL), the authors present estimation results indicating the factors associated with household-level car ownership, i.e., no car, one car, two cars, or more, and test which latent class model performs relatively better than the other. The results from latent class analysis indicate two segments of the population captured in the data: *transit independent (TI)* and *transit friendly (TF)*. Estimation results suggest that several exogenous factors contribute to

increased car ownership level, e.g., higher number of employed adult household members, lower number of children, and lower residential density in the neighborhood the households live in. Moreover, the authors discover that the LSMNL model performs slightly better than the LSOL model.

In addition to latent class approach as a subset of behavioral class models as presented above, several studies have used the application of mixed logit models (MLMs). This approach is the highlight of a 2003 study assessing the household automobile transactions using data from the Toronto Area Car Ownership Study (TACOS) (Mohammadian and Miller 2003). While not necessarily related to vehicle ownership, a comparison between a traditional multinomial logit model, a mixed logit model, and a latent class model (LCM) was the centerpiece of another 2003 study and, therefore, is important to discuss here (Greene and Hensher 2003). In their study, the authors conduct a comparison between the models using data from New Zealand derived from a survey asking car drivers their preferred road environment for long-distance trips. Estimation results indicate that both MLM and LCM outperform the MNL by virtue of evaluating the log-likelihood indicators. Assessing the preferred model between MLM and LCM, however, presents a rather impossible task since “each has its own merits.” From a purely numerical indicator and perhaps only “on this occasion,” however, it appears that LCM offers a “stronger statistical support” than the MLM (Greene and Hensher 2003).

## **DATA**

We presented an exploratory analysis of the NHTS household level data in chapter 2. This section provides more specific data sources and considerations related to the task described in this chapter.



Considering that our final goal for this task is to estimate a vehicle ownership model that can be used in an aggregate model, the research team needed to be able to translate the models based on the disaggregate NHTS data to the aggregate GSTDM structure. In doing so, one of the primary considerations for the research team was matching available NHTS variables with their aggregate counterparts in the 2018 ACS dataset. Following this consideration, and after comparing the available variables in both datasets, the indicators employed in the aggregate linear regression model are *income*, *household size*, *number of employed household members*, and *housing density*.

The research team also used external datasets to complement the 2017 NHTS data. These additional variables are indices representing transit services in a given block group derived from the AllTransit data provided by the Center for Neighborhood Technology. These indicators include the Transit Connectivity Index and the AllTransit performance score. The Transit Connectivity Index is defined as (AllTransit 2018):

*“...the sum of buses/trains per week scaled by overlap of 1/8-mile rings and weighted for each ring (6 for bus and rail, or 3/4 mile) for every stop whose ring intersects the block group. The scaling was optimized by using regression to fit for percent of transit used for journey to work. The result is scaled from 0–100, with zero being no transit and 100 being the best block group in the county.”*

Along a somewhat similar line, the AllTransit performance score is defined as:

*“...a comprehensive score that looks at connectivity, access to land area and jobs, frequency of service, and the percent of commuters who use transit to travel to work. While availability of service and frequency are important aspects of transit,*

*the connection it provides to jobs and other destinations in the region is central in creating an effective transit system.”*

## **METHODOLOGY**

In this section, we elaborate on the methodologies for each model family focusing on the traditional MNL and latent class MNL as the estimation approach under the behavioral-class, and random forest as a family of data-driven, machine learning techniques.

### **Behavioral Modeling Approach**

Behavioral models have long been developed and used in the econometric and, subsequently, travel behavior studies. Depending on the type of the variable of interest, there are several econometric models that can be employed for analysis. Vehicle ownership can be assumed as either a continuous numeric variable, a count variable, or a discrete variable. Studies point to the discrete assumption for the vehicle ownership as a more appropriate variable type assumption and, therefore, recommend conducting vehicle ownership analyses on a discrete choice model (Bhat and Pulugurta 1998).

When considering vehicle ownership as a continuous numeric variable, a linear regression modeling (LRM) framework can be used. Since LRMs are linear in nature, we can aggregate their results to the TAZ level for the GSTDM structure. In contrast, dynamic causal models (DCM) follow a nonlinear structure, and aggregation of their structure would not yield accurate unbiased estimates. We, therefore, first estimate DCM-based models to better capture and explain households' choices, and then develop an LRM to be used further for the GSTDM purposes.

### ***Linear Regression Models***

Linear regression models are perhaps the most simple and well-known modeling technique used in transportation literature and other fields. LRM is a suitable technique when the variable being modeled is continuous (i.e., numeric). In our context, we can consider the number of vehicles owned by a household (0, 1, 2, 3, etc.) as continuous, and use this modeling framework to find the factors influencing a household vehicle count. Equation 1 shows a general formulation of an LRM:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots = \sum_k X_{ik} \beta_{ik} \quad (1)$$

In equation 1,  $y_i$  is the dependent variable's value associated with an individual  $i$ ,  $\beta_k$  is the coefficient associated with variable  $k$ , and  $X_{ik}$  is the value of variable  $k$  for individual  $i$ .

Ordinary least squares (OLS) is used to find the unbiased and efficient coefficients of equation 1.

### ***Discrete Choice Models***

Discrete choice models have been widely used in the transportation literature, in addition to other fields, to model discrete choices or categorical variables. This class of models is based on the definition of a utility for each choice or level of categorical variable and uses the utility parameter to assess the probability by which an option is chosen. Equation 2 shows the general definition of a utility function for a choice or level of categorical variable:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n} \quad (2)$$

In equation 2,  $U_{i,n}$  denotes the utility of option  $i$  for person  $n$ ,  $V_{i,n}$  denotes the deterministic portion of the utility function for the option, and  $\varepsilon_{i,n}$  is the random portion (error term) of the

function. The deterministic portion of the utility is often modeled as a linear-in-parameter function of observed variables in the data ( $\sum_k X_k \beta_k$ ). Assuming a Gumbel distribution for the error term in the utility function, we may compute the probability of choosing option  $i$  using a logit formulation:

$$p(k = i) = \frac{\exp(V_i)}{\sum_k \exp(V_k)} \quad (3)$$

Equation 3's framework is known as the multinomial logit model, as well. In the context of this task, we use the MNL model to investigate vehicle ownership level decisions (4 levels) within household in the 2017 NHTS Georgia add-on dataset.

### ***Latent-class Discrete Choice Models***

Latent-class discrete choice models are an extension of the traditional MNL models and allow for a greater flexibility in dealing with heterogeneity in the data. This class of models, as opposed to the MNL, simultaneously identify latent homogenous subsegments (classes) of the sample and estimate a separate MNL for each latent class. Equation 4 shows the mathematical formulation of a latent-class MNL model:

$$p(y|x, z) = \sum_{c=1}^K p(c|z) p(y|c, x) \quad (4)$$

In equation 4,  $p(c|z)$  denotes the membership submodel of the latent-class MNL, with  $c$ , the latent classes, modeled directly as a function of  $z$ , the covariates, or the membership model's variables. The term  $p(y|c, x)$ , on the other hand, expresses the outcome submodel of the latent-class MNL, where the dependent variable  $y$  (VO in our context) is modeled as a function of  $x$ , a set of explanatory variables, given the latent classes.

## **Data-driven Modeling Approach**

Data-driven or machine learning models are a newer class of models developed largely by computer scientists and statisticians with a focus on handling large data and improving model prediction accuracy. This class of models has become popular in other fields, as well, with several studies in the transportation literature applying them and comparing their results with traditional behavioral approaches. In this task, subsequently, we aimed to test the model performance of data-driven models in estimating vehicle ownership and evaluate how they compare with the behavioral models used in this task. To our knowledge, this is the first time this comparison has been used in a study of vehicle ownership, given that previous studies tend to focus more on mode choice (Zhang and Xie 2008; Ermagun, Rashidi, and Lari 2015). Below, we briefly review the machine learning algorithms selected for this study.

### ***Random Forest***

Random forest, introduced by Breiman (2001), is a supervised machine learning algorithm popular for its ability to handle both regression and classification problems, low number of tuning parameters, and training and prediction speed (Breiman 2001). This class of algorithms is an extension and improvement on the decision tree algorithm where overfitting and correlated independent variables (features) could pose a problem. Denoting the vector of explanatory variables as  $X$  and the dependent variable as  $Y$ , the goal of the algorithm is to find a prediction function  $f(X)$  that minimizes the expected value of a defined loss function  $L(Y, f(X))$ . In the case of a regression application, the squared error loss is usually chosen as the loss function, while in a classification application, a zero-one function is chosen, and the minimization of such loss function results in the estimated model parameters (Cutler, Cutler, and Stevens 2011). The prediction function  $f(x)$  is constructed as a collection of decision trees whose combined output

(averaged in the case of regression, and most frequently predicted class in the case of classification) forms the final output of  $f(x)$ . Further information on splitting criterion, stopping criterion, and other details may be found in Cutler et al. (2011).

Although machine learning models are less conducive to interpretation and inference, we can use a number of methods to investigate the marginal impact and importance of the explanatory variables in the modeling process. Feature importance, as one of these techniques, can assign a relative importance to each variable, quantifying its relative contribution in the prediction of the dependent variable. At each node of a tree, an impurity index, quantifying the homogeneity with respect to the levels of the dependent variable, is computed. The feature importance associated with a specific variable, then, is calculated as the decrease in impurity of a node weighted by share of cases in that node as a result of partitioning on that variable. A higher feature importance, naturally, is associated with a higher association between that variable and the output.

## **RESULTS**

### **Disaggregate Models**

In this section, we present estimation results from the behavioral-class approach, i.e., traditional MNL and latent class MNL, and subsequently data-driven approach, i.e., random forest. We split the dataset into two sets of training and test sets, each comprising 80 percent and 20 percent of the total dataset, respectively, and compare the prediction accuracy of the models on the test set.

### *Traditional MNL*

We first estimate and present the results of the traditional MNL as shown in table 2. The set of explanatory variables used in this model include sociodemographics and built environment, and this model shows an overall reasonable goodness of fit ( $\rho_{EL}^2 = 0.406$ ).

With respect to the results, we find most of the model's coefficients agreeing with our expectation. With respect to race, we see that households identifying as White are more likely to own more vehicles compared to those identifying with other races. A low-income household, furthermore, is more likely to own fewer vehicles compared to a higher income household, a result that agrees with expectations and previous findings. Moreover, a higher number of drivers in a household increases the probability of a household having more cars.

With respect to travel behavior, we see that a higher frequency of using taxi/ridehailing services is associated with owning fewer vehicles in the household. While drawing causality conclusions between VO and ridehailing usage is not possible using this analysis, we nevertheless see a negative association between the two variables in our model.

Finally, we see a clear impact of built environment on household VO. Those households living in rural areas are more likely than their urban/suburban counterparts to own more vehicles. A higher housing density, furthermore, is associated with a lower number of vehicles per household.

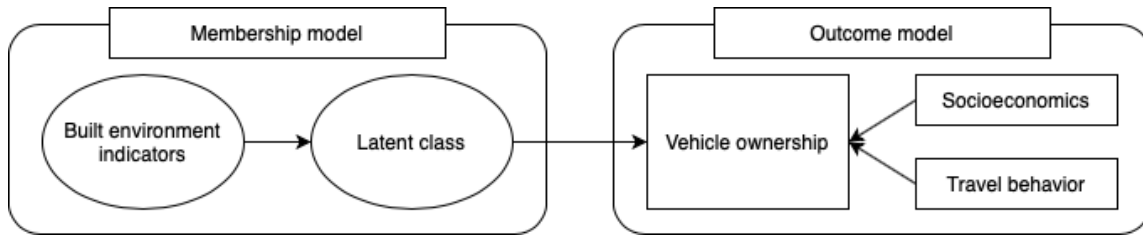
**Table 2. Multinomial logit regression (N=6205).**

Variable	0 Vehicle		1 Vehicle		2 Vehicles		3+ Vehicles	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
Constant	6.397	<0.01	6.192	<0.01	3.086	<0.01	Ref.	-
White Race	-1.194	<0.01	-0.115	0.300	-0.076	0.390	Ref.	-
Low HH Income (<\$35K)	3.298	<0.01	1.700	<0.01	0.553	<0.01	Ref.	-
No. of Drivers in the HH	-7.332	<0.01	-3.822	<0.01	-1.324	<0.01	Ref.	-
Frequency of Using Taxi/Ridehailing	0.198	<0.01	0.047	0.150	0.034	0.210	Ref.	-
Rural Dweller	-1.625	<0.01	-1.050	<0.01	-0.525	<0.01	Ref.	-
Housing Density	0.0004	<0.01	0.0003	<0.01	0.0002	<0.01	Ref.	-
LL <sub>EL</sub> =-8601.96	$\rho_{EL,adjusted}^2=0.406$							
LL <sub>c</sub> =-7636.24	$\rho_{C,adjusted}^2 = 0.331$							
LL <sub>model</sub> =-5096.62								

***Latent-class MNL***

As mentioned, the latent class approach aims to capture heterogeneity in the data by dividing the sample into several probabilistic clusters based on a certain set of parameters. We hypothesize that those living in different built environments tend to make decisions differently regarding their VO, and estimating one set of coefficients for everyone in the sample could not be an appropriate approach. We, therefore, used the built environment variables as the model covariates, and estimated latent class models with different numbers of clusters. A schematic of the overall latent class model is presented in figure 6.





**Figure 6. Schematic. Latent class of this study.**

Based on the AIC<sup>3</sup> statistic and model interpretability, we picked the latent class model with three clusters. Table 3 shows a summary of the membership model. Cluster 1, consisting of 8.6 percent of the total sample, is almost devoid of rural dwellers, and has the highest average housing density of all the clusters. The households in this cluster, with an average of 1.38 vehicles per household, also own the fewest number of vehicles on average compared to the other clusters. Cluster 2, on the other hand, has the highest share of rural dwellers of all the clusters, and accordingly has the lowest average housing density of all the clusters, too. The average number of vehicles owned in this cluster, as expected, is the highest of all clusters at 2.52 per household. Finally, cluster 3, being the largest cluster, shows to have characteristics in between those of its counterparts, with its share of rural dwellers and housing density in between clusters 1 and 2. The average number of vehicles per household, similarly, is between that of the other two clusters.

---

<sup>3</sup> Akaike information criterion.

**Table 3. Membership model parameters and descriptive statistics.**

Model Variables	Descriptive Statistics per Class			Membership Model Parameters		
	Variable Means/Share per Cluster			Class 1	Class 2	Class 3
	Class 1 (8.6%)	Class 2 (35%)	Class 3 (56.4%)	Coef.	Coef.	Coef.
Constants	-	-	-	-1.796	-0.335	0
<i>Covariates</i>						
Rural Dweller	0.004	0.42	0.16	-3.810	0.866	0
Housing Density (person/mi <sup>2</sup> )	1573.10	496.66	1132.63	0.0001	-0.0005	0
<i>Outcome Variable</i>						
HH Vehicle Count	1.38	2.52	1.72	-	-	-

Having discussed the membership model results, we now turn to the outcome model. Table 4 presents the estimation results from the latent class MNL model. The results show a more significant impact of race on clusters 2 and 3, which have a larger share of rural dwellers. Those households identifying as White are more likely to own a higher number of vehicles, agreeing with the overall results of the MNL model. With respect to income, we see an overall similar trend compared to the MNL model, with those households with lower income more likely to own fewer vehicles. The latent class model, however, shows that in cluster 2, with a large share of rural dwellers, low-income households are more likely to own 1 vehicle, as opposed to the other two classes where low-income households are more likely to own 0 vehicle. The number of drivers in the household, in addition, shows a strong association with the household number of vehicles in clusters 2 and 3, indicating that a higher number of drivers in the household is associated with a higher number of vehicles. This effect, however, shows to be statistically weak in cluster 1, although the coefficients' signs and magnitudes do point to the same conclusion. Finally, the impact of taxi/ridehailing usage frequency again points to an overall similar

conclusion compared to the MNL model, albeit with more nuance. In cluster 1, with few rural dwellers, we see that higher frequency of using such services is associated less with households with 1 or 2 vehicles, but more with 0 and 3+ vehicles. This contradictory result can point to the different demographics in nonrural areas who use these services. Studies on ridehailing services point to the younger generation and the higher income as more frequent users of ridehailing services. In clusters 2 and 3, we see that those who use taxi/ridehailing more often tend to own fewer vehicles, although in cluster 2, with more rural dwellers, a higher usage is associated more with 1-vehicle households, while in cluster 3, this variable is associated more with 0-vehicle households.

From the perspective of model performance, we see that the latent class model marginally outperforms the traditional MNL model by a few percentage points after controlling for additional model parameters.

**Table 4. Latent class multinomial regression (N=6205).**

Variables	Cluster	0 Vehicle	1 Vehicle	2 Vehicles	3+ Vehicles
		Coef.	Coef.	Coef.	Coef.
Constant	Cluster 1	-0.292	2.567	2.382	Ref.
	Cluster 2	8.158**	2.756***	2.263***	Ref.
	Cluster 3	19.787***	17.693***	8.734***	Ref.
White Race	Cluster 1	-0.110	0.776	0.869	Ref.
	Cluster 2	-5.022*	-0.304	-0.271	Ref.
	Cluster 3	-1.678*	-0.158	-0.084	Ref.
Low HH Income	Cluster 1	4.095**	1.188	0.085	Ref.
	Cluster 2	1.754	2.143***	0.704***	Ref.
	Cluster 3	5.880***	3.957***	1.364*	Ref.
No. of HH Drivers	Cluster 1	0.347	0.887	1.164	Ref.
	Cluster 2	-14.460**	-2.971***	-1.419***	Ref.
	Cluster 3	-15.564***	-9.706***	-3.604***	Ref.
Frequency of Using Taxi/Ridehailing	Cluster 1	-0.118	-1.020***	-1.055***	Ref.
	Cluster 2	0.278	0.526**	0.511**	Ref.
	Cluster 3	0.319***	0.125	0.0008	Ref.
LL <sub>EL</sub> =-8601.96		$\rho_{EL,adjusted}^2=0.422$			
LL <sub>c</sub> =-7636.24		$\rho_{C,adjusted}^2 = 0.349$			
LL <sub>model</sub> =-4922.73					

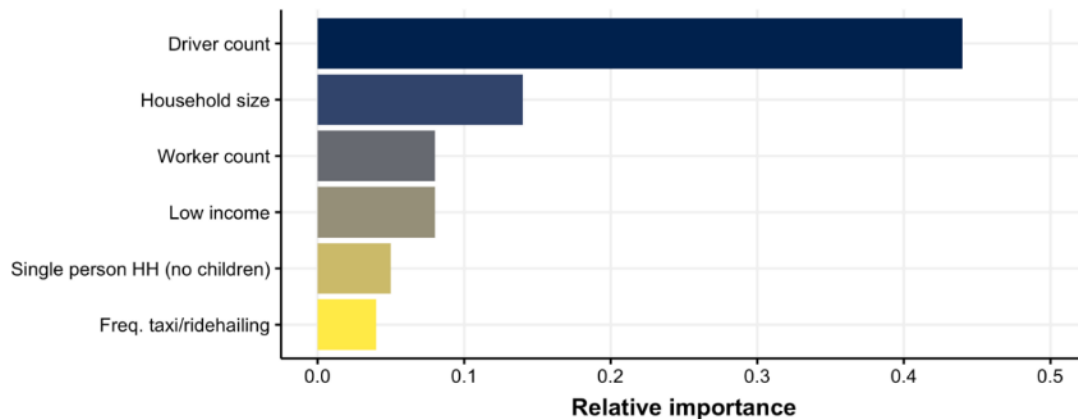
\*\*\*: p-value  $\leq 0.01$ , \*\*: p-value  $\leq 0.05$ , \*: p-value  $\leq 0.10$

### ***Random Forest***

In estimating the random forest model, we used a grid search with k-fold cross validation to find the optimum values of the model's hyperparameters, including the number of trees, maximum depth of a tree, minimum number of samples required to split a node, and minimum sample for a leaf node. We present the results of the finetuned model here. We further tested the model on the test data, as well.<sup>4</sup>

<sup>4</sup> Specifically, we apply these following parameters: RFClassifier (number of trees=300, max tree depth=20, min sample for splitting nodes=30, min sample for a leaf node=3).

Here, we first investigate the feature importance of the variables in the model. Figure 7 presents all the variables with a relative feature importance greater than 0.01. The relative importance of each variable shows to meet the expectation, with the number of drivers in the household showing to be the variable that has the largest importance in predicting VO. This impact also agreed with the model estimation from the traditional MNL model, where the inclusion of the number of HH drivers significantly improved the model fit. The influence of household size and number of HH workers, additionally, come second and third. It should be noted that due to the high correlation between these variables and number of drivers in the household (greater than 0.50), we did not include the three variables together in the model so as not to cause multicollinearity issues. Random forest, however, can handle highly correlated variables without issue.



**Figure 7. Bar graph. Feature importance plot of the random forest model.**

### *Comparison of Prediction Accuracy Between Models*

Having estimated the three models (i.e., traditional MNL, latent class MNL, and random forest), the subsequent step is to compare the prediction accuracy results between models. As mentioned,

we used the train and test method by 20 percent of the dataset to the test set, and ran the estimated models on the test set and compared their prediction accuracy.

The prediction accuracy results as shown in table 5 suggest all three models score within the same range, with random forest scoring the highest, at 69 percent. The latent class MNL, at 68.3 percent, has a prediction accuracy very close to that of the random forest, while the traditional MNL scores the lowest, at 66.9 percent. The finding that the latent class MNL provides a greater prediction accuracy than traditional MNL might not be surprising, given previous studies have found that to be the case when comparing more advanced MNLs against the traditional one to model a variety of outcomes (11, 25, 26).

Although random forest and latent class MNL perform closely on the prediction accuracy, the former requires fewer assumptions regarding the data, and can run faster. On the other hand, latent class MNL provides better insights into the heterogeneity in the data, and can be of more help in policymaking for different regions.

**Table 5. Prediction accuracy across models.**

<b>Model Type</b>	<b>Prediction Accuracy</b>
Traditional MNL	66.9%
Latent Class MNL	68.3%
Random Forest	69.0%

## **Aggregate Model**

### ***Linear Regression Model***

Although the previously discussed disaggregate models provide detailed insights into how households make their VO decisions, transferring their results to the aggregate level is theoretically not possible or requires approximations. A linear regression model, as opposed to the other nonlinear disaggregate model, allows the average disaggregate variables to be used in order to obtain the average dependent variable. In our context, therefore, we can use the average characteristics of a TAZ obtained from the 2018 American Community Survey and use the estimated linear regression model to obtain the average VO per TAZ. We, therefore, first estimate a linear regression model using the set of variables available both to NHTS and ACS, and subsequently input the average values of each GSTDM TAZ into the model and get the average VO per each TAZ.

Table 6 presents the results from linear regression predicting household VO. All the model coefficients behave as expected, with lower income households tending to own fewer vehicles and larger households tending to own more vehicles. Moreover, a higher number of workers in a household is associated positively with a higher number of VO per household.

With respect to the impact of the built environment, we see that, as expected, a higher housing density and better transit accessibility is associated with a lower number of vehicles owned per household.

The model's overall performance, with an adjusted  $R^2$  of 0.363, shows a reasonable fit.

**Table 6. Linear regression predicting household vehicle count (N=8307).**

<b>Explanatory Variable</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>t-value</b>	<b>p-value</b>
(Intercept)	1.510	0.0293	51.623	<0.001
Low HH Income (<\$35k)	-0.620	0.0249	-24.871	<0.001
High HH Income (>\$100k)	0.315	0.0285	11.045	<0.001
HH Size	0.137	0.0089	15.358	<0.001
No. of Workers per HH	0.356	0.0145	24.599	<0.001
Housing Density	-0.000591	0.0000091	-6.519	<0.001
AllTransit Perf. Score	-0.0662	0.00498	-13.303	<0.001

Adjusted R-squared: 0.363

Using the regression model as specified in table 6, we then compute the average vehicle ownership per each GSTDM TAZ. The computed aggregate values are included in an accompanying text file, and are ready to be used in other steps of the GSTDM. We further explore the distribution of the predicted VO at the TAZ level. Table 7 presents a descriptive statistic of the distribution of vehicle ownership in the state of Georgia as predicted by our aggregate model.

**Table 7. A comparison of predicted vehicle counts and observations from the ACS.**

<b>Source</b>	<b>Min.</b>	<b>1<sup>st</sup> Qu.</b>	<b>Median</b>	<b>Mean</b>	<b>3<sup>rd</sup> Qu.</b>	<b>Max.</b>
Predicted	0.528	1.857	2.179	2.230	2.521	5.220



## **CHAPTER 4. INTRODUCING TIME OF DAY IN THE GSTDM**

Specifying the time of day (TOD) at which a trip occurs in the travel demand model allows for more detailed analysis and can thus lead to more effective demand management strategies. Rather than grouping all trips within a 24-hour period, models that incorporate TOD distinguish between trips that occur during peak and off-peak periods, better reflecting congestion effects that might prompt travelers to shift their travel mode or route (Transportation Research Board & National Academies of Sciences, Engineering, and Medicine, 2017). To achieve this level of detail, the research team conducted an extensive literature review and developed a methodology to implement TOD using the 2017 NHTS dataset and the latest external data sources. The proposed method is designed to be compatible with the latest version of the Georgia Statewide Travel Demand Model for practical use.

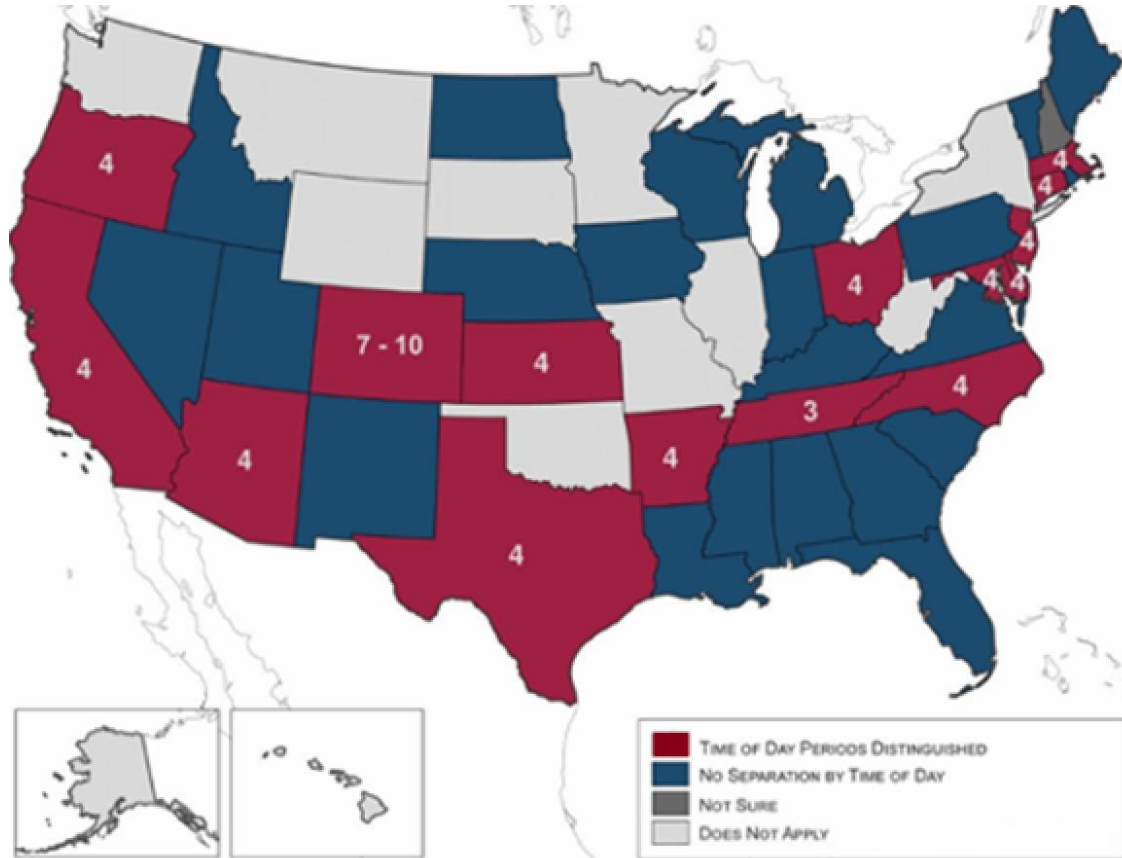
### **LITERATURE REVIEW**

In order to understand how TOD can best be incorporated into the GSTDM, the research team reviewed the methods used in other statewide travel demand models first, then summarized and classified these methods. The sections below provide examples of TOD implementation in other statewide models, followed by a description of the different possible TOD methods and the stages at which they are implemented in a four-step model. Table 8, at the end of this literature review section, summarizes all the TOD examples and methods described here.

#### **TOD in Statewide Travel Demand Models**

Currently, 15 of the 34 states with operational statewide travel demand models account for TOD (Moeckel et al. 2019). The map in figure 8 shows these states in red, with the number of time

periods specified. The following sections describe in more detail the TOD methods used in a few of these models.



**Figure 8. Map. States with TDMs that account for time of day (Moeckel et al. 2019).**

### *Virginia Statewide Travel Demand Model*

The travel demand model for the state of Virginia, developed in 2013, uses NHTS data and traffic counts to calculate time-of-day factors, ultimately dividing trips into four time periods: AM peak (6 AM–9 AM), Midday off-peak (9 AM–3 PM), PM peak (3 PM–6 PM), and Night off-peak (6 PM–6 AM) (Ma and Demetsky 2013). The process for implementing time of day occurs in two steps. In the first step, trips are divided into either peak or off-peak periods after trip generation. The second step, which occurs after mode choice, further divides these trips into

the four time periods listed above. After traffic assignment, a feedback process is used to update the initial time-of-day factors. This process results in a total of 16 factors, one for each combination of trip purpose—home-based work (HBW), home-based other (HBO), non-home-based (NHB), or external—with time period.

### ***North Carolina Statewide Travel Demand Model***

North Carolina’s statewide travel demand model implements TOD following the trip distribution stage. The model uses NHTS data and traffic counts to calculate time-of-day factors for short-distance and long-distance trips, respectively, separating trips into four time periods: AM peak (6 AM–9 AM), Midday off-peak (9 AM–4 PM), PM peak (4 PM–7 PM), and Night off-peak (7 PM–6 AM). In calculating the time-of-day factors, this model is somewhat unique in its use of “trips in motion” (WSP/Parsons Brinckerhoff 2015). This technique divides the day into 15-minute intervals to count the number of trips in progress during each period. In this way, rather than only being counted in the period in which they start or end, trips can be counted in multiple periods, depending on their length.

### ***Other States***

There are several cases that have not adopted TOD implementation in the four-step statewide travel demand modeling approach. For example, Indiana removed the TOD procedures from the statewide model due to file size, model running time, and lack of observation data such as traffic counts. In the case of Florida, some MPO models are taking the TOD implementation into account, but the Florida statewide model is on a daily trip basis.

## **Classification of TOD Methods**

The sections below summarize the possible methods for incorporating TOD into four-step travel demand models. The descriptions of the methods are based mostly on a study completed for the Florida Department of Transportation (Pendyala, 2002).

### ***Method 1: Implementing TOD After the Trip-generation Stage***

In this method, TOD is implemented after the trip-generation step and before trip distribution. Trip generation is, therefore, conducted as in traditional daily models before trips are separated by trip purpose to determine which trips occur during peak periods. The last three steps, trip distribution, mode choice, and assignment, are then performed separately for each period (Pendyala 2002). Implementing time of day at this stage allows for more detailed analysis in these final three steps, as the trips within each time period are more homogenous than when all trips are considered at once.

### ***Method 2: Implementing TOD After the Trip-distribution Stage***

This method incorporates time of day between step two, trip distribution, and step three, mode choice, of the four-step model (Pendyala 2002). As in Method 1 described above, trip generation is conducted for the whole day. However, unlike in Method 1, trip distribution is then determined before dividing trips by purpose and time period. Mode choice and trip assignment are then performed for each period. This method also allows for more detailed mode choice analysis but introduces inconsistency between trip distribution and mode choice, as distribution is based on daily travel speeds while mode choice is based on period-specific travel speeds.

### ***Method 3: Implementing TOD After the Mode-choice Stage***

In this method, the first three steps of the four-step model are performed on the day as a whole. After mode choice is determined, trips are separated by trip purpose and mode to define peak periods before performing trip assignment (Pendyala 2002). Setting the time periods at this stage allows for the consideration of different peak periods for different modes. However, trip distribution and mode choice analysis are less detailed using this method than in Method 1 or 2, as they do not account for time of day.

### ***Method 4: Implementing TOD After the Trip-assignment Stage***

Finally, this method accounts for time of day at the end of the four-step model. In this case, all four steps are performed for the entire day, and the outputs of the trip-assignment stage are analyzed to determine peak periods (Pendyala 2002). As a result, trip assignment does not reflect the variations in speed and volume that occur throughout a day. However, this method is the easiest to implement. This is the method currently employed by the GSTDM.

**Table 8. Summary of TOD implementation in statewide travel demand models.**

<b>State</b>	<b>Model Type</b>	<b>Base Year</b>	<b>Data</b>	<b>TOD Classification (Time Periods)</b>	<b>TOD Implementation Stage</b>
Georgia	Four-step model	2015	NHTS 2009 traffic counts	4 time periods: <ul style="list-style-type: none"> <li>• AM peak (6–10)</li> <li>• Midday (10–15)</li> <li>• PM peak (15–19)</li> <li>• Night (19–6)</li> </ul>	Postprocessing after trip assignment
Colorado	Activity-based model	2010	2010 Front Range Travel Counts	Tour TOD based on 1-hour periods	N/A
North Carolina	Four-step model	2011	NHTS 2009	4 time periods: <ul style="list-style-type: none"> <li>• AM peak (6 AM–9 AM)</li> <li>• Midday off-peak (9 AM–4 PM)</li> <li>• PM peak (4 PM–7 PM)</li> <li>• Night off-peak (7 PM–6 AM)</li> </ul>	After trip distribution (Method 2)
Virginia	Four-step model		NHTS 2009	4 time periods: <ul style="list-style-type: none"> <li>• AM peak (6 AM–9 AM)</li> <li>• Midday off-peak (9 AM–3 PM)</li> <li>• PM peak (3 PM–6 PM)</li> <li>• Night off-peak (6 PM–6 AM)</li> </ul>	Two-stage process: <ul style="list-style-type: none"> <li>• Peak/off-peak split after trip generation</li> <li>• Further divided into 4 time periods after mode choice</li> </ul>

## **METHODOLOGY**

### **Overview**

Based on the literature review and the peer review (Federal Highway Administration 2013), the research team specified the methodology of the TOD implementation, as discussed below.

First, as previously mentioned, the research team followed the ground rule that the proposed travel demand model and its specific methodologies should be embedded in the current version of the GSTDM with compatible modeling methodologies. Given that the trip purpose in the GSTDM, which is classified into two different categories based on trip distance, TOD factors in this research are also separately determined based on trip distance: (1) TOD factors for short-distance trips, and (2) TOD factors for long-distance trips. The same threshold of 50 miles between short- and long-distance trips that was introduced in the Georgia statewide model is applied to this research.

Second, the concept of “trips-in-motion” is adopted to specify time-of-day periods. This method counts the total number of trips for each time bin (e.g., 15 minutes). Importantly, it allows individual trips to be counted multiple times when it occupies multiple time bins, while the traditional counting method considers either start or end time of each trip (that is, each trip must belong to only one time bin based on trip start or end time).

Third, a specific weight indicator proposed by the 2017 NHTS is utilized. The 2017 NHTS includes four weights: household weights, person-level weights, travel-day-level weights, and vehicle weights. The research team only used person-level weights, which are designed to represent all persons in the study area.

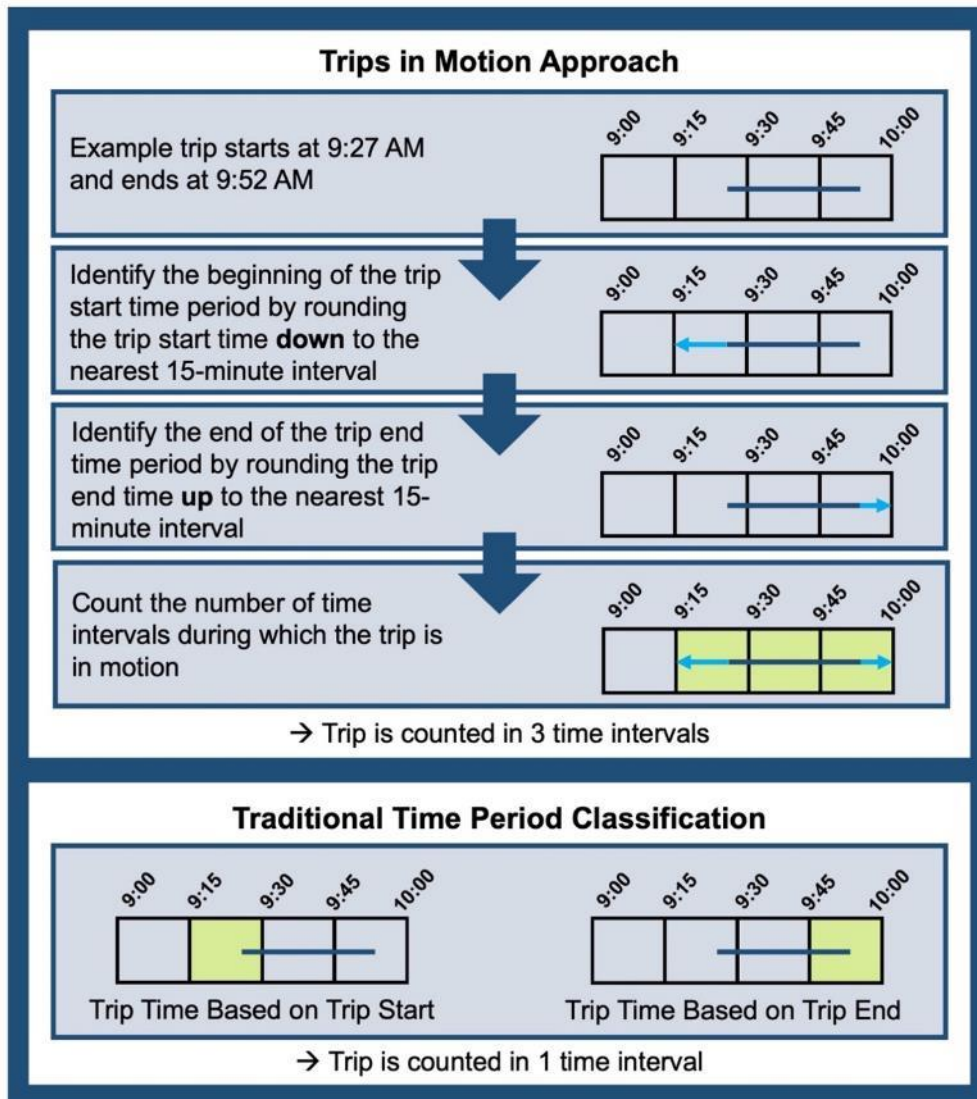
Fourth, the TOD implementation is designed to be conducted after the trip-generation step. Since the current GSTDM method of the TOD implementation is postprocessing after trip assignment (simply dividing daily trips into four time-specific trips), TOD-specific trips cannot be accurately calculated. Instead, the proposed approach can determine trips by TOD more accurately after taking peak time traffic congestion into account.

### **Benchmark Time for TOD**

The research team reviewed which time is the best to determine the time-of-day periods (referred to as *benchmark time* in this research). Traditionally, start time or end time of each trip are considered to define the cut-off of the peak time periods. In this research, however, a newly defined indicator, trips in motion, is adopted. It was introduced in a statewide travel demand model in the North Carolina Department of Transportation (NCDOT) to better account for trips that take place across more than one time period by counting the number of trips in progress during each 15-minute interval. Counting trips in motion, rather than simply trip start or end times, ultimately produces a more accurate estimate of the proportion of trips that occur in each time period. Figure 9 provides an example of how the trips-in-motion approach would differ from a traditional approach in classifying a trip that spans multiple 15-minute time intervals. The entire process of applying the trips-in-motion approach to counting trips is written in an accompanying R script.

There is one thing to keep in mind when using this method. The trips-in-motion approach is beneficial to accounting for traffic congestion situations, but it is not appropriate for vehicle miles traveled (VMT) or traffic counts itself, as that method can overcount the true numbers (e.g., a single trip can be counted multiple times via trips-in-motion).





**Figure 9. Diagram. Comparison of trips-in-motion approach to traditional time period classification.**

### Trip Purpose

Basically, the research team classified the passenger trip purposes into three: home-based work (HBW), home-based other (HBO), and non-home-based (NHB). In the NHTS classification, home-based trip purposes include home-based work (HBW), shopping (HBSHOP), social and recreation (HBSOCREC), and other (HBO), but the latter three purposes were combined into

HBO in this study because reconciled NHTS in Georgia does not have sufficient cases of those three purposes, and there is no significant behavioral difference among them (i.e., they are mostly related to home-based leisure/social trips, being treated as one purpose). Therefore, the TOD factors are determined for three purposes.

## **RESULTS**

### **Number of TOD Periods**

Determining the number of TOD periods in the GSTDM is a starting point of task 2. The number of time periods generally ranges between two (simply peak and off-peak) and five (early morning, AM peak, midday, PM peak, and overnight). Although accuracy increases with the number of time periods used, data needs and computation times do, as well. The decision of how many time periods to include should therefore account for both the level of detail and accuracy needed in the analysis and the data available.

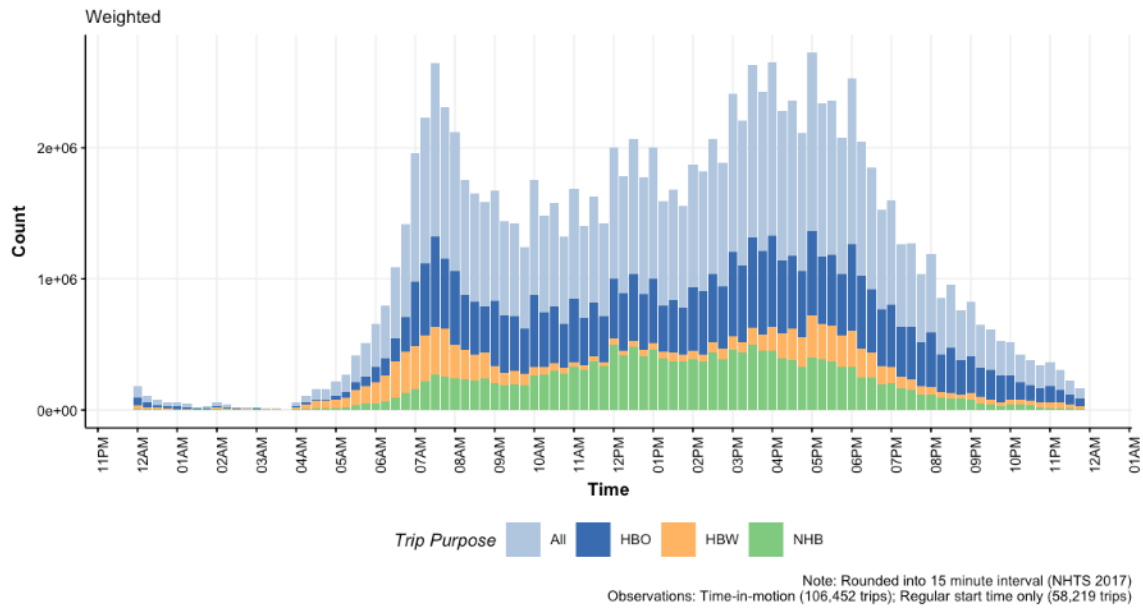
The research team goes through the temporal trip distribution by purpose based on the “trips in motion” approach. Figure 10 shows the resulting temporal trip distribution using this trips-in-motion approach. The figure appears to indicate two dominant time periods where trips occurred, i.e., in the morning and evening. To this end, as a result, the TOD periods in the GSTDM are classified into four time periods as follows:

AM peak: 6 AM–10 AM

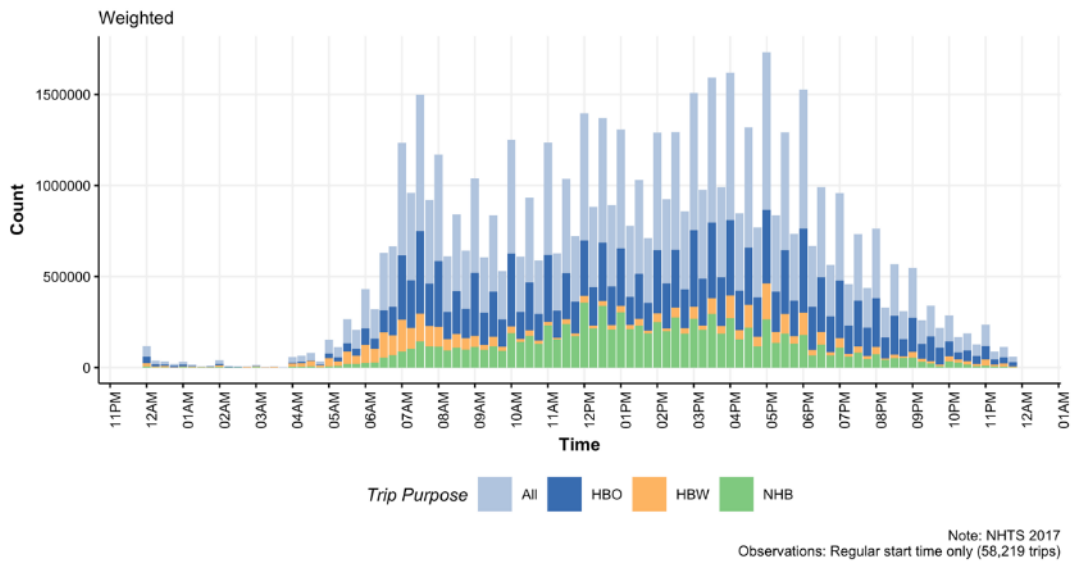
Midday: 10 AM–3 PM

PM peak: 3 PM–7 PM

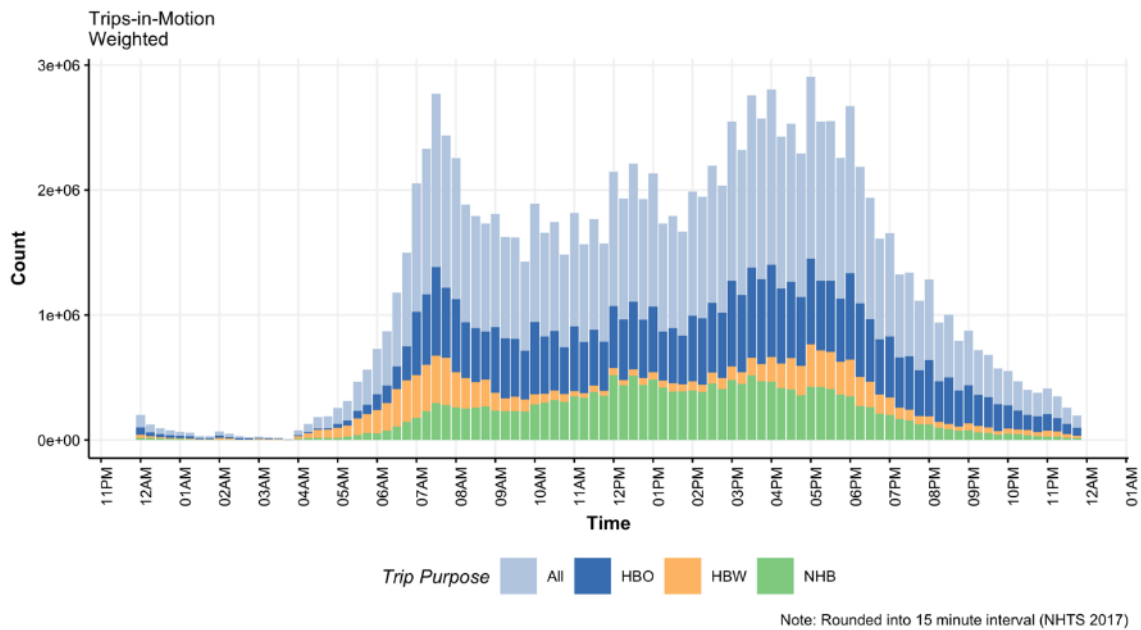
Night: 7 PM–6 AM



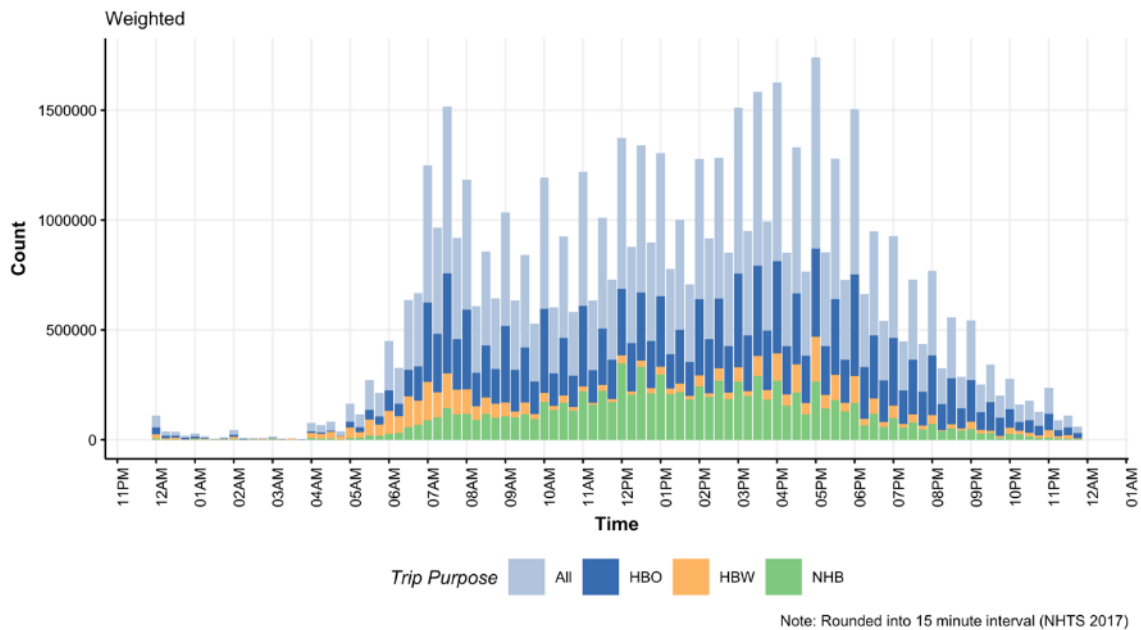
**Figure 10. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, short-distance trips.**



**Figure 11. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach, short-distance trips.**



**Figure 12. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, all trips.**

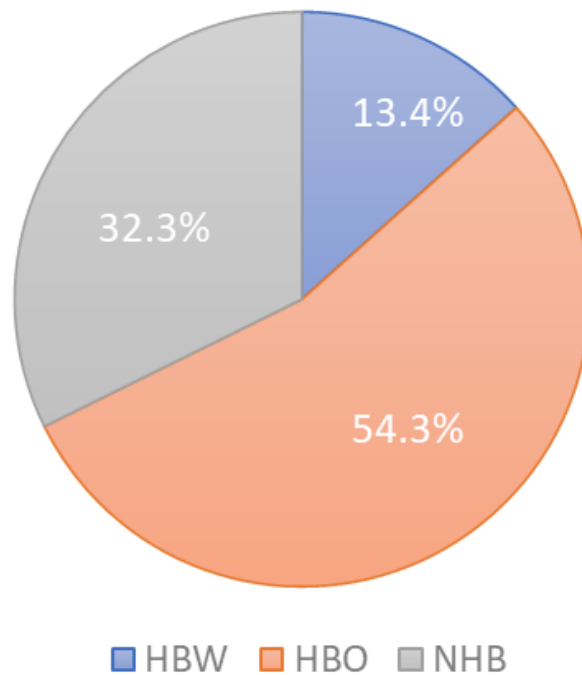


**Figure 13. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach, all trips.**

Two transportation agencies in Georgia, the Georgia Department of Transportation and the Atlanta Regional Commission (ARC), are currently using the same peak time period: 6 AM–10 AM for the AM peak period, and 3 PM–7 PM for the PM peak period. This consistency will allow both agencies to utilize the proposed TOD factors as a reference in travel demand models.

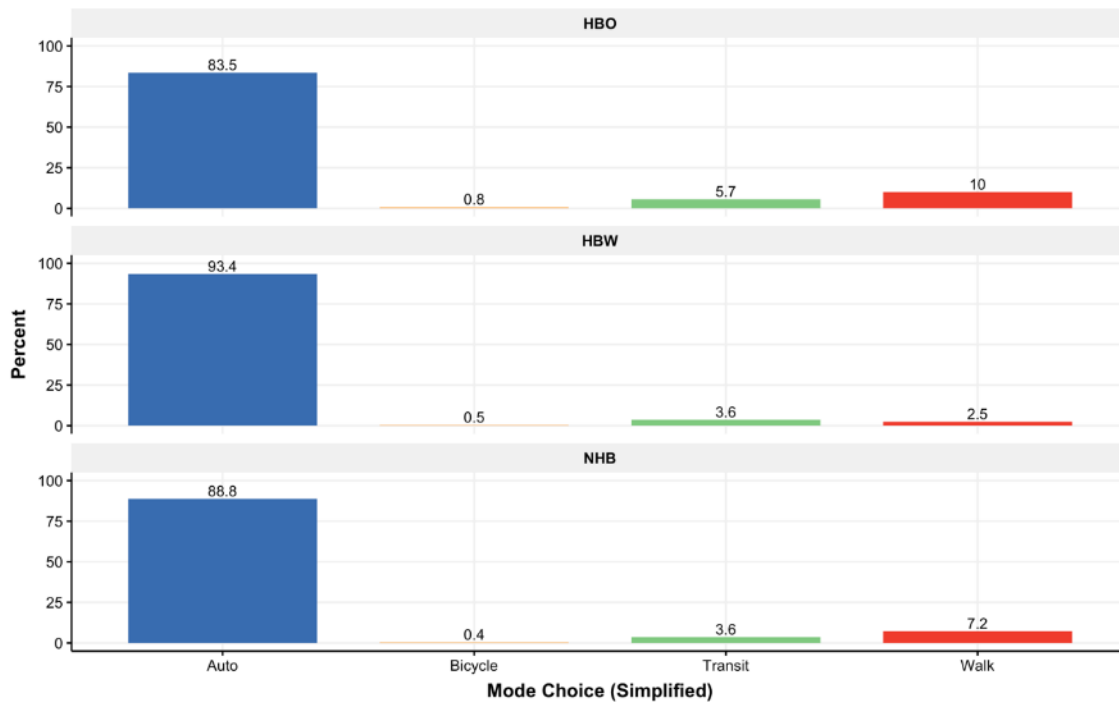
### Short-distance Trips

As mentioned previously, TOD factors for short- and long-distance trips are separately determined using the 2017 NHTS dataset. The profile of short-distance trips, without accounting for the trips-in-motion approach, is presented as follows. HBO trips account for more than half of total trips (54.3 percent, 30,902 trips), followed by NHB (32.3 percent, 19,695 trips) and HBW (13.4 percent, 7,622 trips).

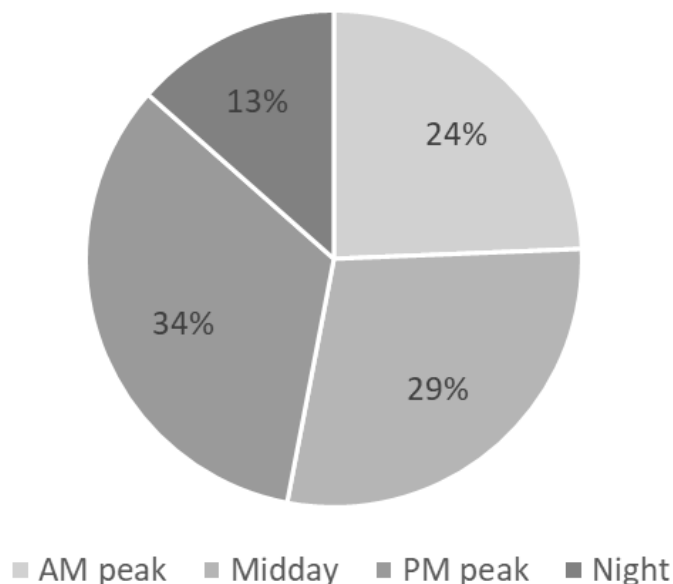


**Figure 14. Pie graph. Shares of total short-distance trips by purpose.**

Breaking down the shares of total short-distance trips by purpose and mode choice indicates a largely consistent finding where most trips were made by auto. That is, as shown in figure 15, auto is the dominant mode choice, exceeding 80 percent from the total share for each trip purpose category. This observation is particularly relevant for the HBW trips where there were 7,288 auto trips (93.4 percent) out of 7,604 HBW trips. Trips classified as HBO had the lowest share of auto trips at 83.5 percent, or 26,618 out of 30,841 HBO trips. Moreover, there were 17,834 auto trips (88.8 percent) out of 19,587 NHB trips.



**Figure 15. Bar graphs. Shares of total short-distance trips by mode and purpose.**



**Figure 16. Pie graph. Shares of total short-distance trips by time period (all purposes).**

Trip shares by time period following the trips-in-motion approach are shown in table 9. First, there are two pronounced peaks for HBW. AM and PM peaks account for 38.7 and 34.8 percent respectively (73.5 percent of daily trips), whereas off-peak trips account only for 26.5 percent, which is a plausible distribution of commuting trips. The pattern of HBO trips is considerably different from that of HBW; its PM peak and Night are fairly similar (32.7 and 16.0 percent) to those of HBW, but the portion of AM peak for HBO is markedly lower than that for HBW. Accordingly, the share of Midday for HBO is significantly higher than that for HBW. This is because characteristics of various types of home-based trip purposes other than HBW, including social, recreation, and shopping trips that mostly occur in the daytime, are all mixed up. For example, people usually leave their home late morning or around noon (i.e., after the AM peak) to meet people (social) or to visit some places (social, recreation, or shopping). With respect to NHB, the overall pattern is similar to the HBO distribution. A majority of trips are concentrated in Midday (39.9 percent), followed by PM peak (34.1 percent), while AM peak and Night

account only for 18.0 and 7.9 percent, respectively. Since most NHB trips are treated as a part of a trip chain and not a major component in a tour trip, it is likely to rely on home-based trips and supposed to replicate characteristics of home-based trips. Although the AM peak of HBW accounts for 38 percent of total HBW trips, it does not significantly affect NHB distribution because HBW constitutes a relatively small portion (17 percent) of total trips, meaning that NHB trips are heavily affected by characteristics of HBO trips.

**Table 9. Trip shares by time period and purpose (short-distance trips).**

	<b>AM Peak (%)</b>	<b>Midday (%)</b>	<b>PM Peak (%)</b>	<b>Night (%)</b>
<b>HBO</b>	40.9	52.0	58.6	25.9
<b>HBW</b>	91.2	22.9	79.4	30.4
<b>NHB</b>	29.7	75.3	58.2	11.9

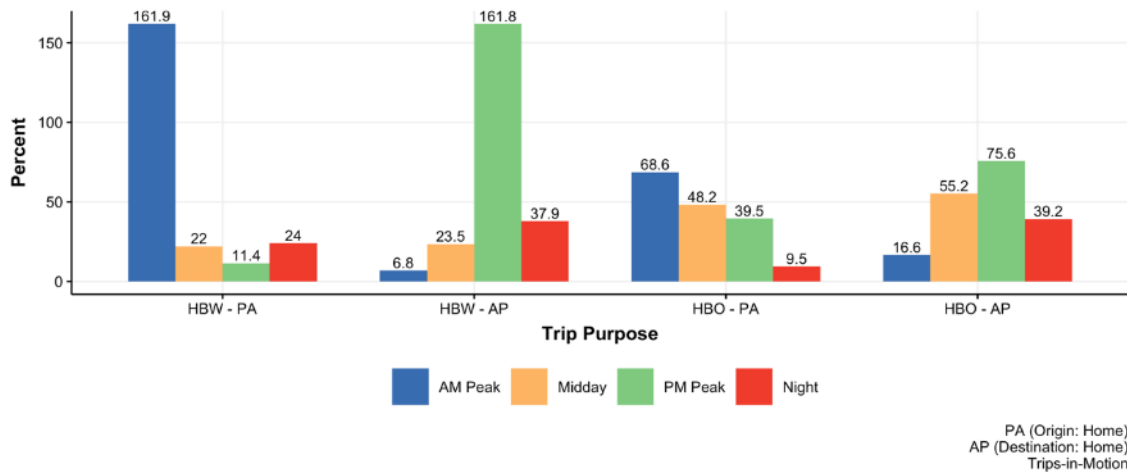
Note: The sum of shares by purpose exceeds 100 percent because the denominator is the total number of trips while the numerator is the sum of multiple-counted trips based on trips-in-motion.

### ***Spatial Distribution of Trips by Time Period***

Basically, the amounts of production and attraction trips depend highly on time of day, and their pattern is significantly opposite from each other. For example, most HBW trips are generated during the peak period, but directionality of those trips during the AM peak is mostly “from home to workplace” (referred to as PA in this report). Likewise, most trips during the PM peak are “from workplace to home (AP). It means that a different set of TOD factors needs to be applied to the total number of PA and AP trips by TAZ after the trip-generation stage to convert daily total trips to period-specific trips. In order to examine directionality, the research team further conducted a series of data visualization exercises to observe the spatial distribution of the *home-based* trips by mapping the number of trips whose trip origin is home (PA) or trip destination is home (AP) at the census tract level by trip purpose and peak time period (for



example, refer to figure 13). Figure 17 illustrates shares of short-distance home-based trips by purpose and TOD, which is further divided into PA and AP. It shows that 71 percent of HBW PA trips are generated during the AM peak, while 67 percent of HBW AP trips are concentrated in the PM peak. Regarding NHB trips, it does not have directionality between production and attraction, and there is no clear home-based source producing or attracting trips (all NHB trips just have origin and destination). Therefore, TOD factors for NHB trips are not divided into PA and AP, but simple TOD factors specified into four time periods (refer to table 4) are applied to TOD implementation. Detailed interpretation and implications of map visualization are described by purpose as follows.

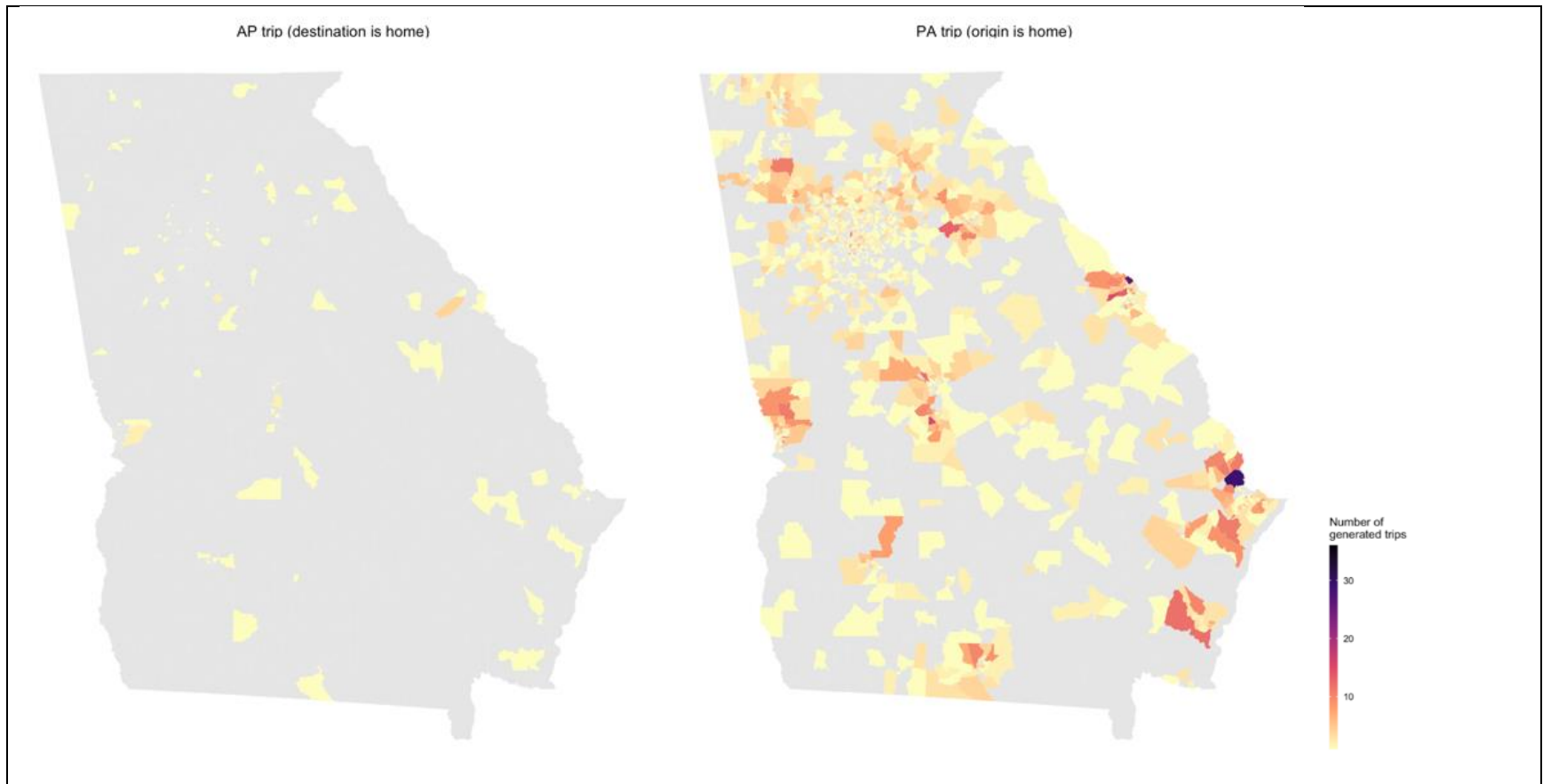


**Figure 17. Bar graph. Share of short-distance trips by trip purpose and directionality.**

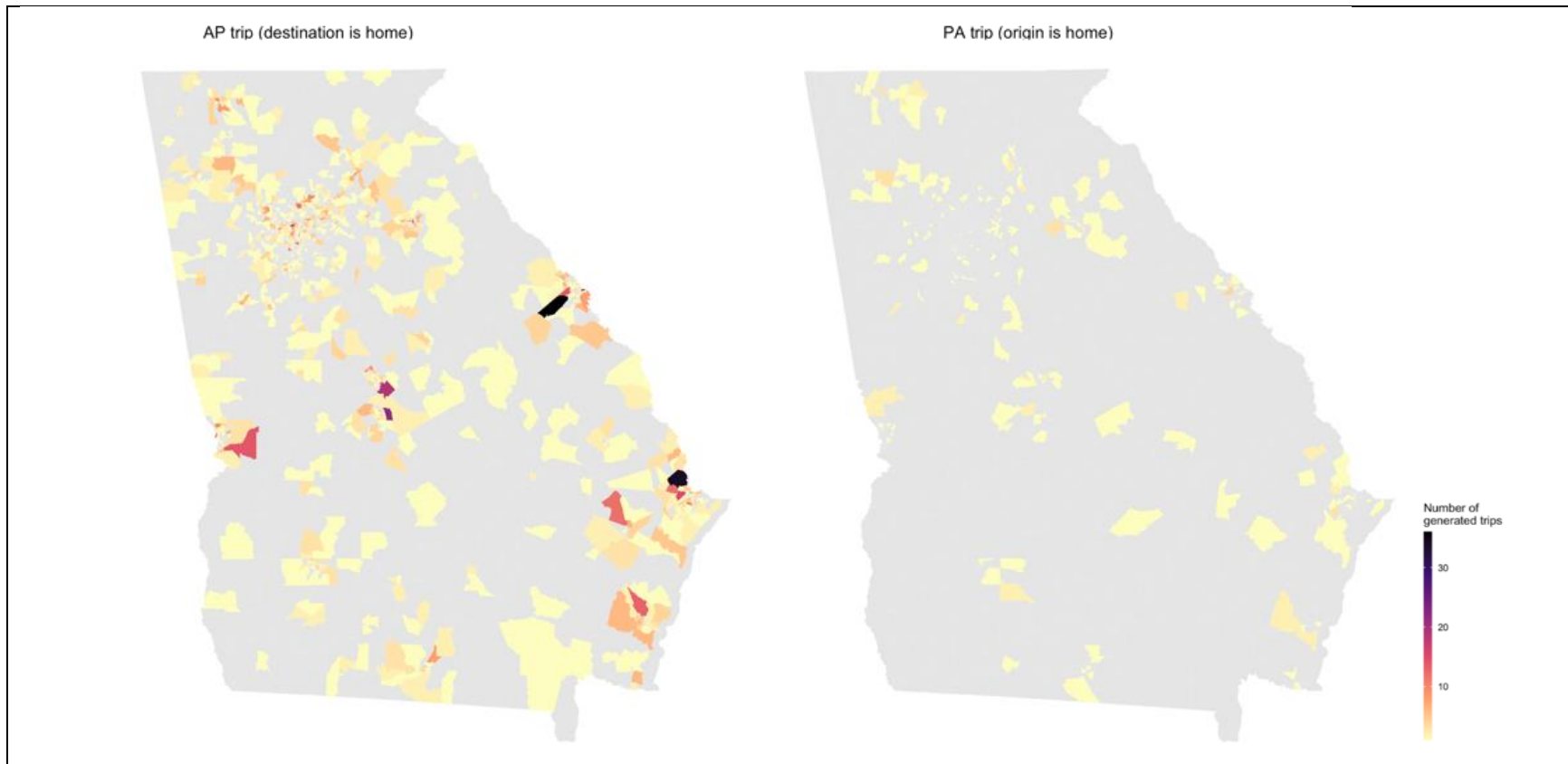
*Home-based work (HBW)*: Figure 18 and figure 19 shows the spatial distribution of the generated HBW trips at the census tract level during the AM and PM peaks, respectively; the darker the TAZs are, the more the trips are generated. The figure on the left side illustrates the number of

trips whose *origin* is home (i.e., PA trip), aggregated by origin TAZ, and the other one presents the number of trips whose destination is home (i.e., AP trip), aggregated by destination TAZ.

By and large, all four maps demonstrate typical patterns. On one hand, there are substantially more trips originated *from home to work* (PA trips) than generated trips *from work to home* (AP trips) during the AM peak period; on the other hand, there are significantly more trips originated *from work to home* during the PM peak period. It means that a larger portion of HBW trips should be allocated for PA trips at AM peak and for AP trips at PM peak.



**Figure 18. Maps. Spatial distribution of generated HBW short-distance trips at the census tract TAZ level during the AM peak period.**



**Figure 19. Maps. Spatial distribution of generated HBW short-distance trips at the census tract TAZ level during the PM peak period.**

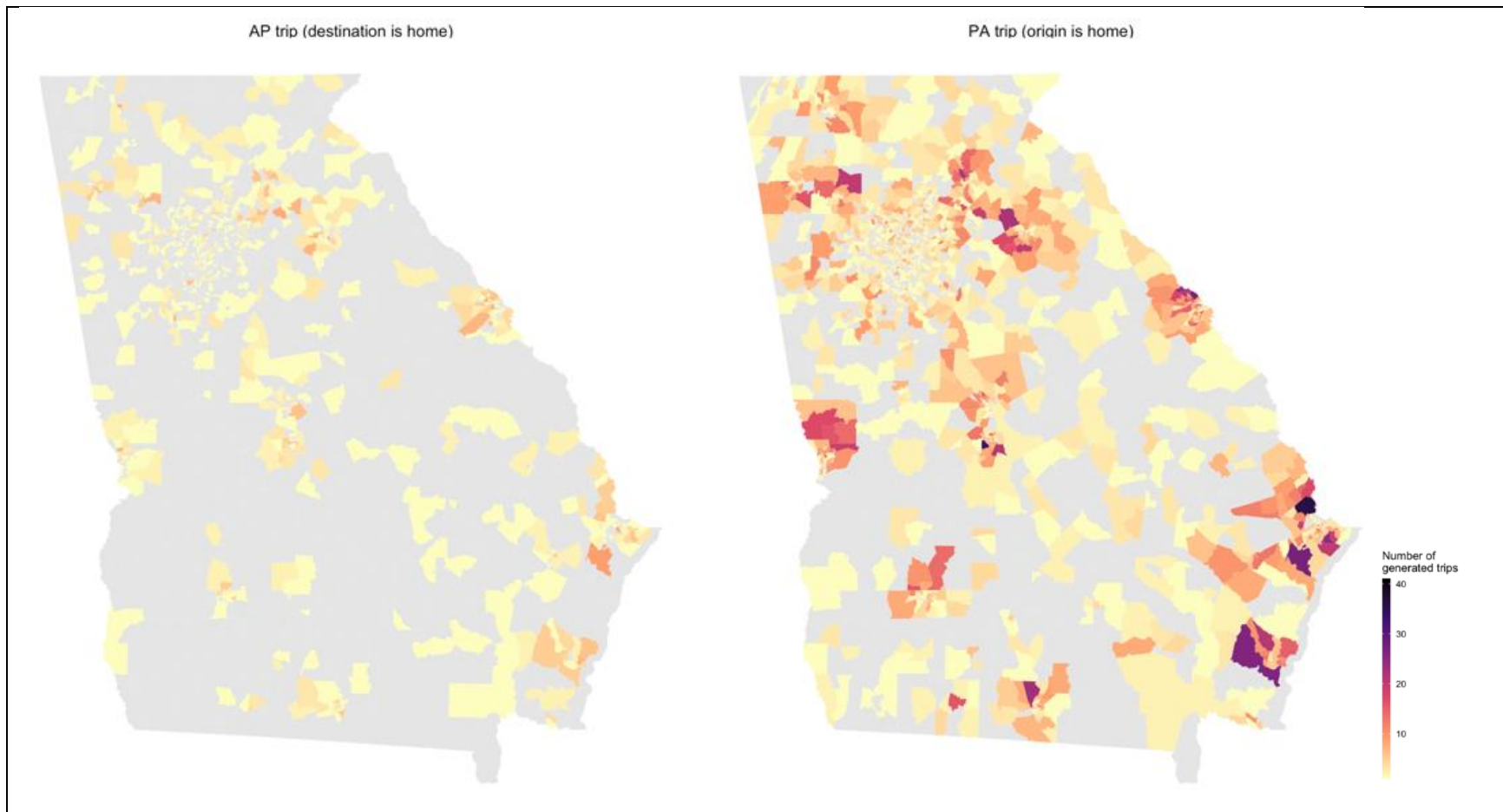
Table 10 corroborates those findings. It shows that a more substantial number of trips are originating from home to work during the AM peak (2,898 trips) than during the PM peak (239 trips) and more trips returning home during the PM peak (2,276) than the AM peak (116).

**Table 10. Number of generated HBW trips by directionality, short-distance trips.**

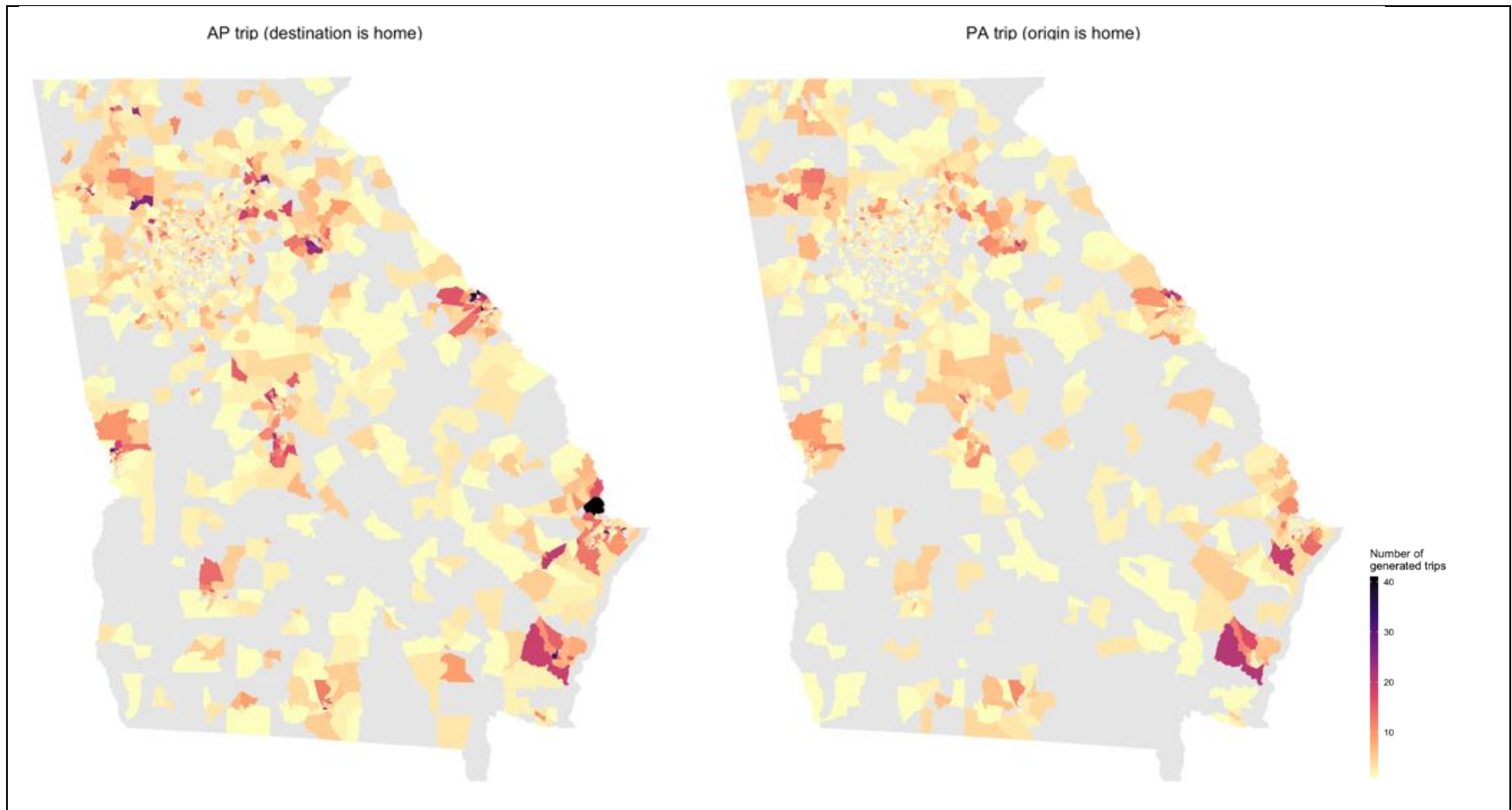
	<b>AM Peak</b>	<b>Midday</b>	<b>PM Peak</b>	<b>Night</b>
<b>HBW-PA</b>	2,898	473	239	496
<b>HBW-AP</b>	116	444	2,276	564

PA (Origin: Home); AP (Destination: Home)

*Home-based other (HBO)*: Based on the same analytical process for the HBW trips as presented above, the research team also investigates the spatial distribution of the generated HBO trips at the census tract level during the AM and PM peaks. As presented in figure 20 and figure 21, overall directionality of HBO trips between PA and AP shows extremely opposite patterns in both peak time periods. Following the expectation, most HBO trips originating from home, which include trips such as dropping children off to school, shopping, and recreation, tend to occur during the AM peak rather than PM peak. Similarly, and again following the expectation, the analysis also presents that there are substantially more HBO trips heading toward home during the PM peak in comparison to the AM peak period.



**Figure 20. Maps. Spatial distribution of generated HBO short-distance trips at the census tract TAZ level during the AM peak period.**



**Figure 21. Maps. Spatial distribution of generated HBO short-distance trips at the census tract TAZ level during the PM peak period.**

Table 11 further substantiates the characteristics of period-specific HBO trips. It shows the observed number of HBO trips originating from or heading to home during the different time period throughout the day. Considerably more trips originate from home during the AM peak (5,724 trips) than during the PM peak (3,507 trips) and more trips return home during the PM peak (6,247) than the AM peak (1,549). Interestingly, a sizable portion of HBO AP trips occurs during Midday (4,856); this makes sense given that shopping or leisure trips can occur between two peak time periods.

**Table 11. Number of generated HBO trips by directionality, short-distance trips.**

	<b>AM Peak</b>	<b>Midday</b>	<b>PM Peak</b>	<b>Night</b>
<b>HBO-PA</b>	5,724	3,910	3,507	833
<b>HBO-AP</b>	1,549	4,856	6,247	3,149

PA (Origin: Home); AP (Destination: Home)

***Non-home Based (NHB)***

As mentioned previously, unlike the HBW and HBO trips which have directionality, NHB trips traditionally do not consider directionality in determining TOD factors. Neither production nor attraction are a base spot generating trips, so it is impossible to define directionality in the context of a trip-based model. Given that reason, the research team does not take directionality into account (i.e., not dividing trips into PA and AP) in determining TOD factors for NHB trips. Thus, the simple share of NHB trips by TOD period is presented in table 12.

**Table 12. Number of generated NHB trips, short-distance trips.**

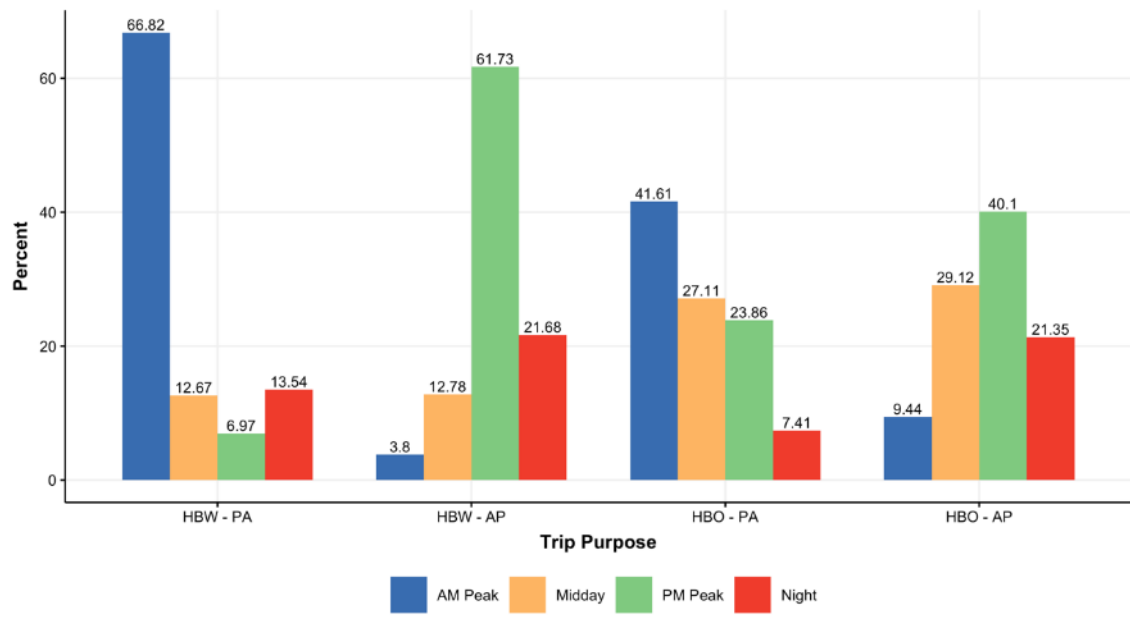
	<b>AM Peak</b>	<b>Midday</b>	<b>PM Peak</b>	<b>Night</b>
<b>NHB</b>	3,634	9,372	6,418	1,514



### ***Final TOD Factors***

Before determining final TOD factors, the research team first scales individual trips up using the person-level weights (“WTPERFIN” variable) provided by NHTS data. Based on the basic TOD factors stratified by four time periods as presented above, the specified TOD factors for production to attraction (PA) and attraction to production (AP) are calculated after considering the directionality, as shown in figure 22 and table 13.

Not surprisingly, HBW shows the opposite distribution of trips between PA and AP. The PA and AP trips for the AM peak constitute 66.8 and 3.8 percent of daily trips, respectively, while those for the PM peak account for 6.9 and 61.7 percent, respectively. On the other hand, the overall distribution of HBO is quite different. There is a clear large portion of trips at AM peak for PA (41.6 percent) and at PM peak for AP (40.1 percent), but midday trips also account for almost 30 percent of daily trips, which is consistent with the result above. Regarding NHB, final TOD factors for AP are not presented in table 13 because of a lack of an anchor station and directionality. In order to calculate period-specific traffic volumes for NHB AP trips, the TOD factors for NHB PA can be applied. That is, TOD factors for NHB AP are 17.8, 42.9, 31.2, and 8.1 for AM peak, Midday, PM peak, and Night, respectively.



PA (Origin: Home)  
 AP (Destination: Home)

**Figure 22. Bar graph. TOD factors, short-distance trips (weighted).**

**Table 13. TOD factors, short-distance trips (weighted).**

	Volume		Distribution	
	PA	AP	PA	AP
<b>HBW</b>				
<b>AM Peak</b>	1,391,472	68,058	66.82	3.80
<b>Midday</b>	263,863	228,745	12.67	12.78
<b>PM Peak</b>	145,226	1,104,743	6.97	61.73
<b>Night</b>	281,923	388,068	13.54	21.68
<b>Sum</b>	<b>2,082,484</b>	<b>1,789,614</b>	<b>100</b>	<b>100</b>
<b>HBO</b>				
<b>AM Peak</b>	3,039,749	771,153	41.61	9.44
<b>Midday</b>	1,980,590	2,378,571	27.11	29.12
<b>PM Peak</b>	1,743,121	3,275,450	23.86	40.10
<b>Night</b>	541,466	1,743,938	7.41	21.35
<b>Sum</b>	<b>7,304,926</b>	<b>8,169,112</b>	<b>100</b>	<b>100</b>
<b>NHB</b>				
<b>AM Peak</b>	1,823,084	- <sup>a</sup>	17.80	- <sup>a</sup>
<b>Midday</b>	4,395,230	-	42.92	-
<b>PM Peak</b>	3,193,822	-	31.19	-
<b>Night</b>	827,232	-	8.08	-
<b>Sum</b>	<b>10,239,368</b>	<b>-</b>	<b>100</b>	<b>-</b>

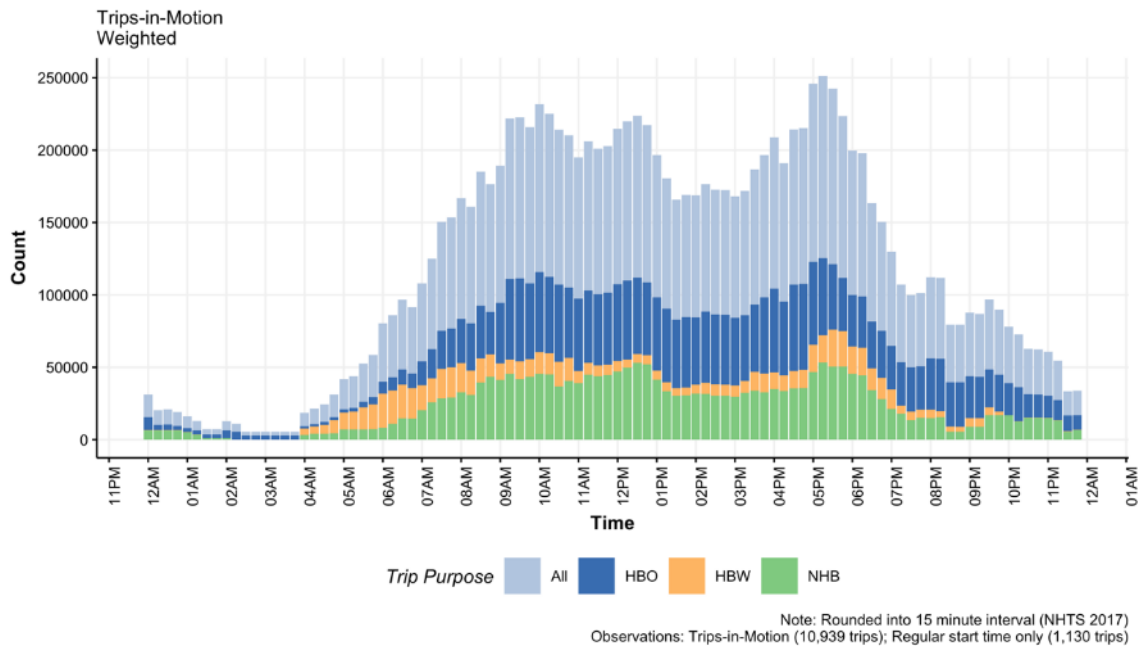
<sup>a</sup> Final traffic volumes and corresponding TOD factors for NHB AP trips are not presented due to the lack of directionality, but TOD factors for NHB PA can be applied to the same category of NHB AP.

### Long-distance Trips

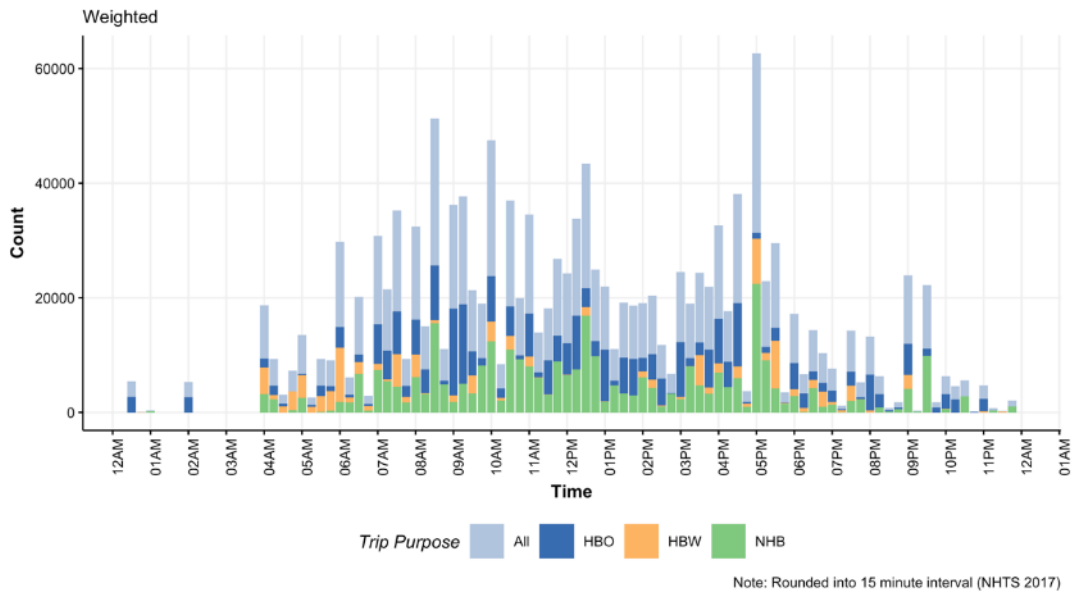
Following the process to compute TOD factors for short-distance trips, the same process with the trips-in-motion approach was conducted to determine TOD factors for long-distance trips that extended more than 50 miles.<sup>5</sup> Figure 23 presents the results of applying the trips-in-motion approach on long-distance trips to showcase the distribution of trips that occupies the roads on a

<sup>5</sup> The analysis excludes air travels and trips that occurred entirely out of Georgia boundaries. This led to a fewer number of trips used in the analysis (n=1,130) from the initial observations (n=1,338).

15-minute time bin. There are two prominent peaks of long-distance trips throughout the 24-hour timeframe at morning and evening peaks, which is similar to the observations from short-distance trips. Also, the distribution of trips based on a regular approach (figure 24) demonstrates the traditional pattern with two peaks during peak time periods. What differentiates the observation of the temporal distribution between the short- and long-distance trips in terms of the trips-in-motion approach is the notion that the “valley” between the two peaks in the long-distance trips (figure 23) seems to be less steep in comparison to the one observed in the short-distance trips (figure 10). This is an expected discrepancy since long-distance trips cover longer distances and durations than the short-distance trips and thereby would likely cover more 15-minute time bins, which spread farther than short trips and in turn yield the less pronounced “valley” between the two peak time periods.

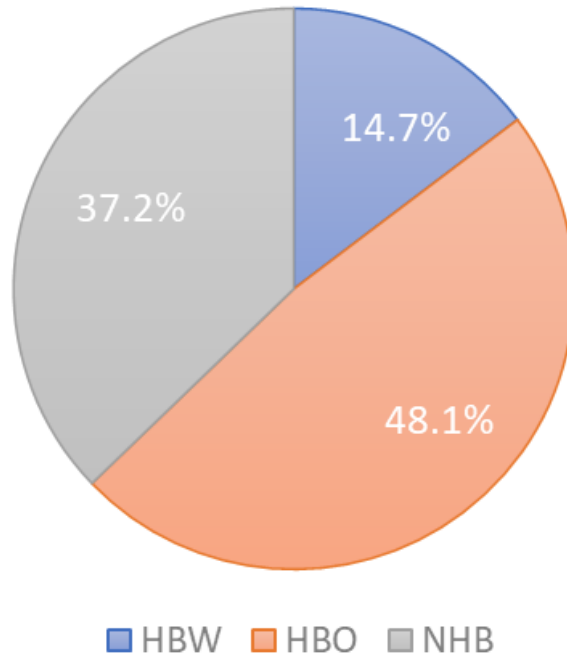


**Figure 23. Stacked histogram. Temporal distribution of trips by purpose based on “trips in motion” approach, long-distance trips.**



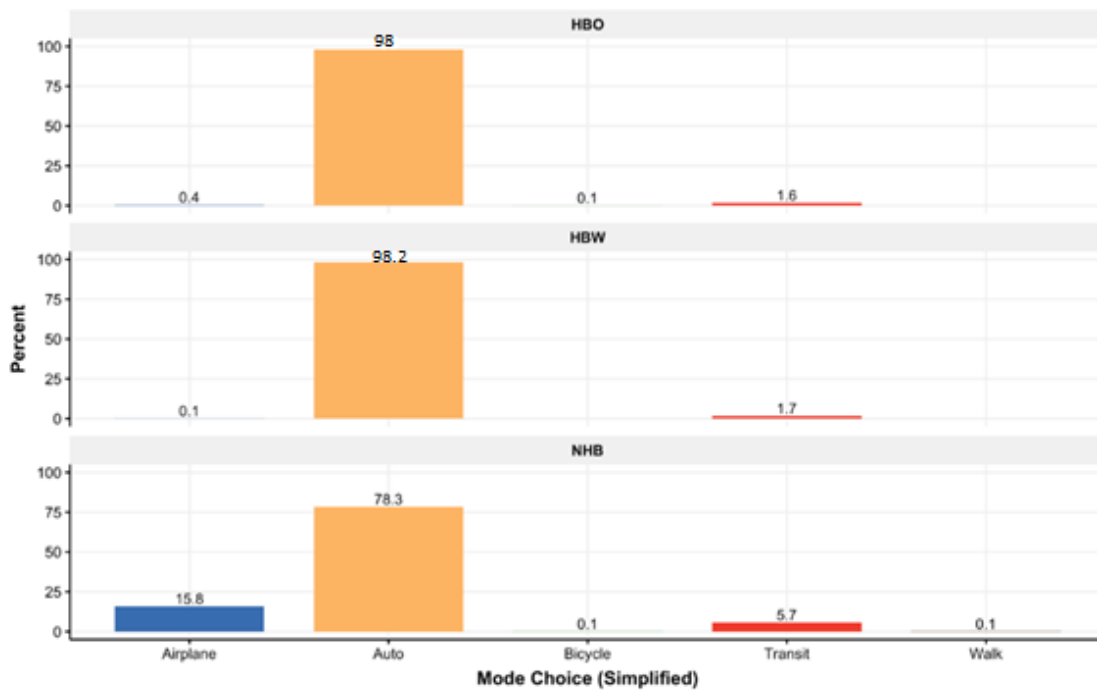
**Figure 24. Stacked histogram. Temporal distribution of trips by purpose based on a regular approach, long-distance trips.**

There is no significant difference in the purpose share between short- and long-distance trips, as illustrated in figure 25. HBO trips account for almost half of total trips (48.1 percent), followed by NHB (37.2 percent), and HBW (14.7 percent).



**Figure 25. Pie graph. Share of total long-distance trips by purpose.**

Breaking down the shares of total long-distance trips by purpose and mode demonstrates that a vast majority of travelers choose a travel mode of auto across purposes (figure 26). Specifically, 98 percent of trips for HBO and HBW are auto, and the share of auto trips for NHB constitutes 78.3 percent of total NHB trips. On the other hand, shares of other modes, including airplane, bicycle, transit, and walk, appear almost negligible; their portions are mostly lower than 2 percent, except for NHB whose shares of airplane and public transit are 15.8 and 5.7 percent, respectively.



**Figure 26. Bar charts. Share of total long-distance trips by mode and purpose.**

Trip shares by time period based on the trips-in-motion approach are presented in table 14. For instance, in comparison to the shares for short-distance trips (table 9), the difference between AM peak and Midday appears to be less dramatic. While the shares of HBW trips during the AM peak and Midday in the short-distance trips are 38.71 and 10.7 percent, respectively, those in the long-distance trips are 34.08 and 18.45 percent, respectively. Another finding that characterizes the long-distance trips is the greater share of trips occurring during the Night period than the ones observed in the short-distance trips. This is particularly apparent for the NHB trips where there were 14.36 percent in the long-distance trips in contrast to the 7.93 percent as observed in the short-distance trips.

**Table 14. Trip shares by time period and purpose (long-distance trips).**

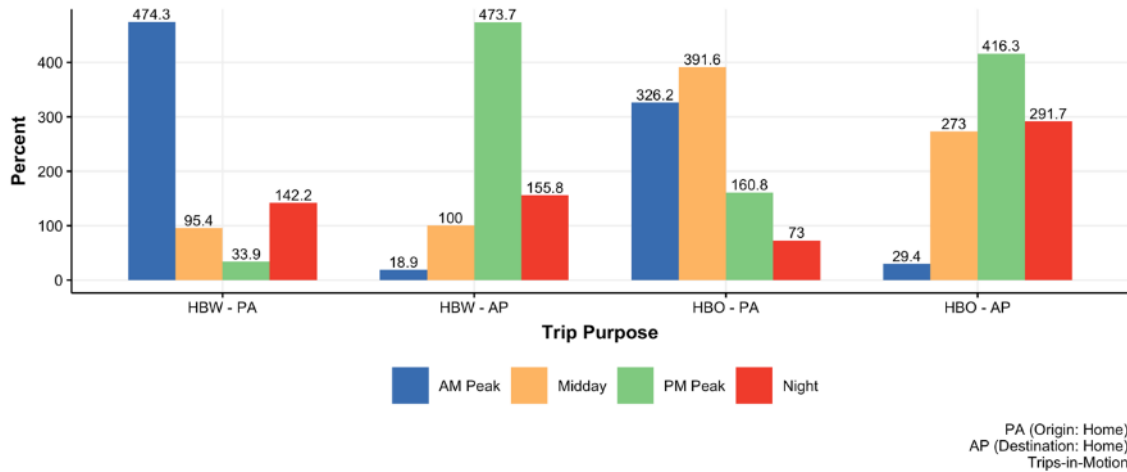
	<b>AM Peak (%)</b>	<b>Midday (%)</b>	<b>PM Peak (%)</b>	<b>Night (%)</b>
<b>HBO</b>	190.6	351.3	305.2	180.2
<b>HBW</b>	257.2	100.5	246.2	149.5
<b>NHB</b>	183.9	423.8	302.6	141.3

*Spatial Distribution of Trips by Time Period*

As described above, the amounts of production and attraction trips which take directionality into account depend heavily on time of day, and their pattern is significantly opposite from each other (figure 27). Not surprisingly, most HBW trips occur during the two peak time periods, but the pattern between PA and AP is reversed, as expected. Of trips from home to workplace, 65.4 percent are generated during the AM peak, whereas only 0.7 percent of trips from home to workplace are generated during the PM peak. Likewise, the share of HBW AP trips (from workplace to home) at the PM peak accounts for 60.2 percent, while the share of HBW AP trips at the AM peak constitutes only 4.7 percent.

Note that PA and AP classifications for NHB are not presented here, as NHB trips do not have directionality (i.e., there is no anchor station for NHB trips, so neither production nor attraction point can be defined).

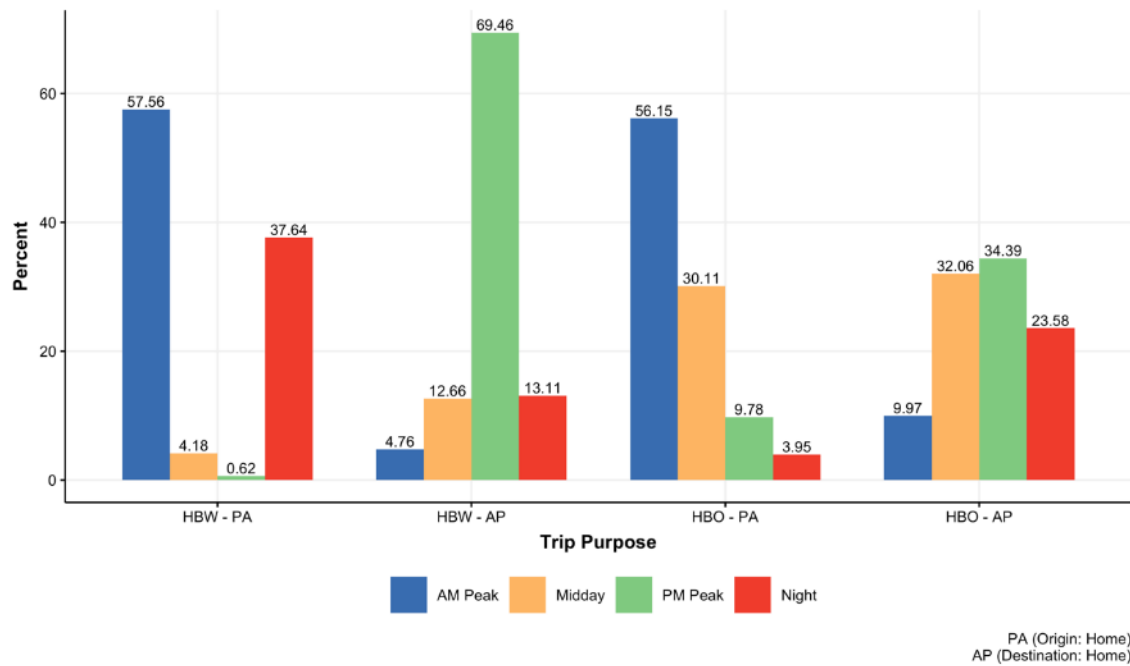




**Figure 27. Bar charts. Share of total long-distance trips by directionality.**

### ***Final TOD Factors***

As the research team did for finalizing the short-distance TOD factor, individual long-distance trips are weighted based on the person-level weights (“WTPERFIN”), which is provided by the original NHTS dataset. Summing all trips by TOD and purpose, the research team calculate the final TOD factors after considering directionality, as presented in figure 28 and table 15. The final factors (the right two columns in table 15) can be used in the traditional four-step travel demand model. In particular, purpose- and period-specific number of trips can be obtained using the factors after the trip-generation step in the modeling process, along with a proposed TOD implementation method (refer to the TOD Implementation After Trip Generation subsection below).



**Figure 28. Bar chart. TOD factors, long-distance trips (weighted).**

**Table 15. TOD factors, long-distance trips (weighted).**

	Volume		Distribution	
	PA	AP	PA	AP
<b>HBW</b>				
<b>AM Peak</b>	31,183	2,366	57.56	4.76
<b>Midday</b>	2,265	6,291	4.18	12.66
<b>PM Peak</b>	339	34,508	0.62	69.46
<b>Night</b>	20,392	6,515	37.64	13.11
<b>Sum</b>	<b>54,179</b>	<b>49,680</b>	<b>100</b>	<b>100</b>
<b>HBO</b>				
<b>AM Peak</b>	62,231	13,626	56.15	9.97
<b>Midday</b>	33,908	43,824	30.11	32.06
<b>PM Peak</b>	11,010	47,004	9.78	34.39
<b>Night</b>	4,453	32,228	3.95	23.58
<b>Sum</b>	<b>111,602</b>	<b>136,682</b>	<b>100</b>	<b>100</b>
<b>NHB</b>				
<b>AM Peak</b>	74,369	-	31.00	- <sup>a</sup>
<b>Midday</b>	76,676	-	31.96	-
<b>PM Peak</b>	60,139	-	25.07	-
<b>Night</b>	28,718	-	11.97	-
<b>Sum</b>	<b>239,902</b>	<b>-</b>	<b>100</b>	<b>-</b>

<sup>a</sup> The final traffic volumes and corresponding TOD factors for NHB AP trips are not presented due to the lack of directionality, but TOD factors for NHB PA can be applied to the same category of NHB AP.

### Through Trips (External to External Trips)

A through trip refers to interregional travel for which both trip ends are located out of Georgia. Since the dataset used in this research only includes trips within Georgia or partly in Georgia (referred to as internal trips for which either trip end or both ends is located in Georgia or both), TOD factors for external trips must be defined using external data sources. The research team amassed data collecting traffic counts; as a result, the source from the Traffic Analysis & Data Application (TADA), which provides data collected from the Georgia Traffic Monitoring

Program, located on public roads is selected. To determine representative TOD factors for external trips, traffic counts need to be aggregated. There are several ways to sum up the external trips counted by loop detectors, and the research team reviewed two different approaches: (1) using the most representative month and date with the least variation, and (2) annual average traffic counts.

Regarding the first approach, two traffic adjustment factors, monthly and daily factors, are used to investigate what month and date show the least variation (i.e., representative), thereby choosing the most representative day to collect traffic count data given that day. GDOT (2018) calculated both factors by dividing the annual average daily traffic (AADT) by the daily (or monthly) average traffic, meaning how much traffic volume for each day (or month) need to be adjusted to determine the AADT. The result shows that Monday, April, and September demonstrate the least fluctuation (refer to table 16 and table 17).

Based on the result above, specific traffic counts by TOD could be obtained by choosing the most representative specific date or computing the average of every applicable date. However, it might raise a question about reliability of this method because there is no clear guideline specifying the best method to aggregate traffic counts by TOD (i.e., it depends highly on the analysts' decision and judgment). Thus, the research team has chosen the second approach. Hourly average traffic counts from the Georgia state border collected from January 1 to December 31, 2017, are used to determine the final TOD factors for through trips.

**Table 16. 2016 daily factors by road hierarchy in Georgia.**

Road Type <sup>a</sup>	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	1.22	1.01	0.98	0.98	0.94	0.89	1.04
2	1.22	1.01	0.99	0.98	0.94	0.88	1.05
3	0.92	1.08	1.15	1.10	1.01	0.82	0.97
4	1.37	0.97	0.92	0.92	0.91	0.90	1.17
5	1.26	1.00	0.99	0.98	0.94	0.87	1.04
6	1.16	1.02	1.01	0.99	0.95	0.88	1.04
7	1.43	0.98	0.93	0.92	0.92	0.88	1.13
8	1.36	0.99	0.94	0.93	0.92	0.90	1.10
9	1.19	1.01	0.98	0.97	0.95	0.91	1.04
Mean	1.24	1.01	0.99	0.97	0.94	0.88	1.06
SD	0.15	0.03	0.07	0.05	0.03	0.03	0.06

Note: The highlighted row shows the least variation of traffic counts.

<sup>a</sup> 1=collectors and locals in rural, 2=arterials in rural, 3=interstates in rural, 4=collectors in urban, 5=arterials in urban, 6=interstate in urban, 7=arterials in urban, 8=arterials in Atlanta, 9=interstates in Atlanta.

**Table 17. 2016 monthly factors by road hierarchy in Georgia.**

Road Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	1.10	1.09	0.99	0.99	0.96	0.98	0.97	0.96	0.97	0.90	1.07	1.07
2	1.11	1.03	0.99	0.98	0.97	0.98	0.99	1.00	0.99	0.92	1.01	1.05
3	1.17	1.12	0.97	0.98	0.96	0.92	0.88	1.02	1.03	0.99	1.00	1.04
4	1.08	1.01	0.96	0.97	0.96	1.00	1.02	0.95	0.98	0.98	1.06	1.06
5	1.10	1.01	0.98	0.98	0.96	0.98	0.99	0.97	1.00	0.96	1.04	1.05
6	1.10	1.04	0.96	0.97	0.97	0.96	0.94	1.00	1.02	1.01	1.03	1.03
7	1.07	0.98	0.96	0.96	0.97	1.00	1.01	0.98	1.00	1.00	1.04	1.05
8	1.07	1.00	0.97	0.98	0.97	0.99	1.02	0.97	0.98	0.96	1.05	1.06
9	1.07	1.02	0.98	0.98	0.97	0.97	0.99	0.99	1.00	0.99	1.04	1.03
Mean	1.10	1.03	0.97	0.98	0.97	0.98	0.98	0.98	1.00	0.97	1.04	1.05
SD	0.03	0.04	0.01	0.01	0.01	0.02	0.04	0.02	0.02	0.04	0.02	0.01

Note: The highlighted row shows the least variation of traffic counts.

<sup>a</sup> 1=collectors and locals in rural, 2=arterials in rural, 3=interstates in rural, 4=collectors in urban, 5=arterials in urban, 6=interstate in urban, 7=arterials in urban, 8=arterials in Atlanta, 9=interstates in Atlanta.

In order to count trips traveling across Georgia borders, 14 traffic count stations out of 18 stations in TADA are selected, as presented in table 18 and figure 29. Those stations are located on main arterials or interregional freeways. Traffic counts are summed up by hour across 14 stations (figure 30), then the share of each TOD is calculated, as presented in table 19.

**Table 18. List of traffic count stations for external trips.**

#	ID	Roadway type	Description
1	083-0194	Inter (rural)	I-59/SR406 bn Pudding Rdg Rd & SR136 MP 8
2	083-0209	Inter (rural)	I-24/SR409 bn TN SL & I-59, Trenton, Dade Co
3	083-0214	Inter (urban)	I-24 bn TN State Line & SR299 W Side, Chattanooga
4	047-0114	Inter (urban)	I-75 btwn SR146 & Tennessee line
5	147-0287	Inter (rural)	I-85 btwn Hart/Franklin Co. Line & SR77 Whitworth Rd
6	245-0218	Inter (urban)	I-20 E of I-520 @SC state line, Augusta
7	245-0233	Inter (urban)	I-520 S of SC state line & SR28 nr Foster Ln, Augusta
8	051-0387	Inter (urban)	I-95, 2 mi N of SR-21 (Augusta Rd) @ SC state line
9	039-0218	Inter (urban)	I-95: bn FL SL & St Marys Rd, Kingsland, GA
10	065-0125	Arterial (rural)	US-441/SR89/Barton St S of SR94, Fargo, Clinch Co
11	185-0227	Inter (rural)	I-75/SR401 @FLA SL, Lake Park, Lowndes Co
12	087-0103	Arterial (rural)	US-27/SR1 S of US-27BU/SR1BU/E Griffin Ave
13	145-0234	Inter (urban)	I-85/SR403 1 mi E of AL state line, West Point
14	143-0126	Inter (rural)	I-20 btwn Alabama state line & SR100 Veterans Mem Hwy

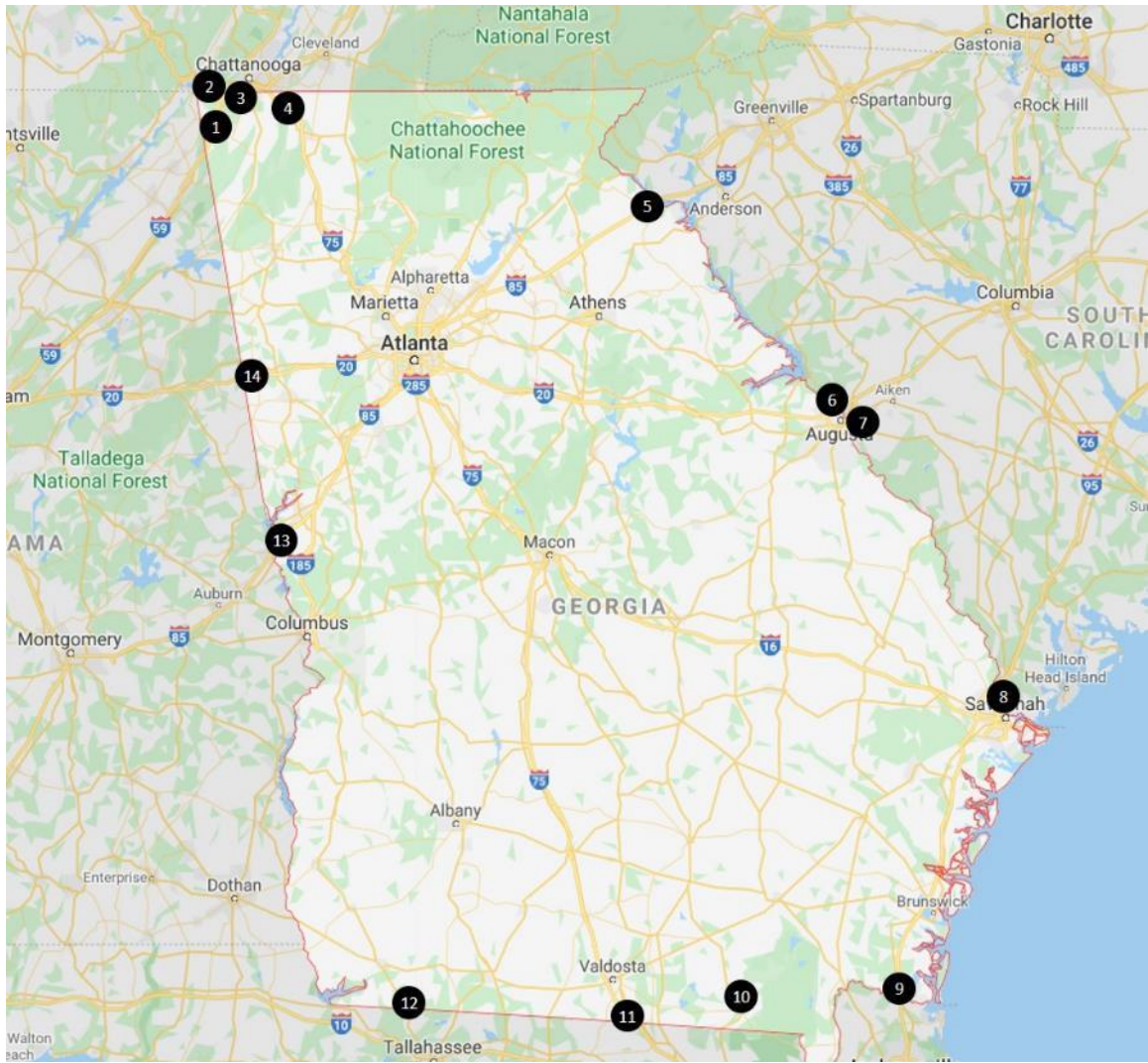


Figure 29. Map. Location of traffic count stations in Georgia.

Site ID	Site Name	Location	Direction	00:00:00	01:00:00	02:00:00	03:00:00	04:00:00	05:00:00	06:00:00	07:00:00	08:00:00	09:00:00	10:00:00	11:00:00
000000390218	039-0218	30.75257, -81.65076	All directions	784	627	560	617	923	1674	2624	2885	2926	3550	4048	4257
000000390218	039-0218	30.75257, -81.65076	North	405	321	290	322	482	930	1487	1521	1559	1945	2212	2276
000000390218	039-0218	30.75257, -81.65076	South	379	306	270	295	440	744	1137	1364	1367	1605	1836	1982
000000470114	047-0114	34.98511, -85.20087	All directions	1297	1018	924	1023	1418	2275	4130	5862	5347	5233	5727	6105
000000470114	047-0114	34.98511, -85.20087	North	604	470	430	482	735	1343	2602	3580	3081	2732	2906	3038
000000470114	047-0114	34.98511, -85.20087	South	692	547	494	542	683	932	1529	2281	2266	2501	2821	3067
000000510387	051-0387	32.2002, -81.18769	All directions	755	586	606	572	749	1107	1913	2544	2806	3092	3523	3860
000000510387	051-0387	32.2002, -81.18769	NorthEast	410	311	314	277	353	597	1160	1492	1575	1661	1855	2062
000000510387	051-0387	32.2002, -81.18769	SouthWest	345	275	292	295	395	511	753	1052	1230	1431	1667	1799
000000650125	065-0125	30.67896, -82.558	All directions	7	7	8	12	24	43	71	76	93	96	96	102
000000650125	065-0125	30.67896, -82.558	NorthWest	4	3	4	5	10	21	41	35	38	45	50	54
000000650125	065-0125	30.67896, -82.558	SouthEast	3	4	4	7	14	22	31	41	55	52	46	48
000000830194	083-0194	34.82233, -85.53962	All directions	206	165	152	156	216	336	499	635	656	759	859	934
000000830194	083-0194	34.82233, -85.53962	NorthEast	106	85	77	83	124	200	317	380	334	371	422	466
000000830194	083-0194	34.82233, -85.53962	SouthWest	100	80	75	73	92	137	182	255	322	388	437	468

Figure 30. Screenshot. Data format of hourly averages report from Georgia state border in 2017.

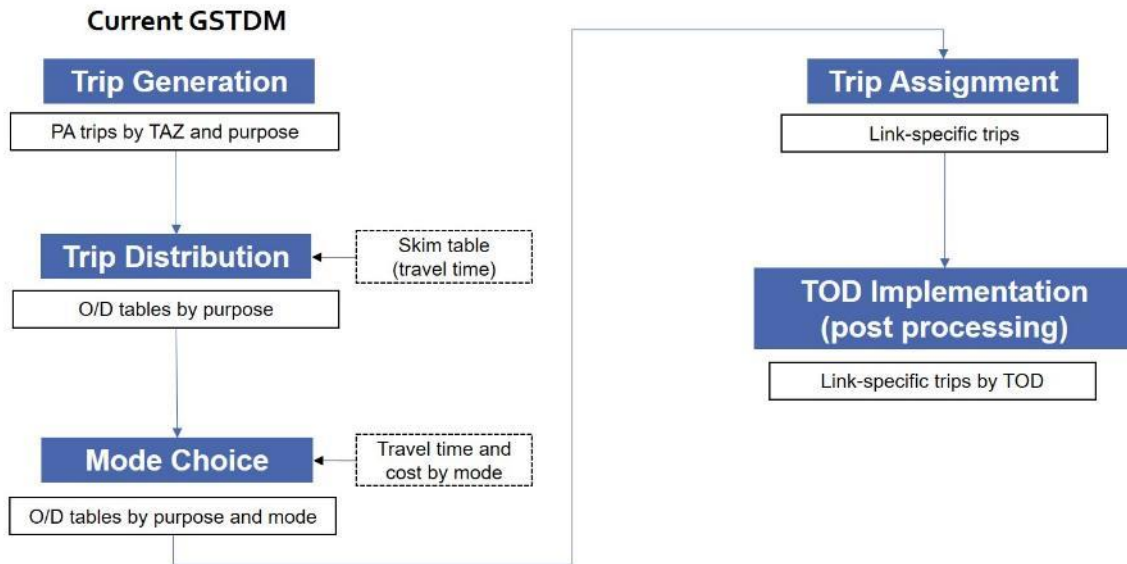
**Table 19. Final TOD factors for through trips.**

TOD	AM	Midday	PM	Night
TOD Factors	18.6	30.6	26.3	24.5

Note: The trips-in-motion approach is not applied to through trips as each trip in traffic counting data was counted only one time when it passed detectors.

### TOD Implementation After Trip Generation

In the current GSTDM, TOD implementation is conducted after the trip-assignment step, through appropriate postprocessing, to calculate AM and PM peak period trips (figure 31). That is, the current approach simply divides the GSTDM daily trip outputs, which are obtained after trip assignment, into period-specific traffic volumes. The process includes: (1) creating the AM and the PM peak trips from the daily trip tables using the TOD factors, (2) updating the network with peak-period capacities, and (3) performing trip assignment for two peak periods.



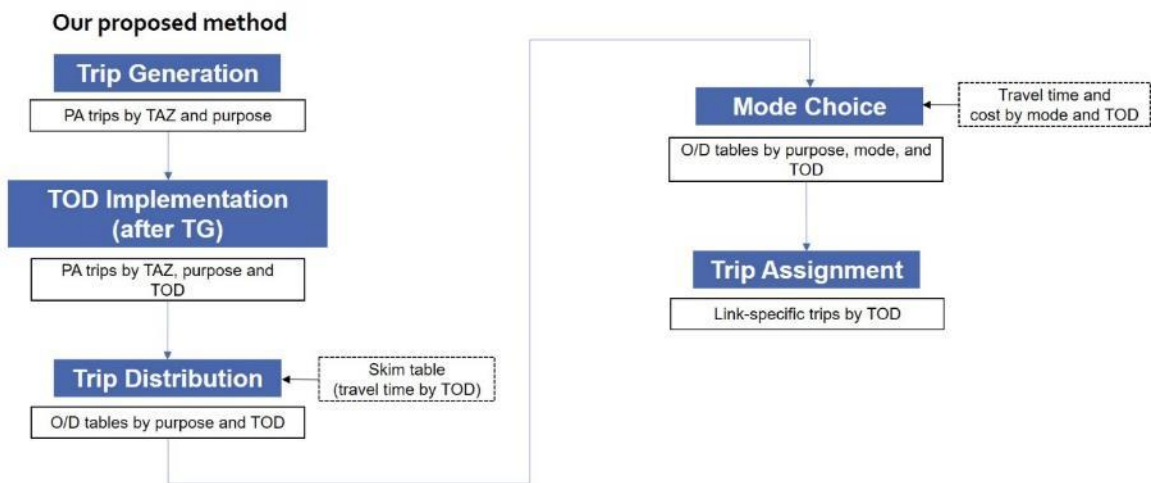
**Figure 31. Flowchart. TOD implementation in the current GSTDM.**

One limitation of this approach is that TOD factors cannot be applied in the intermediate steps of the analysis, such as after the trip-generation or distribution steps. A problem of the



postprocessor results in that it cannot account for peak time congestion in the process of travel demand estimation since travel time and cost variables (summarized in the skim matrices) are defined based on the 24-hour travel patterns, meaning that there is a discrepancy between the four-step process and the postprocessing.

Therefore, the research team proposed a method allowing modelers to specify traffic congestion conditional on each peak time period in the trip-distribution, mode-choice, and trip-assignment steps by applying the TOD implementation right after the trip-generation step to predict more accurate peak-period traffic volumes (figure 32). The key improvement of this approach is to calculate TOD-specific travel time and cost accounting for traffic congestion on roadways. Consequently, TOD-specific trip-distribution and mode-choice analyses can be carried out using travel time and cost variables by TOD. Of course, it requires extensive time and effort to compute TOD-specific mode attributes, trip tables, and statistical modes. Thus, it has to be discussed with modelers, practitioners, and decision-makers to meet their requirement.



**Figure 32. Flowchart. Proposed method of the TOD implementation in GSTDM.**

## **CHAPTER 5. EVALUATING THE INCLUSION OF A DESTINATION CHOICE MODEL**

The trip-distribution step in the four-step travel demand modeling framework has shown to be the largest source of error in travel demand modeling (Zhao and Kockelman 2002), therefore mandating further improvements in the accuracy of this step in the GSTDM. Although the gravity model, largely due to its simplicity and theory-based application, has been the prevailing method of distributing trips in most regional models, it tends to underperform compared to other newer methods of trip distribution. Among these other trip-distribution methods, specifically, destination choice models have gained more traction in regional models and have shown to improve trip-distribution accuracy (Mishra et al. 2013). The goal of this task, therefore, is to investigate and evaluate the application of a destination choice model in the GSTDM, and provide guidance on the implementation of this model in the overall statewide modeling framework.

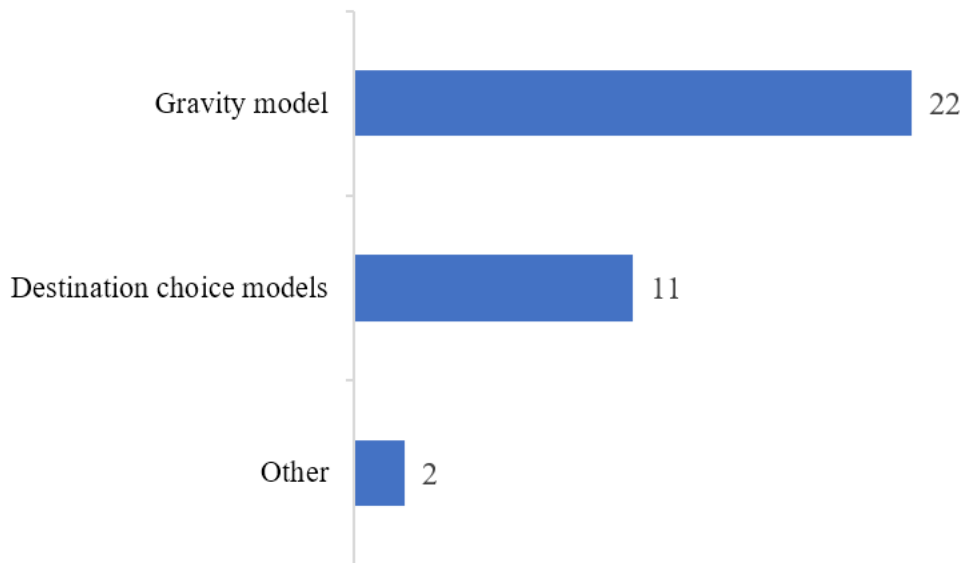
GDOT Report 18-24 (Kash, Mokhtarian, and Circella 2021) provides a basic exploration of trip patterns by location of travel in Georgia, such as trip frequencies between MPO tiers and vehicle miles traveled of trips in Georgia. Readers are encouraged to refer to GDOT Report 18-24 for further initial exploration of the 2017 NHTS data.

### **CURRENT STATUS OF TRIP DISTRIBUTION IN THE GSTDM**

The GSTDM currently uses the gravity model structure to distribute trips. The output of the trip-generation step is segmented by trip purpose (HBW, HBO, and NHB), and a separate gravity model is estimated for each segment. The current version of the GSTDM does not consider time-of-day in its trip distribution step.

## REVIEW OF OTHER STATEWIDE MODELS

A recent survey of statewide models in the U.S. shows that the gravity model continues to be the dominant method of trip distribution (Moeckel, Donnelly, and Ji 2019), with destination choice models the second-most popular method. As figure 33 shows, 22 states use the gravity model structure, while 11 use the destination choice model. These 11 states include Arizona, California, Idaho, Iowa, Maryland, New Hampshire, Ohio, Oregon, Tennessee, Wisconsin, and Florida. Of the 11 states that use (logit-based) destination choice models, 8 only use it for short-distance trips and use the gravity model for long-distance trips. Such a combination generally ensures the largest model sensitivities for the trip-distribution step (Moeckel, Donnelly, and Ji 2019). In line with such conclusions, we recommend that GDOT keeps its long-distance gravity model, and consider using a destination choice model for the short-distance trips.



**Figure 33. Bar graph. Distribution of the U.S. statewide travel demand models based on their trip distribution model.**

## **DATA PREPARATION**

Similar to the other tasks in this project, the researchers used the 2017 NHTS as the main data source for the estimation and evaluation of the destination choice model. We used the 2017 NHTS *trip* file for the state of Georgia, and used the geocoded origin and destination of the trips to associate them with the TAZ structure used in the GSTDM.

The next step, and arguably the most important step in the data preparation for destination choice models, involved defining the destination choice set. Although it might be feasible to include all available destination TAZs in the destination choice for trips (especially in smaller models), it becomes increasingly difficult in larger statewide models. Statewide models often include thousands of TAZs, and considering all of them as possible destinations can both run counter to intuition (since an individual making a trip cannot possibly consider thousands of options to make a decision) and be intractable in terms of data preparation and estimation. We, therefore, devised a scheme to limit the destination choice set of the trips to a more manageable number.

### **Choice Set Formation**

In order to limit the size of destination choice sets in our model, we took two steps. First, for each origin TAZ, since we were dealing with short-distance trips, we only included destination TAZs within its 50-mile radius. Afterward, we used a probability sampling process to sample a limited number of destinations for each origin. Previous studies in the literature use varying destination choice set sizes, with set sizes ranging from 0.7 to 14.5 percent of all the available alternatives (Kim and Lee 2017). In this study, we test different choice set sizes (i.e., 10, 20, and 30) and evaluate their impact on the model specification and fit.

The two common probability sampling processes used in destination choice studies include simple random sampling and importance sampling. In random sampling, each alternative (or in our context, a TAZ), has an equal probability of entering the choice set. In importance sampling, however, we assign a weight (or importance) to each alternative based on its attraction level, and carry out a weighted sampling based on the calculated weights (Ben-Akiva and Lerman 1985). In this task, we use the importance sampling scheme to form the destination choice set, since previous studies point out that this sampling method is superior to the simple random sampling in destination choice studies (Bowman and Ben-Akiva 2001).

### ***Importance Sampling***

Conceptually, importance sampling operates on the assumption that not all alternatives are created equally. In other words, for a given origin, and as an example, destinations that are closer or are more attractive (more densely populated or more employment opportunities) should have a higher probability of entering the choice set than others. We, therefore, define an importance function to assign a weight to each destination based on the origin TAZ. For an origin zone  $i$ , we define the weight of destination  $j$  using the following function:

$$w_{ij} = a_j \times \exp\left(\frac{-2d_{ij}}{d}\right) \quad (5)$$

In the above formula,  $w_{ij}$  denotes the weight of destination  $j$  with respect to origin  $i$ , and  $a_j$  is the attraction level of destination  $j$  and is defined as the summation of its population and employment opportunity counts.  $d_{ij}$  is the distance between the origin and destination zones, and  $d$  is the average distance between the TAZs in the region.

Based on the calculated weights, we assign an importance probability to each destination TAZ of an origin using equation (6):

$$p_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} \quad (6)$$

In the above equation,  $p_{ij}$  denotes the importance probability of destination  $j$  with respect to origin  $i$ . Given the calculated importance probabilities, we carry out a weighted random sampling (Efrimidis and Spirakis 2008) of all the available alternatives for an origin. Literature shows that this sampling can be done both *with* or *without* replacement (Kim and Lee 2017). Although sampling without replacement tends to be the default sampling method, it can result in dependent alternatives. Studies, however, show that these two methods provide similar results when the population is large and the number of sampled alternatives is less than 5 percent of the total population. Considering that the number of available alternatives for a TAZ tends to be around 1000 in our model, and we are sampling less than 50 alternatives, both methods should provide similar results. We, nevertheless, carry out the weighted sampling *with* replacement for this study.

When carrying out a *weighted* sampling scheme for a discrete choice model, a correction term must be added to the utility function to ensure that the model yields unbiased estimates. We discuss this correction more in detail in the Method section.

### **Dataset Augmentation**

The next step in the data preparation process involved adding the travel impedance variable associated with each trip. We used the 2015 GSTDM distance skim matrix, available in the

Cube<sup>®</sup> model files, and joined the distances between origin–destination (O–D) TAZs for each trip to our data.

Another set of variables commonly used in destination choice models are size variables (or size terms), which include those such as employment size and population of a TAZ that help define its attraction as a destination. We used the base-year socioeconomic data used in the 2015 GSTDM and added variables on population, household, employment (by category), and TAZ size to our destination choice dataset.

Furthermore, we added some household-level characteristics, such as vehicle ownership and annual household income, to the dataset. These variables help capture the heterogeneity in destination choice among different segments of the population with differing mobility levels. We included these variables both at the household level and also at the aggregate (TAZ) level so as to make the implementation more flexible. We used the results of task 1 (see chapter 3) to add the aggregate vehicle ownership (average vehicle ownership per TAZ) measure and used ACS data to obtain aggregate income distribution per TAZ.

Finally, we used the Census TIGER shapefiles with geographical feature data to identify TAZs with special features, such as parks and trails, military bases, airports, and colleges, and joined these features to our dataset. These indicators help better capture the trip-distribution patterns in the model and act as “geographical” constants in the logit-based model, especially since the inclusion of alternative-specific constants (as discussed below) are computationally intractable.

## METHOD

Destination choice models are often formulated using random utility theory, with the logit framework used to compute the probability by which each destination (zone) might be chosen for a specific trip. The probability of zone  $j$  being selected for trip  $n$  conditional on the choice set  $c_n$  and a set of explanatory variables  $x$  is:

$$p_n(j|c_n, x) = \frac{\exp(V_{jn} + \ln q(C_n|j))}{\sum_{j \in c_n} \exp(V_{jn} + \ln q(C_n|j))} \quad (7)$$

In the above equation,  $V_{jn}$  is the utility function associated with destination  $j$  for trip  $n$ , and is defined as:

$$V_{jn} = \alpha_j + \sum_k \beta_k d_{jn}^k + \sum_l \beta_l x_{nl} d_{jn} + \gamma_j \ln(\text{Size}_j) \quad (8)$$

$d_{jn}^k$  denotes the distance polynomial term of order  $k$  for destination  $j$  and trip  $n$  and  $\beta_k$  is the associated coefficient.  $x_{nl}$  denotes the socioeconomic variables ( $l = 1, \dots, L$ ) of the tripmaker  $n$  (such as income or vehicle ownership), which is interacted with the distance term, with  $\beta_l$  showing the associate coefficient of each interaction term. Finally,  $\text{size}_j$  shows the attraction variable set for destination  $j$ . The reason the natural log transformation of the attraction variables is usually used in the model is to allow for a direct linear relationship between the size variables and trip distribution shares given the exponential function of the logit formulation.

The second term in the exponential function in equation 7, i.e.,  $\ln q(C_n|j)$ , is the sampling correction term needed to obtain unbiased estimates in the presence of nonrandom sampling of alternatives. This correction is calculated as follows (Frejinger, Bierlaire, and Ben-Akiva 2009):



$$\ln q(C_n|j) = - \ln \frac{p_n(j)}{k_{jn}} \quad (9)$$

In equation 9,  $p_n(j)$  is the sampling probability of alternative  $j$  for trip  $n$  (equation 6), and  $k_{jn}$  is the number of times alternative  $j$  is drawn for the choice set of trip  $n$ . This probability is calculated before model estimation, and enters the estimation process with a fixed coefficient. In the Results section below, we have added the  $\ln \frac{p_n(j)}{k_{jn}}$  term directly to the utility function and fixed its coefficient to  $-1$ .

## RESULTS

In this task, we estimated separate models based on trip purpose and time of day. There are three trip purposes in the GSTDM framework: home-based work trips, home-based other trips, and non-home-based trips. For each of these trip purposes, we estimated four destination choice models based on the defined time-of-day trip periods. As discussed in chapter 4, the developed four time-of-day trip periods include the AM peak, Midday, PM peak, and Night trips. We, therefore, estimated a total of 12 different destination choice models.

### Home-based Work Models

The total number of HBW trips in our dataset is 6912. As mentioned, the HBW trips are divided into four time-of-day segments. Table 20 shows the destination choice model for the home-based work trips during the morning (AM) peak period. The impedance term used in the model, as discussed previously, is the distance among the TAZs. This term is used in the polynomial to help better capture the nonlinear relationship between the destination choice and distance. Based

on the signs of the polynomial terms, we see a negative relationship between the choice of destination and its distance to the origin of a decision-maker.

We further interacted two socioeconomic characteristics with the distance term to capture the impacts of important socioeconomic variables on destination choice. As the results show, households with lower income tend to choose closer destinations, while those with a higher number of vehicles are more likely to travel farther for their work.

Furthermore, we see that TAZs with a lower population density are more likely to be among the destination choices in the morning peak trips. The reason is the work nature of trips in the morning, when the majority of trips are headed to work locations as opposed to heading back home. We also see that TAZs with a higher nonretail employment concentration tend to attract more workers during the morning peak, while nonretail employment shows a weak (negative) association with higher likelihood of a TAZ to be chosen.

Finally, the geographical constant terms in our model show that TAZs with military bases or colleges in them are more likely to be chosen as morning peak work destinations, while TAZs with commercial airports are less likely.

**Table 20. Destination choice model for the HBW, AM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.178	0.010	-17.67	<0.001
Distance Squared	0.00078	0.0002	4.89	<0.001
Distance×Low-Income (<\$25K)	-0.0150	0.008	-1.81	0.070
Distance×VO	0.000632	0.003	1.98	0.048
<i>Size Terms</i>				
Ln(Population Density)	-0.12031	0.015	-8.20	<0.001
Ln(Retail Jobs)	-0.0225	0.017	-1.33	0.183
Ln(Non-Retail Jobs)	1.336	0.023	57.38	<0.001
<i>Geographical Constants</i>				
Parks	0.0565	0.065	0.87	0.387
Colleges	0.27897	0.062	4.51	<0.001
Commercial Airports	-0.261	0.123	-2.12	0.034
Military Bases	4.509	0.335	13.46	<0.001
N= 3007				
LL(0) = -8854.20				
LL(β) = -7774.33				
$\rho^2 = 0.122$				

<sup>a</sup> SE is the standard error

Table 21 shows the destination choice model for HBW trips in the PM peak period. With respect to the impact of distance and socioeconomics on destination choice, we see a similar pattern as discussed for the AM peak trips: more distant TAZs are less likely to be chosen and those living in higher income households or with higher number of vehicles are more likely to travel longer to their destinations. In contrast to the trips in the morning period, however, we see that those TAZs with higher population densities are more likely to be a chosen destination. This result

points to the fact that most trips in the PM peak are headed toward home, and TAZs with a higher residential population density tend to be among the destinations.

With respect to employment, we see that the number of manufacturing and retail jobs have an insignificant relationship with the PM peak destination choice, while TAZs with a higher number of agricultural and service jobs are more likely to be among the destinations.

Among the geographical constants, finally, we see that the college constant is the only significant one, with its negative sign indicating that TAZs with a college in them are less likely to be among HBW trip destinations in this time period.

**Table 21. Destination choice model for the HBW, PM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.189	0.010	-18.280	<0.001
Distance Squared	0.000450	0.000	2.570	0.010
Distance×Low-Income(<\$25K)	-0.0298	0.009	-3.270	0.001
Distance×VO	0.0229	0.003	7.020	<0.001
<i>Size Terms</i>				
Ln(Population Density)	0.0418	0.017	2.510	0.012
Ln(Retail Jobs)	-0.00645	0.020	-0.330	0.744
Ln(Service Jobs)	0.540	0.027	19.900	<0.001
Ln(Manufacturing Jobs)	-0.0144	0.017	-0.860	0.389
Ln(Agricultural Jobs)	0.304	0.021	14.830	<0.001
<i>Geographical Constants</i>				
Parks	-0.0870	0.078	-1.120	0.264
Colleges	-.339	0.087	-3.880	<0.001
Commercial Airports	0.0632	0.142	0.440	0.657
Military Bases	0.346	0.509	0.680	0.497
N = 2471				
LL(0) = 7402.45				
LL(β) = 6639.08				
$\rho^2 = 0.103$				

<sup>a</sup> SE is the standard error

Table 22 and table 23 show the results for the Midday and Night periods. In both models, we see similar relations between distance and socioeconomic variables with destination choice, with the difference that the Night period model does not have the polynomial distance term due to statistical insignificance. In the Midday period model, we see that retail job counts, unlike the

other models, have a positive association with destination choice, indicating that HBW trips to retail jobs are more likely to take place during the Midday period.

**Table 22. Destination choice model for the HBW, Midday period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1	Fixed Parameter		
<i>Impedance and Interaction Terms</i>				
Distance	-0.270	0.020	-13.580	<0.001
Distance Squared	0.00199	0.000	6.170	<0.001
Distance×Low-income(<\$25K)	-0.0280	0.018	-1.550	0.122
Distance×VO	0.0106	0.006	1.660	0.098
<i>Size Terms</i>				
Ln(Population Density)	-0.0526	0.026	-2.000	0.045
Ln(Retail Jobs)	0.0642	0.032	2.020	0.043
Ln(Non-Retail Jobs)	0.923	0.041	22.460	<0.001
<i>Geographical Constants</i>				
Parks	0.159	0.119	1.330	0.183
Colleges	0.1241	0.118	1.050	0.293
Commercial Airports	0.0064	0.216	0.030	0.976
Military Bases	3.828	0.549	6.970	<0.001
N = 924				
LL(0)= 2711.07				
LL(β)= 2454.19				
ρ <sup>2</sup> = 0.095				

<sup>a</sup> SE is the standard error

**Table 23. Destination choice model for the HBW, Night period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.134	0.012	-11.030	<0.001
Distance×Low-income(<\$25K)	-0.0294	0.011	-2.720	0.007
Distance×VO	0.0121	0.005	2.650	0.008
<i>Size Terms</i>				
Ln(Population Density)	-0.0745	0.022	-3.440	0.001
Ln(Retail Jobs)	0.00568	0.027	0.210	0.836
Ln(Non-Retail Jobs)	0.978	0.035	27.970	0.000
<i>Geographical Constants</i>				
Parks	0.169	0.105	1.620	0.106
Colleges	-0.188	0.117	-1.610	0.108
Commercial Airports	-0.366	0.199	-1.840	0.066
Military Bases	4.827	0.466	10.360	<0.001

N=1120

LL(0)= 3301.02

LL(β)=3129.41

$\rho^2 = 0.052$

<sup>a</sup> SE is the standard error

### **Home-based Other Models**

The total number of HBO trips in our dataset is 29,764 cases. Similar to the HBW case, we segmented the trips into four segments based on the time-of-day periods. Table 24 shows the destination choice model for HBO trips during the morning peak. The distance polynomial variables show that, as expected, further TAZs are less likely to be chosen. Unlike other models, however, the impact of socioeconomic variables (through their interaction with the distance term) was too statistically weak to be included in the model.

We, furthermore, observe a positive association between both the population density and retail and nonretail employment counts and a higher likelihood of being chosen the destination, indicating that for HBO trips in the morning period, residential activity and employment activity both play a positive role in trip attraction.

The geographical constants, in addition, show no statistical significance except for the military-base indicator. The TAZs with a military base in them are more likely to be destinations for HBO trips in the morning.

**Table 24. Destination choice model for the HBO, AM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.411	0.005	-78.670	<0.001
Distance Squared	0.00473	0.000	38.480	<0.001
<i>Size Terms</i>				
Ln(Population Density)	0.468	0.018	26.510	<0.001
Ln(Retail Jobs)	0.0394	0.013	3.030	0.002
Ln(Non-Retail Jobs)	0.743	0.018	40.870	<0.001
<i>Geographical Constants</i>				
Parks	-0.00181	0.052	-0.030	0.973
Colleges	-0.0330	0.050	-0.670	0.505
Commercial Airports	0.0992	0.090	1.100	0.270
Military Bases	2.585	0.309	8.370	<0.001
N = 6824				
LL(0) = 20006.64				
LL( $\beta$ ) = -14567.02				
$\rho^2 = 0.272$				

<sup>a</sup> SE is the standard error



The PM peak model, as shown in table 25, shows similar relationship between distance and likelihood of being chosen as the destination. Moreover, we see a positive association between residential density and nonretail employment and destination choice in the PM peak period, while retail employment shows a statistically insignificant relationship in this period's model.

With respect to geographical constants, we see that TAZs with parks, commercial airports, and military bases are more likely to be among the destinations in the PM peak than the ones with colleges.

**Table 25. Destination choice models for the HBO, PM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1	.....(Fixed	Parameter).....	
<i>Impedance and Interaction Terms</i>				
Distance	-0.365	0.006	-61.72	<0.001
Distance Squared	0.00400	0.000	34.69	<0.001
<i>Size Terms</i>				
Ln(Population Density)	0.875	0.017	50.20	<0.001
Ln(Retail Jobs)	-0.00447	0.011	-0.40	0.686
Ln(Non-Retail Jobs)	0.408	0.016	26.31	<0.001
<i>Geographical Constants</i>				
Parks	0.0836	0.043	1.96	0.050
Colleges	-0.298	0.045	-6.59	<0.001
Commercial Airports	0.206	0.073	2.83	0.005
Military Bases	1.291	0.308	4.19	<0.001
N = 9645				
LL(0) = 28313.04				
LL( $\beta$ ) = -21143.55				
$\rho^2 = 0.253$				

<sup>a</sup> SE is the standard error

Table 26, showing the results for the Midday period trips, describes overall similar relationships as the previous models. The notable differences include the statistically significant interaction of the low-income household indicator with distance, indicating that lower income families tend to travel shorter distances to get to their HBO destinations.

**Table 26. Destination choice model for the HBO, Midday period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.3472	0.005	-76.990	<0.001
Distance Squared	0.00356	0.000	32.180	<0.001
Distance×Low-Income(<\$25K)	-0.0124	0.005	-2.500	0.013
<i>Size Terms</i>				
Ln(Population Density)	0.68916	0.017	41.430	<0.001
Ln(Retail Jobs)	0.0479	0.011	4.280	<0.001
Ln(Non-Retail Jobs)	0.500	0.016	31.860	<0.001
<i>Geographical Constants</i>				
Parks	0.0372	0.043	0.860	0.390
Colleges	-0.148	0.045	-3.290	0.001
Commercial Airports	0.140	0.076	1.850	0.064
Military Bases	1.792	0.282	6.360	<0.001
N = 8965				
LL(0) = 26294.88				
LL(β) = -20448.01				
$\rho^2 = 0.222$				

<sup>a</sup> SE is the standard error

Table 27 shows the results for the Night period trips. The results, in general, are in line with the previous models, and point out that a higher vehicle ownership is associated with choosing

farther destinations, and TAZs with commercial airports and military bases, all else equal, are more likely to be among chosen destinations.

**Table 27. Destination choice model for the HBO, Night period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.382	0.010	-40.070	<0.001
Distance Squared	0.00420	0.000	29.420	<0.001
Distance×VO	0.00788	0.003	2.510	0.012
<i>Size Terms</i>				
Ln(Population Density)	0.932	0.026	35.430	<0.001
Ln(Retail Jobs)	-0.0807	0.016	-4.960	<0.001
Ln(Non-Retail Jobs)	0.418	0.023	18.410	<0.001
<i>Geographical Constants</i>				
Parks	0.0287	0.061	0.470	0.640
Colleges	-0.268	0.064	-4.170	<0.001
Commercial Airports	0.212	0.112	1.880	0.060
Military Bases	1.819	0.392	4.640	<0.001
N = 4330				
LL(0) = 12749.99				
LL(β) = -9950.89				
ρ <sup>2</sup> = 0.220				

<sup>a</sup> SE is the standard error

### **Nonhome-based Trip Models**

The total number of NHB trips in our data is 18,610 cases. Similar to the previous two subsections, we will discuss the NHB destination choice with four segmented time-of-day models.

Table 28 and table 29 show the results for the NHB destination choice model during the morning and afternoon peak periods. The polynomial distance term in both models shows a negative association with chosen destination, and the negative sign of the interaction of low-income household and distance shows that lower income households are less likely to travel farther to get to their destination.

The size terms, in addition, show that the TAZs with a higher population density or employment count are more likely to be among the chosen destinations. With respect to the geographical constant terms, we see that military bases in both models, all else equal, are more likely to be among the chosen destinations, while the TAZs with parks are less likely to be chosen in the afternoon peak model.

**Table 28. Destination choice model for the NHB, AM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.338	0.007	-46.510	<0.001
Distance Squared	0.00395	0.000	24.520	<0.001
Distance×Low-income(<\$25K)	-0.0138	0.008	-1.790	0.073
<i>Size Terms</i>				
Ln(Population Density)	0.249	0.022	11.420	<0.001
Ln(Retail Jobs)	0.0358	0.019	1.890	0.059
Ln(Non-Retail Jobs)	0.957	0.026	37.000	<0.001
<i>Geographical Constants</i>				
Parks	0.0780	0.072	1.080	0.281
Colleges	0.0180	0.068	0.270	0.790
Commercial Airports	0.145	0.123	1.180	0.237
Military Bases	3.950	0.301	13.130	<0.001

N = 3010

LL(0) = 8822.44

LL( $\beta$ ) = -7066.33

$\rho^2 = 0.199$

<sup>a</sup> SE is the standard error

**Table 29. Destination choice model for the NHB, PM peak trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.324	0.005	-61.760	<0.001
Distance Squared	0.00362	0.000	29.670	<0.001
Distance×Low-income(<\$25K)	-0.0293	0.007	-4.150	<0.001
<i>Size Terms</i>				
Ln(Population Density)	0.412	0.018	22.900	<0.001
Ln(Retail Jobs)	0.237	0.014	16.630	<0.001
Ln(Non-Retail Jobs)	0.614	0.019	31.560	<0.001
<i>Geographical Constants</i>				
Parks	-0.111	0.053	-2.080	0.037
Colleges	-0.0264	0.050	-0.520	0.601
Commercial Airports	0.106	0.095	1.110	0.265
Military Bases	1.441	0.475	3.030	0.002

N = 5774

LL(0) = 16964.27

LL( $\beta$ ) = -13688.61

$\rho^2 = 0.193$

<sup>a</sup> SE is the standard error

Table 30 and table 31 show the results for the Midday and Night period trips, with the results being similar to the previous models.

**Table 30. Destination choice model for the NHB, Midday period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.394	0.00472	-83.70	<0.001
Distance Squared	0.00469	0.00011	42.27	<0.001
<i>Size Terms</i>				
Ln(Population Density)	0.316	0.0139	22.65	<0.001
Ln(Retail Jobs)	0.241	0.0120	20.01	<0.001
Ln(Non-Retail Jobs)	0.699	0.0161	43.28	<0.001
<i>Geographical Constants</i>				
Parks	-0.176	0.0458	-3.84	0.0001
Colleges	0.0225	0.0416	0.54	0.5891
Commercial Airports	0.0125	0.0768	0.16	0.8703
Military Bases	3.479	0.296	11.72	<0.001
N = 8537				
LL(0) = 25074.96				
LL( $\beta$ ) = -18291.41				
$\rho^2 = 0.270$				

<sup>a</sup> SE is the standard error

**Table 31. Destination choice model for the NHB, Night period trips.**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>a</sup></b>	<b>Z-value</b>	<b>P-value</b>
Sampling Correction Term	-1		Fixed Parameter	
<i>Impedance and Interaction Terms</i>				
Distance	-0.349	0.011	-31.780	0.000
Distance Squared	0.00425	0.000	16.860	0.000
<i>Size Terms</i>				
Ln(Population Density)	0.327	0.035	9.260	0.000
Ln(Retail Jobs)	0.185	0.030	6.120	0.000
Ln(Non-Retail Jobs)	0.746	0.041	18.320	0.000
<i>Geographical Constants</i>				
Parks	-0.0974	0.108	-0.900	0.367
Colleges	-0.0963	0.102	-0.950	0.344
Commercial Airports	-0.0860	0.226	-0.380	0.703
Military Bases	4.545	0.816	5.570	0.000
N = 1289				
LL(0) = 3786.54				
LL( $\beta$ ) = -3050.38				
$\rho^2 = 0.194$				

<sup>a</sup> SE is the standard error

## **SUGGESTIONS FOR MODEL IMPROVEMENT**

Although in this task we explored the viability of destination choice models and how they can be applied in the context of Georgia trip patterns, we saw that some models showed a lower goodness of fit than others. An important step in improving the accuracy of these models is to adopt a better impedance variable. In our modeling, we used distance as the main impedance variable, but travel time can help with model predictions if it is used appropriately. We recommend that in further iterations of these models, congested and uncongested travel times be used instead of distance as the main impedance variable according to the time of day of each



model. In addition, further and more detailed geocoding of geographical features of TAZs can better capture unique destination patterns among trips.

## **CHAPTER 6. DEVELOPMENT OF A MODE CHOICE MODEL**

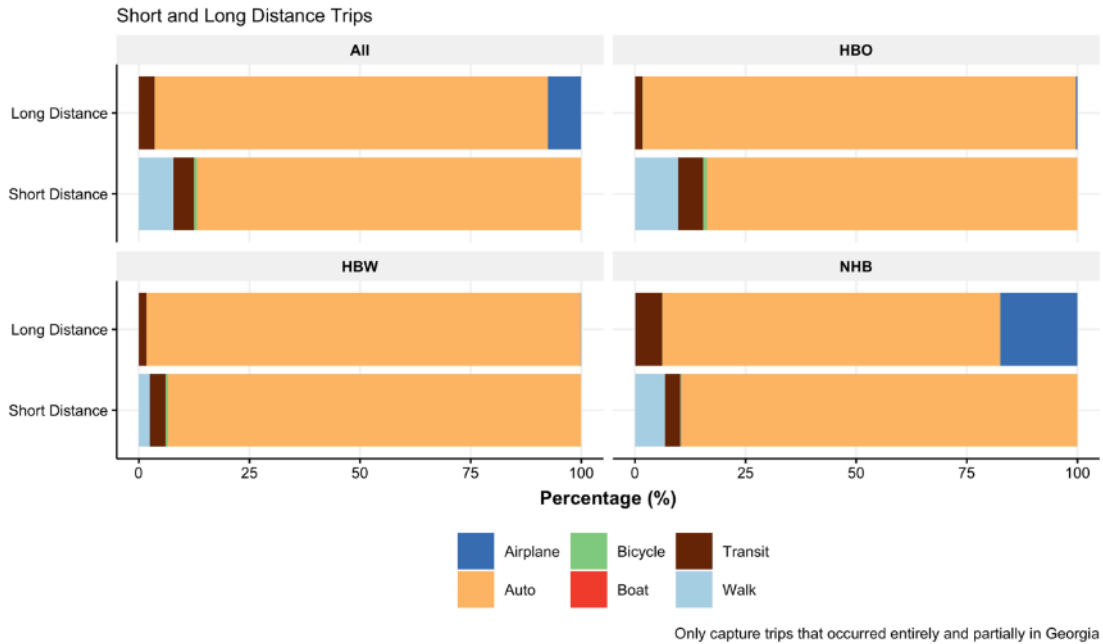
Understanding mode choice is a critical component in most travel demand modeling across a variety of geographical scale, and the GSTDM is not an exception. As part of this research project, the research team seeks to enhance the collective understanding of a mode choice component in the GSTDM. The current mode choice model in the GSTDM consists of two purpose-specific parts (business and nonbusiness models) and includes only mode attribute characteristics. Since the latest version of NHTS has been released and additional data sources are available, a more sophisticated modeling approach and corresponding choice model can be developed.

GDOT Report 18-24 (Kash, Mokhtarian, and Circella 2021) provides an initial exploration of mode use in the 2017 NHTS Georgia add-on. Specifically, they discuss the mode share of all the recorded travel modes in the 2017 NHTS Georgia add-on, and further compare mode shares across MPO tiers in Georgia. Readers are encouraged to refer to GDOT Report 18-24 for further discussion on the mode share overview in Georgia.

### **COMPOSITION OF THE MODE CHOICE MODEL**

The mode choice model is divided into short- and long-distance models based on the trip length with a threshold of 50 miles. That is, short-distance trips refer to trips that cover the distance for 50 miles or less, whereas long-distance trips cover the distance for more than 50 miles. Figure 34 presents the share of mode for travel by distance. As indicated, one primary differentiating characteristic between short- and long-distance trips is that, on one hand, short-distance trips had a nonnegligible share of nonmotorized trips (e.g., walking and bicycling). On the other hand, the air mode accounts for a significant portion of long-distance trips. It apparently shows why

distance-based segregation is needed: totally different mode choice composition and relevant heterogeneous behavioral characteristics.



**Figure 34. Stacked bar graphs. Share of mode choices for short- and long-distance trips by trip purpose.**

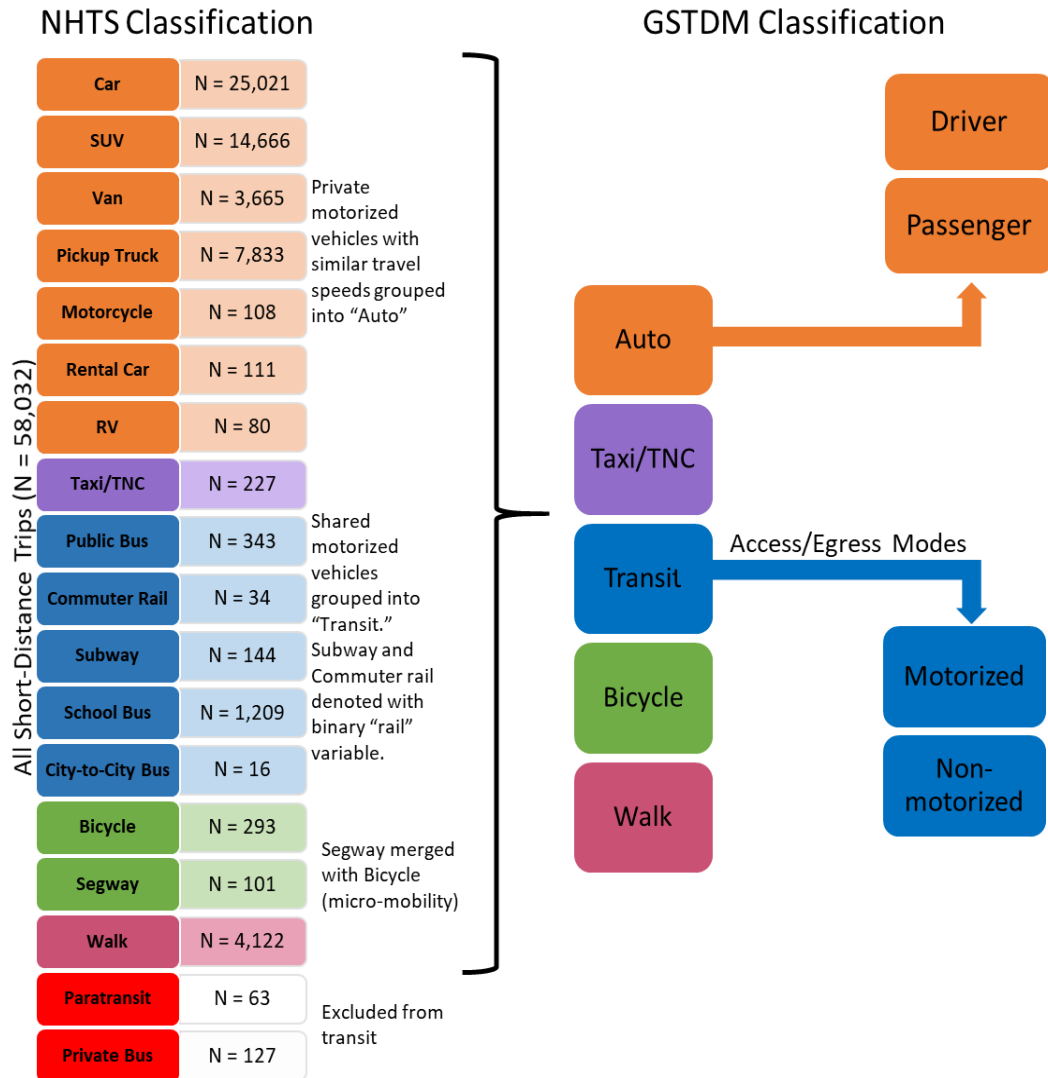
## SHORT-DISTANCE TRIPS

This section elaborates on the data assembly process for short-distance trips, descriptive statistics of select indicators in the dataset, the methodology for estimating the models, and the results as the basis for evaluating travelers' behavior in Georgia.

### Data Assembly Process

As mentioned above, the research dataset for task 4 is constructed by incorporating alternative, nonchosen modes for each trip observation. In assembling the data, the research team incorporates a framework that combines the existing primary data (i.e., the NHTS add-on for

Georgia) with emerging data sources derived primarily from Google API (application programming interface) needed to incorporate the nonchosen modes. In doing so, the research team has developed a script-based, automated process as applicable in the R Studio environment to obtain travel time information for these following modes, i.e., auto, transit, bike, and walk, using Google Maps Distance Matrix API. The decision to incorporate these modes stems from these factors: (1) the research team considers these modes as an appropriate representation of the available modes for travelers in Georgia, following the framework as presented in figure 35; (2) Google API provides a robust travel time estimation for these modes; and (3) leveraging the Google API, the choice of these modes enables a scalable application within and across geographies in Georgia.



**Figure 35. Classification chart. Linking mode classification between NHTS and GSTDM.**

***Travel Time Information: Google API Query***

The query process to obtain travel time requires the following sets of information: date, time, origin, and destination of the trips. This particular level of comprehensive information necessary to obtain travel time of the nonchosen modes for each trip highlights the benefits of adopting the add-on component of the NHTS data. The research team highly values the availability of the NHTS add-on data component for Georgia and recommends that GDOT consider continuously adopting the add-on component in the foreseeable future.

One particular aspect to note is that Google API would not allow a query for historical dates; that is, either at this moment or in the future (Google, 2021). To address this apparent shortcoming, the research team added 5 years from the actual date when the trip took place. Since most of the trips in the Georgia NHTS add-on took place between 2016 and 2017, the query process, therefore, functions as if the trips were in 2021 to 2022. The logic behind selecting this 5-year timeframe stems from the practicality and the current COVID-19 pandemic. From the practicality perspective, the project was started in mid-2019 and the team initiated the query process in Spring 2020. From the perspective of the current COVID-19 pandemic, running the query process using the year of 2020 might bias the estimation due to the notion that travel demand generally declined in Georgia and elsewhere in most parts of the country in 2020. In terms of time, we adopt the hours and minutes from the NHTS as it is. Thus, an example of the adopted timestamp for running a query would be instead of using, for instance, 2016-04-18 5:47 PM, we use the following: 2021-04-18 5:47 PM.

The query process has also benefited from the considerable level of granularity of the origin and destination for each trip as available in the NHTS add-on. While the data do not disclose or provide the exact XY coordinates of each origin and destination combination, the data provide census tract identifiers. The presence of census tract identifiers has allowed the research team to construct spatially embedded XY coordinates at the centroid of each census tract used to run the query process.

### ***Data Cleaning and Wrangling***

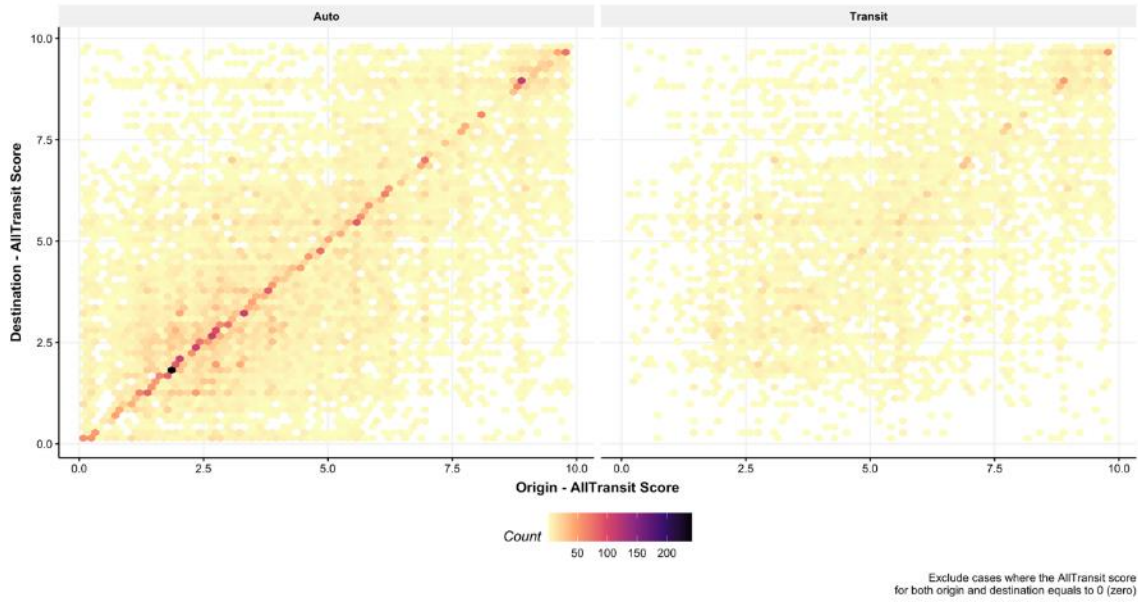
Having collected travel time information from Google API, the research team applies a suite of data cleaning and wrangling processes to ensure the applicability of the dataset to run the mode

choice analyses, particularly for short-distance trips that occurred entirely or partially in Georgia.

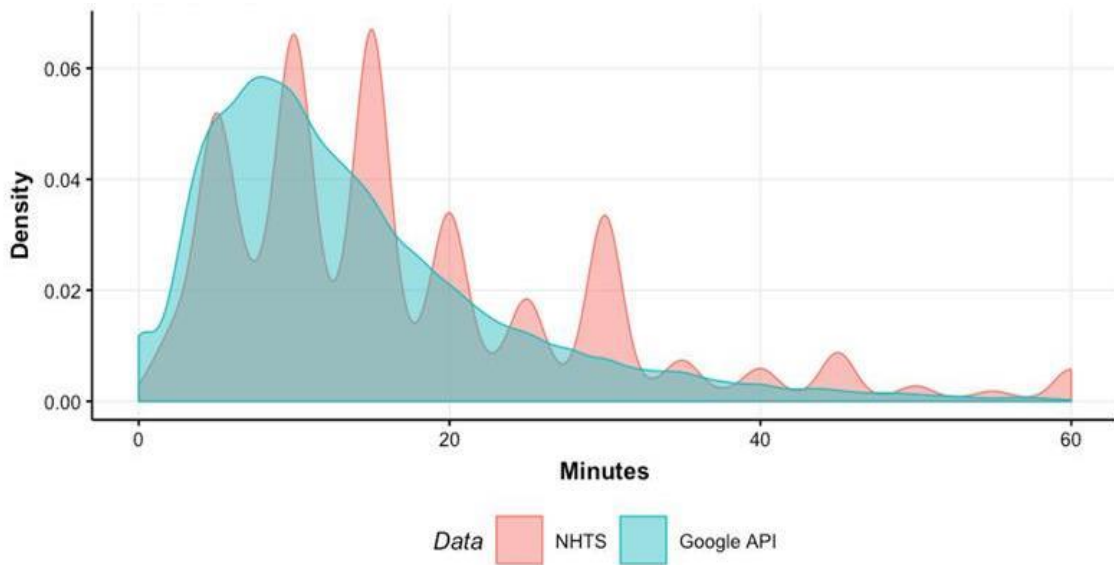
The research team applies several layers of the cleaning and wrangling process that involves:

- (1) evaluating the obtained travel time information derived from Google API, and
- (2) comparison between the NHTS and Google API.

In evaluating the obtained travel time information derived from Google API, the research team identifies one notable issue: the potentially imprecise travel time information from the NHTS add-on data. As indicated in figure 37, the distributional density of travel time on the NHTS add-on data tends to follow an arbitrary 5-minute interval of travel time. One potential explanation of this issue is due to the circumstance where the survey instrument relied largely on the respondents' memorization of the travel time on a particular mode. This subsequently could lead to the rounding of travel time information into a 5-minute interval, i.e., 5, 10, 15, 20 minutes, and so on as observed in the data. The research has, therefore, explored the data further to evaluate whether this observation holds true across the primary modes of travel incorporated in the short-distance trips dataset.



**Figure 36. Scatter plot. AllTransit score for the origin and destination combination by mode (auto and transit), short-distance trips. Larger values denote greater transit presence.**

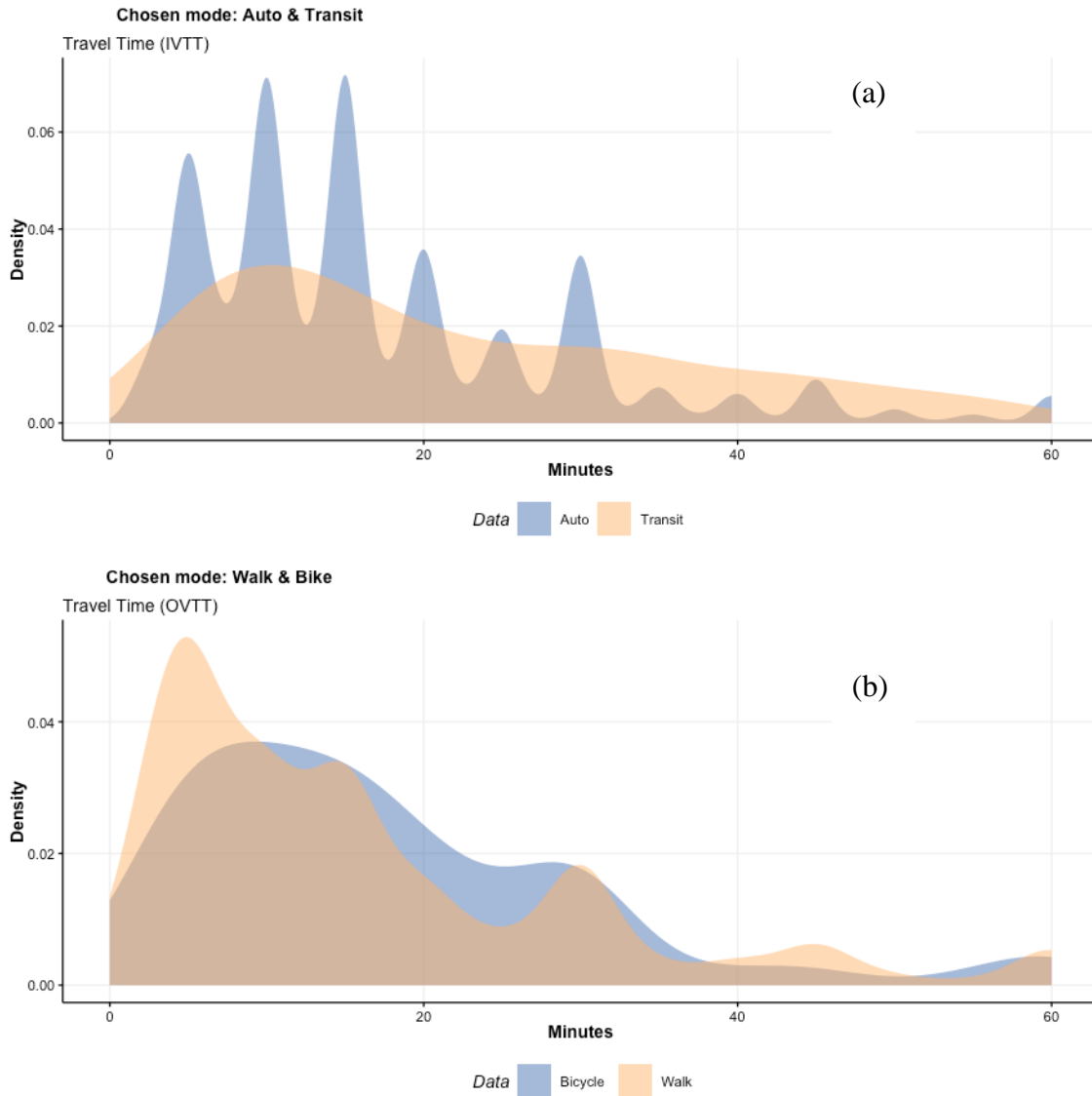


**Figure 37. Area graph. Density distribution of travel time between NHTS and Google API.**

This observation is particularly telling when observing the distributional density of travel time on auto as shown in figure 38(a). As a note, this issue is not unique to travel time on auto, but can



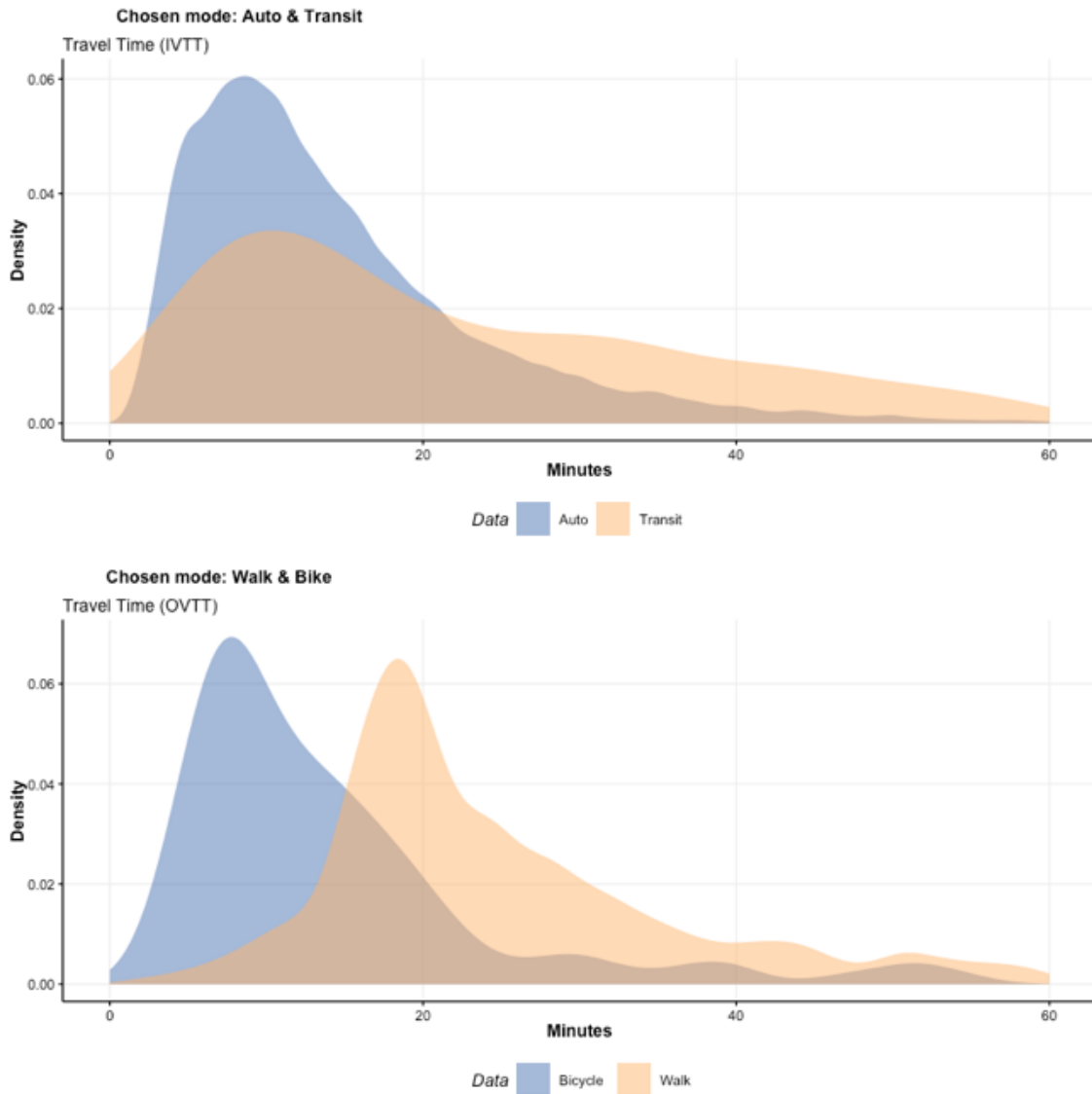
also be observed in other modes, e.g., walking and bicycling, as shown in figure 38(b), albeit with a reasonably less distinct characteristic than auto.



**Figure 38. Area graphs. Density distribution of travel time from NHTS add-on data, short-distance trips.**

Given the prevalence of rounded travel time into a 5-minute interval, the research team has sought to compare the NHTS add-on data with travel time information derived from Google API, as evidenced in the analysis as shown in figure 37. Figure 39 shows the distributional density of

travel time derived from Google API. Comparing the observation from figure 44 with the one from figure 38, it is apparent that the distributional density of travel time from Google API tends to follow a smoothed distribution, which might be more representative of the actual observation of the trips captured in the NHTS data.



**Figure 39. Area graphs. Density distribution of travel time derived from Google API, short-distance trips.**

Given that travel time derived from Google API yields a more realistic representation of the trips, the research team has decided to adopt the travel time from Google API into the short-distance trips dataset.

### ***Fuel Cost, Parking, and Transit Fare***

Another challenge arising from incorporating nonchosen travel modes for estimating mode choice models is calculating fuel cost and transit fare. The research team has incorporated data from various sources and applied several logics and assumptions to address this challenge.

*Fuel Cost.* In terms of fuel cost, the research team makes use of several existing indicators in the NHTS add-on dataset and data from a vehicle fuel efficiency database.<sup>6</sup> The research team links information of vehicle make and brand in the NHTS add-on data (i.e., the vehicle module at the household level) with the vehicle fuel efficiency database. To compute the estimated fuel cost for both chosen and nonchosen auto trips, the research team multiplied the estimated fuel consumption per mile on city road, instead of highway, with the distance covered for the corresponding trip considering the gasoline prices at that time. For instance, a traveler living in a household owning a Honda CR-V and using the car to cover a distance of 9.5 miles would spend an approximate \$0.57 in fuel cost given the estimated vehicle efficiency of a Honda CR-V at 36.41 miles per gallon (mpg) and gasoline prices at 219.5 cents per gallon.

*Parking.* In terms of parking cost, the research team follows several layers of logic. One particular consideration is the mode choice; that is, we assume travelers would not spend any out-of-pocket monetary parking cost if the travelers in question did not drive or ride as a

---

<sup>6</sup> The database contains information of vehicle fuel efficiency for thousands of cars by brand and make in the U.S. Due to the proprietary nature of the database, contact the research team to gain access to this database.

passenger in a car trip. Another consideration is the destination of the auto trips; that is, we set forth a scenario on the estimated parking cost based on the locational characteristics of the destination. In classifying the locational characteristics, and subsequently the estimated parking cost, the research team follows the area classification as adopted in the ‘URBANICITY’ variable in the NHTS add-on data. This variable, derived from Claritas (2020), captures the rural–urban continuum in the NHTS data that consists of rural, small town, suburban, second city, and urban. Using this variable, the research team has, therefore, assigned parking cost where parking cost is assumed as zero in the rural, small town, and suburban areas; the hourly parking rate for areas classified as second city and urban are accordingly assigned based on the research team’s tailored investigation.

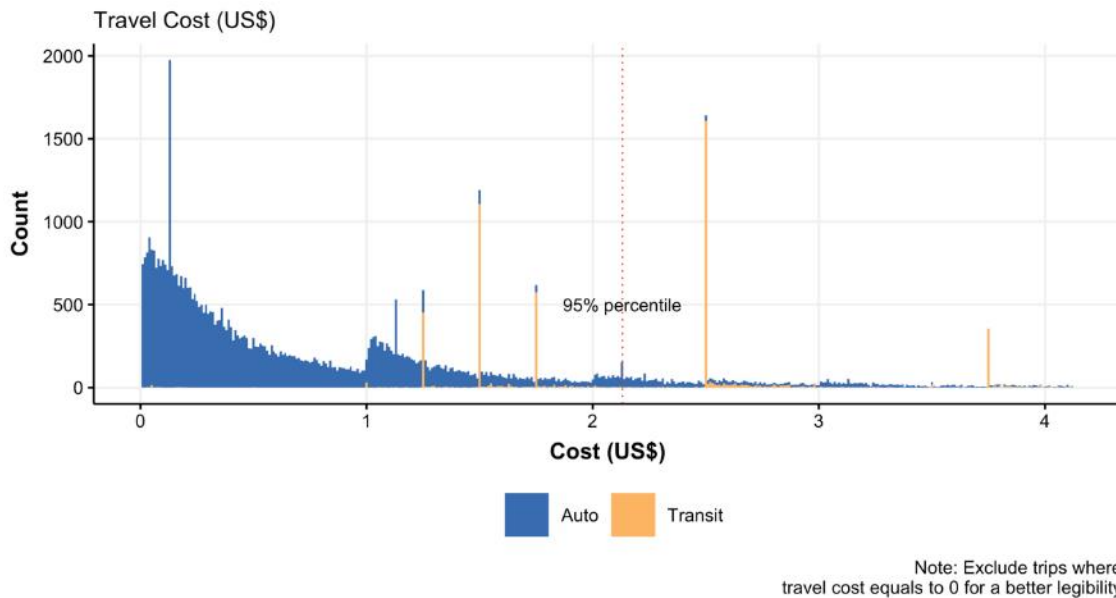
In estimating parking cost, the research team also considers the estimated parking duration. In probing the duration, the research team has computed the hourly interval before the auto traveler took the subsequent trips. Using the estimated parking duration, the research team multiply the duration with the hourly rate following area classification as mentioned above. Moreover, the research team also differentiates between parking over an extended period (e.g., daily parking) and short-term period. That is, if parking duration would last for more than 4 hours, the research team assigns \$19 daily rate in the corresponding areas where parking cost is levied.

*Transit Fare.* In determining the estimated transit fare, the research team has compiled transit fare for each transit agency in Georgia, as well as in the adjacent states where the trips crossed the state boundary (table 32). Google API has proven to be useful for this process as it could return the information of the transit agency for the corresponding trip.

**Table 32. Transit fare in Georgia.**

Transit Agency	Travel Cost (\$)
Atlanta Street Car	1
Cherokee Area Transit	1.25
Cobb Linc	2.5 / 5
Gwinnet County Transit	2.5 / 3.75 / 5
GRTA Xpress	3 / 4
Marta	2.5
Athens	1.75
Augusta	1.25
Savannah	1.5

*Combined Travel Cost.* Considering several aspects related to travel cost as described above, figure 40 depicts a comparison of frequency distribution of travel cost (USD) between auto and transit. As shown in the figure, on one hand, travel costs for auto tend to be distributed heavily on the left side, where most are not more than \$1. On the other hand, travel costs for transit tends to be distributed sparsely, and more or less follow the \$0.5 interval from one distribution to another.



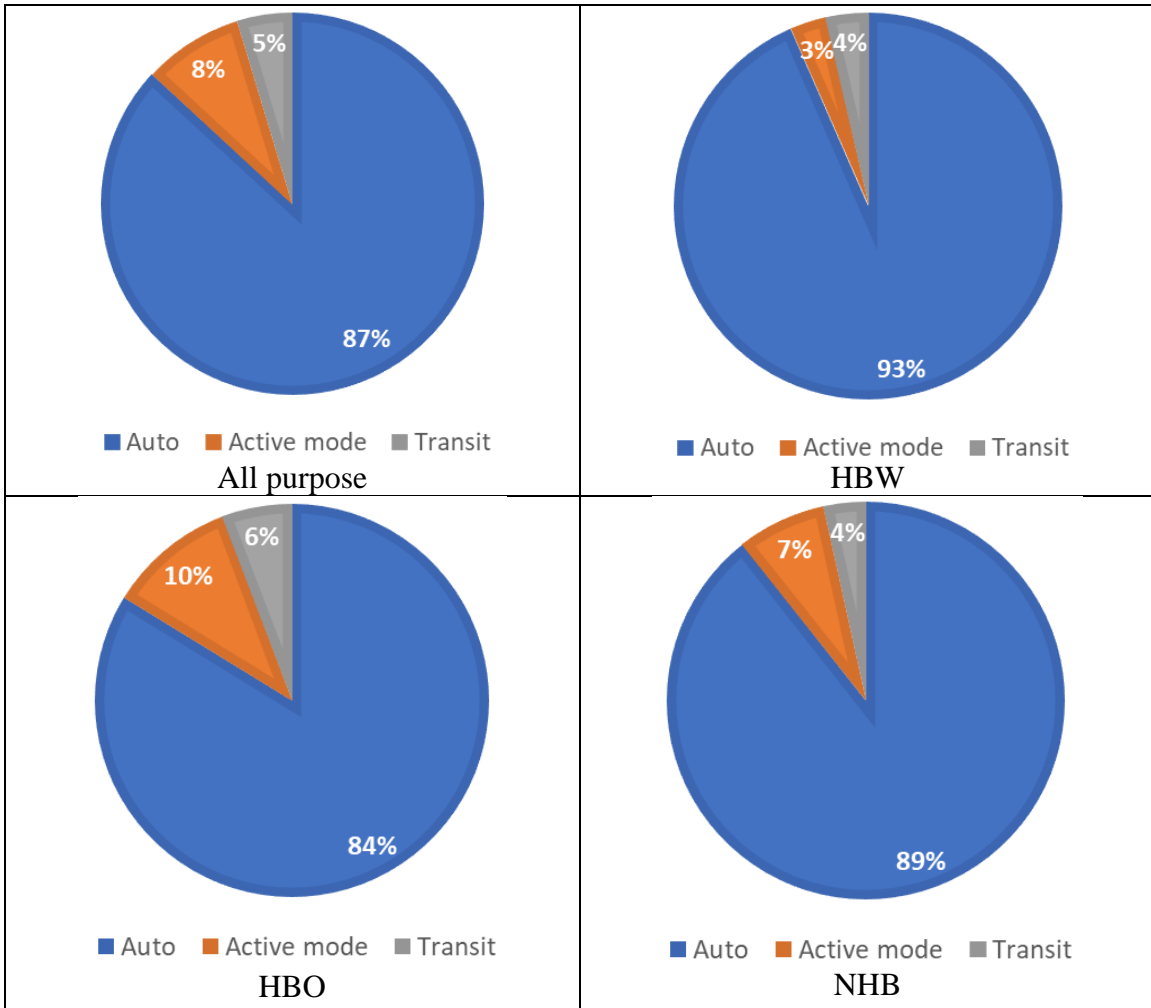
**Figure 40. Histogram. Estimated travel cost for auto and transit.**

### ***Access and Egress Time and Mode***

An additional consideration in assembling the data is developing considerably realistic access and egress information for both chosen and nonchosen transit trips. For instance, a traveler who owns a car is assumed to use the car for accessing the station; therefore, it might be unrealistic to assume that the traveler in question would walk covering lengthy distances to access transit stops. In addressing this aspect, the research team adopts this following framework. If access/egress distance is longer than 1 mile, it is assumed that travelers take transit or auto depending on mode availability for first or last mile travel (labeled as *public transit with a motorized mode*). Otherwise, all travelers are assumed to walk to reach a station or destination (labeled as *public transit with a non-motorized mode*).

### ***Mode Share Distribution for Short-distance Trips***

Figure 41 shows the mode share distribution for short-distance trips that occurred entirely and or partially in Georgia. The primary finding is as expected since a strong majority of travelers in Georgia use automobile as their primary mode of travel across different trip purpose.



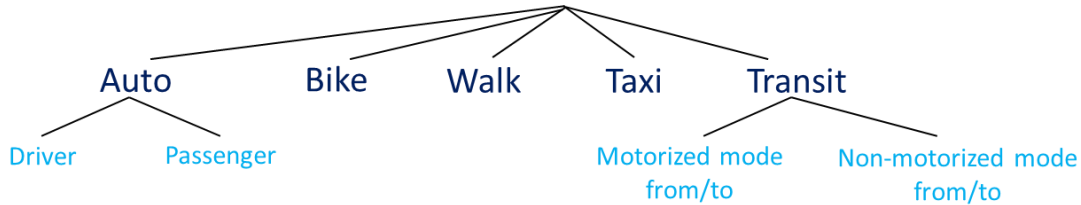
**Figure 41. Pie graphs. Share of mode choices for short-distance trips by trip purpose.**

## Model Specification

### *Mode Choice Set*

As presented in figure 35, the research team regroups the two-level mode choice set, including five modes at the upper level (i.e., auto, bike, walk, taxi, and transit). At the lower level, auto is divided into driver and passenger, and transit is also split into two specific modes based on the access/egress mode indicating whether those first- and last-mile modes are motorized or

nonmotorized. Thus, the final model choice set includes seven specific modes, as illustrated in figure 42.



**Figure 42. Model diagram. Mode choice set with nested structure.**

The research team also constructs an unequal choice set, considering what alternatives are physically available for given situations. Traditional mode choice models have a universal choice set that all travelers are assumed to have equal mode choice options regardless of their mode accessibility. This assumption is considerably unrealistic and not straightforward. Thus, for those who have different modal options, the mode choice set needs to be adjusted to reflect the circumstance. For example, a traveler who does not have access to transit at home cannot choose the transit mode at the beginning of the home-based trip. It was also motivated by some paradoxical cases in the NHTS dataset; there are travelers choosing car even though they did not own a car. In this case, the car mode needs to be excluded from the choice set, and other possible options are required to define the mode they used (e.g., passenger or rental car). To create the unequal choice set, every trip in the Georgia add-on sample are examined, then only the available mode options are added in the choice set, which is the called individual-specific choice set.



## **Explanatory Variables**

The final set of explanatory variables consists of three categories, including mode attributes, socioeconomic demographics (SED) traits, and accessibility.<sup>7</sup> There are multiple sources that were used to define the explanatory variables. As described above, mode attributes are mainly collected by combining 2017 NHTS trip information and Google API. The SED traits are from the 2017 NHTS (individual and household information). Accessibility indicators are defined using AllTransit data provided by the Center for Neighborhood Technology.

Table 33 demonstrates mode attribute variables selected in the final model. Travel time is divided into in-vehicle travel time (IVTT) and out-of-vehicle time (OVTT). All modes have IVTT, but only car and transit have OVTT. Regarding transit, three specific OVTT indicators, including terminal, waiting, and transfer times, are merged altogether to calculate total OVTT. Travel cost is a single variable combining fuel cost, parking fee (only for car), and transit fare (only for transit). In the case of bike mode, travel cost is not assigned based on the assumption that all cyclists own and use their bike to travel.

---

<sup>7</sup> In initial exploration with various model specifications, some built environment characteristics were also considered, but proved to not be significant at the 10 percent level. Therefore, they were excluded from the final model.

**Table 33. Mode attribute variables of the mode choice model.**

Variable	Mode	Description
In-vehicle Travel Time (IVTT) <sup>a</sup>	Car	Driving time
	Taxi	
	Transit <sup>g</sup>	On-board time
	Bike	Biking time
	Walk	Walking time
Out-of-vehicle Travel Time (OVTT)	Car <sup>b</sup>	Urban = 5 minutes
		Suburban = 3 minutes
	Rural = 1 minute	
Terminal time	Transit <sup>c,g</sup>	(1) Walking time from/to a station or
		(2) driving time from/to a station if trip length is longer than 1 mile
Waiting time, Transfer time	Transit <sup>g</sup>	High frequency = half of headway
		Low frequency = 5 minutes
Fuel <sup>d</sup>	Car	$\frac{\text{Trip distance (mile)}}{\text{Gas mileage (mpg)}} \times \text{gas price (\$/g)}$
Toll fee	Car	Zero (not considered)
Travel Cost (US dollars)	Car	Average daily parking fee for
		(1) those who are full-time workers and work in urban areas, or
	(2) those who are university (or graduate) students	
	Car	Urban = \$4/hr (\$24/day more than 6 hrs)
Second city = \$1		
	Other = \$0	
Fare (monthly pass) <sup>f</sup>	Transit <sup>g</sup>	Average daily transit fare (e.g., \$4/day) for those who are full-time workers or students
Fare (one-time)	Transit <sup>g</sup>	Assign transit fares depending on the transit line

<sup>a</sup> IVTT was obtained from Google API.

<sup>b</sup> Car depends on locations of origin and destination.

<sup>c</sup> Transit was obtained from Google API. In particular, the threshold of 1 mile is determined based on the Transit Capacity and Quality of Service Manual (TCQSM).

<sup>d</sup> Official fuel economy data including EVs and hybrid cars were used.

<sup>e</sup> Parking fee (monthly) was applied to specific trip purposes (commute to work or attend school).

<sup>f</sup> Full-time workers or students were assumed to purchase the monthly pass for all trips.

<sup>g</sup> Information about available transit lines from Google API and relevant transit fare information, which were manually collected from MPO agencies, are combined.

Table 34 shows selected SED. Those variables are mostly identified based on the Georgia add-on sample. It is classified into individual and household categories. In addition, two additional

variables (*active driver* and *vehicle sufficiency*) are created to explain whether individuals are able to drive and own the available personal automobiles in their household.<sup>8</sup>

Concerning accessibility, seven specific variables are initially selected to estimate the final model, indicating to what extent public transit is accessible to travelers (table 35) as introduced in chapter 2.

**Table 34. Socioeconomic variables of the mode choice model.**

<b>Category</b>	<b>Variable</b>	<b>Description</b>
Individual	Female	1 = Yes, 0 = No
	Age	Continuous
	Under 16 years	1 = Yes, 0 = No
	Education	Categorical
	Employment	1 = worker, 0 = nonworker
Household	Active driver	1 = Yes, 0 = No
	Household size	Continuous
	Vehicle ownership	Continuous
	Income	Categorical
	Number of drivers	Continuous
Attitudes	Vehicle sufficiency	1 = Yes, 0 = No
	Perceived health status	1 (excellent) to 5 (poor)

Note: All socioeconomic variables are from the 2017 NHTS dataset.

---

<sup>8</sup> If individuals do not have enough vehicles in their household, they are not able to choose auto even though they hold a driver's license and have the ability to operate it. It clearly represents vehicle accessibility.

**Table 35. Accessibility variables of the mode choice model.**

<b>Category</b>	<b>Variable</b>	<b>Description</b>
AllTransit Score	Performance score	1 (poor) to 10 (excellent)
	Walkable neighborhood	Continuous
Transit Accessibility	# of jobs (workers) accessible in 30 mins transit ride	Continuous
	Transit connectivity index	0 (poor) to 35 (excellent)
	Transit trip per week	Continuous
	# of transit stops within ½ mile	Continuous
	# of high frequency transit routes within ½ mile	Continuous

Note: Accessibility variables were defined using AllTransit data provided by the Center for Neighborhood. Technology.

## **Estimation Results**

### ***Mode Choice Model for All Purposes***

The final estimation result is presented in table 36. The research team only presents the all-purpose model, not purpose-specific mode choice models such as HBW, NHB models. Initial exploration includes various model specifications with multinomial logit forms, but those results did not provide an acceptable goodness of fit, or some key variables (e.g., travel time or cost) were not significant. In addition, the ultimate goal of this research is to propose an improved modeling approach for the statewide model in Georgia (not a microscopic behavior model); thus, the research team has concluded that purpose-specific models are not necessary, and the all-purpose model can be used for future prediction and analysis in the context of the GSTDM.

The final model exhibits a decent model fit (adjusted rho squared = 0.377). All explanatory variables other than the alternative specific constant for taxi are statistically significant at the 5 percent level, with the expected signs. The model clearly shows that: (1) travelers are more likely to prefer a personal mode to public transit or active modes; (2) mode attributes including travel time and cost have negative impact on mode choice which is plausible and aligns with

extensive literature on mode choice behavior; (3) driving availability (i.e., *active driver* and *vehicle sufficiency*) is a critical factor in choosing car, implying that building an unequal choice set to take individual mode availability into account is imperative to describe mode choice behavior at a deeper level; and (4) transit accessibility is also a key variable to define mode availability and the corresponding model mode choice mechanism by accounting for whether individuals are able to access transit.

**Table 36. Mode choice model for all purposes (nested logit form).**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>c</sup></b>	<b>Z-value</b>	<b>P-value</b>
<i>Alternative Specific Constant</i>				
Drive	1.035	0.075	13.885	<0.001
Passenger	0.415	0.093	4.477	<0.001
Bike	-1.271	0.160	-7.935	<0.001
Walk <sup>a</sup>	-	-	-	-
Taxi	-0.056	0.612	-0.091	0.464
Public transit (motorized AE)	-1.397	0.170	-8.211	<0.001
Public transit (nonmotorized AE)	-1.862	0.109	-17.074	<0.001
<i>Mode Attributes</i>				
IVTT_auto	-0.147	0.002	-66.354	<0.001
IVTT_passenger	-0.143	0.002	-61.025	<0.001
IVTT_taxi	-0.102	0.009	-10.811	<0.001
IVTT_PT	-0.014	0.002	-7.303	<0.001
OVTT_passenger	-0.064	0.011	-6.023	<0.001
OVTT_taxi <sup>b</sup>	-0.172	0.036	-4.786	<0.001
Total time_bike	-0.164	0.007	-23.305	<0.001
Total time_walk	-0.057	0.001	-40.562	<0.001
Cost	-0.108	0.010	-10.974	<0.001
<i>SED</i>				
Female	-0.188	0.073	-2.568	0.005
Age	-0.011	0.002	-5.555	0.000
Worker	0.349	0.093	3.755	0.000
Active driver	0.121	0.028	4.33	0.000
Under 16 years	1.807	0.089	20.235	0.000
Vehicle sufficiency	1.229	0.053	23.303	0.000
<i>Accessibility</i>				
Average performance score	0.074	0.015	5.101	0.000
<i>Nesting Parameters</i>				
Auto	0.992	0.033	29.784	0.000
Public transit	0.285	0.049	5.786	0.000

N = 53,814

LL(c) = -39,670.34

LL(β) = -24,715.74

Adjusted  $\rho^2$  (MS) = 0.377

<sup>a</sup> Walk is the reference mode.

<sup>b</sup> Coefficients of OVTT for public transit with motorized and nonmotorized access/egress modes are equal to IVTT\_auto and IVTT\_walk.

<sup>c</sup> SE is the standard error.

In the meantime, the research team identifies a fundamental limitation on the practical application of the proposed mode choice model. Value of travel time savings (VOTTS) is generally utilized to evaluate mode choice model from the economic perspective by using travel time and cost coefficients. In the present research, VOTTS for auto is considerably higher than expected. It is generally known that the range of VOTTS for auto ranges somewhat varies, but most previous studies and guidelines suggest \$10 to \$40 per hour based on socioeconomic conditions of each country. In particular, The United States Department of Transportation (USDOT 2016) proposes that VOTTS for local travel are \$13.6/hour, \$25.4/hour, and \$14.1/hour for personal, business, and all purposes, respectively. It also suggests that VOTTS for intercity travel are \$19.00/hour (personal), \$25.40/hour (business), and \$20.40 (all purpose). However, VOTTS for auto in the proposed model are \$80/hr on average (VOTTS<sub>drive</sub> = \$82/hr, VOTTS<sub>passenger</sub> = \$79/hr) which is substantially higher than the U.S. guidance. On the other hand, VOTTS for other modes show relatively reasonable ranges; VOTTS for taxi and public transit are \$57/hr and \$8/hr, respectively.

After discussion, the research team determined that extremely short-distance auto trips may result in the biased estimation. In most cases, those auto trips are likely to be a part of a tour, meaning that they are dependent on other major, primary auto trips, so people accordingly use car regardless of travel impedance (time and cost). In this case, travel time and cost do not affect mode choice decision, and even may lead to higher VOTTS because those factors are not critical in choosing the car mode, and people would choose car even though travel time and cost for auto is higher than other modes (i.e., the impact of travel time/cost can be biased). On the other hand, trips made by taxi and public transit are relatively independent, meaning that they are less

affected by tour characteristics and other modes/trip purpose within the same tour, resulting in more plausible estimations of time value.

To examine variations of VOTTS and see how those short auto trips affect VOTTS, additional mode choice models stratified by trip purpose are developed after excluding such short auto trips. Three different thresholds are applied: 0.5, 1, and 2 miles, and VOTTS for IVTT and OVTT are calculated across scenarios (table 37). However, the result shows that there is no significant improvement on VOTTS; VOTTS for auto and taxi are still higher than the generally accepted range.

**Table 37. Comparison of VOTTS by model.**

Model	Driving distance	Purpose	VOTTS (\$/hr)						$\rho^2$ (EL)	$\rho^2$ (MS)	$\theta_5$ (<1)
			IVTT (auto)	IVTT (taxi)	IVTT (PT)	OVTT (taxi)	OVTT (PT_w)	OVTT (PT_a)			
1	All cases	All purpose	82	57	8	96	32	82	0.718	0.377	Yes
2	> 1 mile	All purpose	84	66	10	11	39	84	0.719	0.374	Yes
3	> 1 mile	HBW	62	46	13	56	22	62	0.850	0.320	Yes
4	> 1 mile	Non-HBW	84	66	8	133	38	84	0.699	0.358	Yes
5	> 0.5 miles	All purpose	109	82	12	136	44	109	0.718	0.398	Yes
6	> 0.5 miles	HBW	-	-	-	-	-	-	-	-	-
7	> 0.5 miles	Non-HBW	91	67	9	134	37	91	0.700	0.378	Yes
8	> 2 miles	All purpose	76	70	11	114	42	76	0.724	0.434	Yes

## LONG-DISTANCE TRIPS

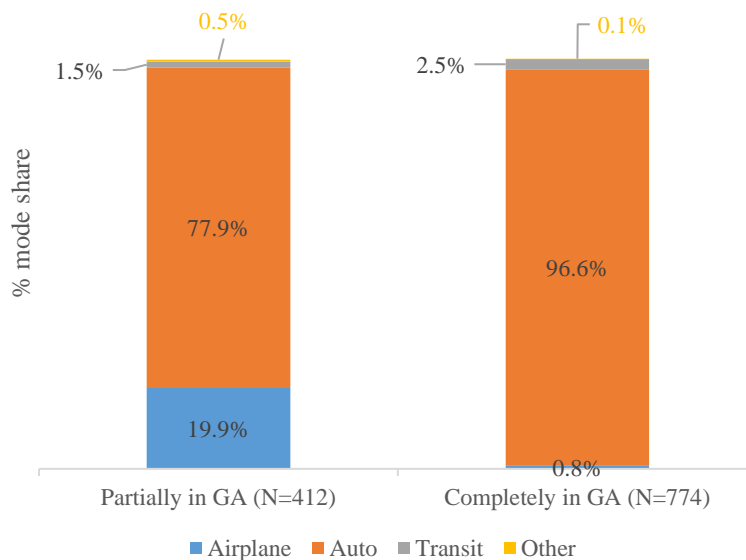
In contrast to the short-distance trip dataset in which we had a sufficient number of trips in the 2017 NHTS Georgia add-on to estimate a mode choice model, the total number of long-distance trips was too small and unimodal to allow for a robust model estimation. The research team tried to gather more long-distance data from other similar states, but given the timeline of the project, could not gather enough long-distance data from the other identified states. In this section,



therefore, we provide an exploratory analysis of the long-distance trips in Georgia, and provide recommendations on how GDOT can proceed with data collection and model estimation in the future.

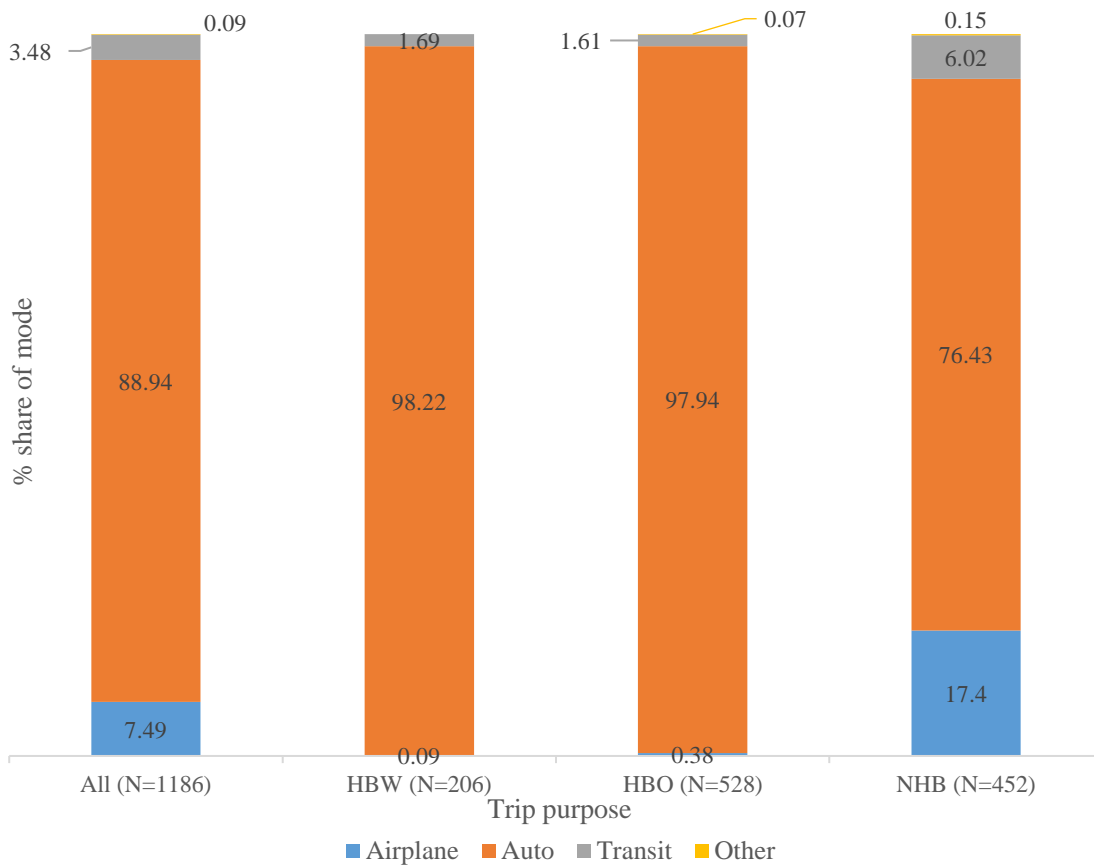
### Exploratory Analysis of 2017 Georgia NHTS Long-distance Trips

There are a total of 1186 long-distance trips in our dataset, of which, 774 (65.3 percent) took place completely and 412 (34.7 percent) took place partially in Georgia. Figure 43 shows the mode share distribution of long-distance trips based on whether they happened completely or partially in Georgia. As this figure shows, most of the long-distance trips that happened completely in Georgia were accomplished with an automobile (96.6 percent), with very few cases completed using transit or airplane. For the trips taking place partially in Georgia, however, airplanes show a 20 percent mode share, a significant difference compared to the mode share of trips completely within Georgia.



**Figure 43 Stacked bar graph. Mode share distribution of long-distance trips based on location in Georgia.**

Figure 44 shows the distribution of all long-distance trips by mode and trip purpose. Automobile is the dominant mode for long-distance travel for all trip purposes, especially for HBW and HBO purposes. Air travel and transit, moreover, constitute larger shares of mode choices for NHB travel than the other purposes, although automobile is still the dominant mode of travel for this trip purpose. Other modes of travel, including bicycle and walking, constitute negligible shares in all long-distance trip purposes.



**Figure 44. Stacked bar graph. Mode share for all long-distance trips by trip purpose in Georgia.**

Table 38, moreover, shows the distribution of (all of Georgia’s) long-distance trip distances by mode of travel. As this table shows, trips completed on an airplane have the highest median, while transit trips have the lowest median of all modes.

**Table 38. Distribution of long-distance trip distances by mode of travel in Georgia.**

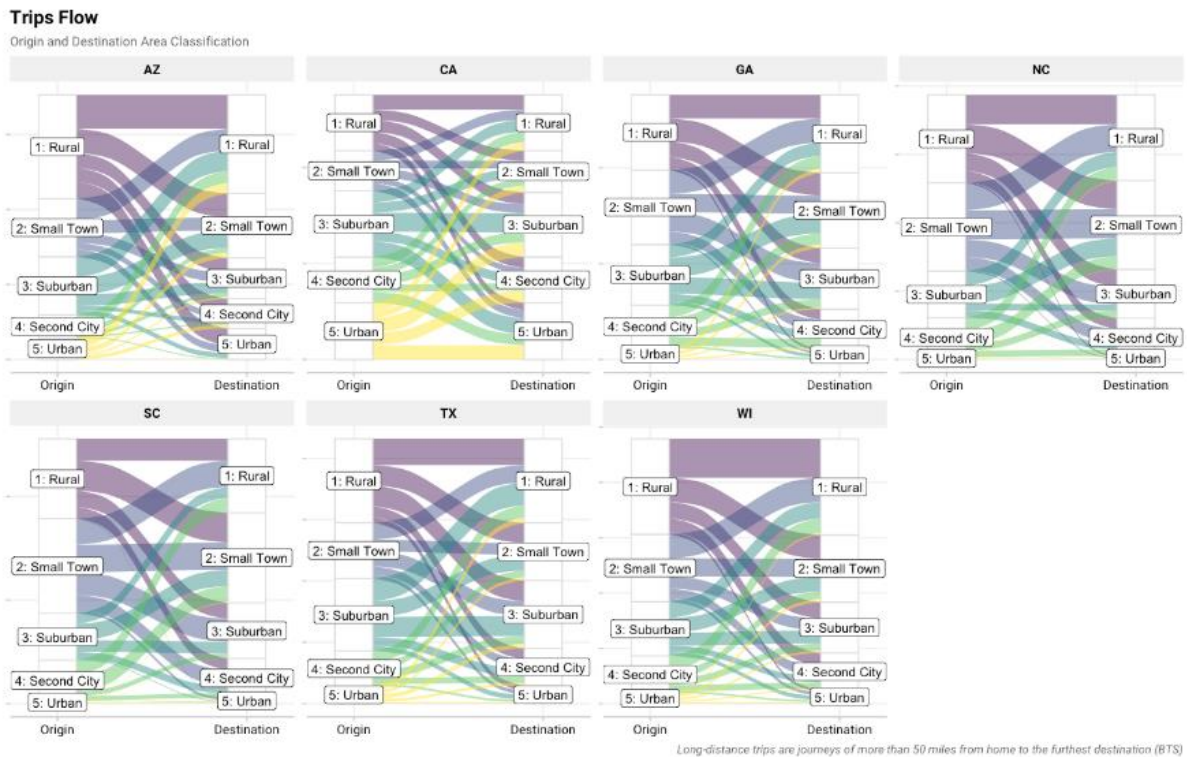
<b>Travel Mode</b>	<b>Descriptive Statistics of Trip Distances (miles)</b>			
	<b>Min.</b>	<b>Median</b>	<b>Mean</b>	<b>Max.</b>
All	50.02	93.73	172.47	5035.61
Automobile	50.02	87.72	127.41	866.73
Airplane	83.14	605.31	959.18	5035.61
Transit	50.98	78.29	193.61	822.36
Other	75	320.1	320.1	565.1

### **Augmenting Data with Comparable States**

Given the relatively few cases of long-distance trips in the NHTS Georgia add-on data, an additional consideration for assembling long-distance trips dataset is by including long-distance trips data derived from other states that also have the add-on component. This approach might be somewhat challenging since whether states participate in the add-on data collection could be driven by unobserved factors. Notwithstanding, several criteria can be applied to identify comparable states that have add-on data and use it to augment the long-distance trips dataset.

One of the proposed approaches is by identifying the trips flow characteristics as illustrated in figure 45. In particular, this approach could shed light on the directionality of the trips based on the geographic characteristics of the origin and destination. As indicated, the trips flow

characteristics of Georgia mimic those of North Carolina, South Carolina, and to some extent Arizona, where the ratio of the trip origin and destination geographic characteristics combination appear to be somewhat similar. On the contrary, most long-distance trips in California originated and ended in urban areas; thus, incorporating long-distance trips from the California data to the augmented long-distance trips might not be an appropriate decision. A similar observation can be made for the Wisconsin add-on data where a substantial share of long-distance trips occurred between rural areas.



**Figure 45. Diagrams. Trip flows by rural–urban continuum area classification for select states that have an NHTS add-on data component.**

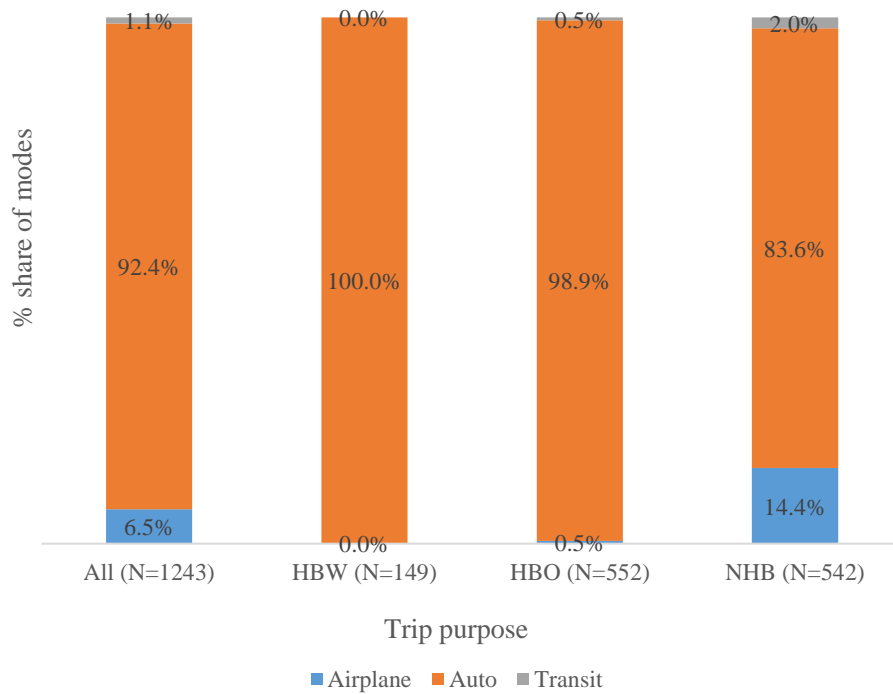
In addition to this data-driven approach, another method that can be considered is through a qualitative assessment of a given state’s urban spatial structure. This entails an assessment of whether the state has a dominant metropolitan area that dwarfs the rest of the metro or urban

areas within that state. Case in point would be Georgia itself that hosts the Atlanta metro, while the second largest metro, i.e., Augusta, pales in comparison to Atlanta.

The research team contacted the identified states' DOTs, but given timeline of this project could only obtain data from the South Carolina (SC) DOT. We recommend that future efforts continue the collection of datasets from the identified similar states, while in the following section, we provide an initial exploratory analysis of the SC NHTS long-distance trip data, and compare it to that of Georgia.

### *Exploring 2017 South Carolina NHTS Long-distance Trips*

There are 1243 cases of long-distance trips in the 2017 NHTS SC add-on data, the mode share for which is shown in figure 46. Similar to Georgia, automobile trips in South Carolina constitute a strong majority of all long-distance trips. The share of air travel, in addition, at 6.5 percent, is approximately equal to that of Georgia. In addition, air travel constitutes a more significant portion of long-distance NHB trips than other purposes, an observation also matched by the state of Georgia.



**Figure 46. Stacked bar graph. Mode share for all long-distance trips by trip purpose in SC.**

Table 39 also investigates the distribution of the long-distance trip distances in the SC dataset. This distribution is overall quite similar to Georgia's, with median and mean of all trips approximately corresponding to those of Georgia. The mean and median of automobile and air travels are also similar to Georgia's, but transit trips in South Carolina appear to be on average longer. We should caution that the number of transit long-distance trips in both datasets are very small, so making any conclusions based on these small numbers may be difficult.

**Table 39. Long-distance trip distance distribution by mode of travel in SC.**

<b>Travel Mode</b>	<b>Descriptive Statistics of Trip Distances (miles)</b>			
	<b>Min.</b>	<b>Median</b>	<b>Mean</b>	<b>Max.</b>
All	50.01	97.53	187.63	6555.40
Automobile	50.01	89.88	125.97	2588.85
Airplane	88.28	611.71	1067.28	6555.40
Transit	50.90	119.88	157.08	458.80

Overall, this exploratory analysis of South Carolina’s long-distance trip data confirms its similarities to Georgia’s, and that future analysis can augment Georgia’s dataset with South Carolina’s to estimate a more robust long-distance mode choice model.

## **CHAPTER 7. ADDITIONAL IMPROVEMENT TO THE GSTDM: PROPOSING A TOUR-BASED APPROACH**

In the traditional trip-based framework, each trip is treated as independent, meaning that there is no interaction between trips. In a broad sense, however, some trips are closely connected to each other, in particular when they belong to the same trip chain (a series of trips made by people sequentially). Krizek (2003) also pointed out that there are two specific problems with the traditional trip-based model; it treats each trip in an isolated manner, and it does not account for travel combining multiple purposes. The following is a simple case showing why the traditional trip-based model is not appropriate in explaining the changed trips. If a commuter takes public transit to come to the office, that person is not allowed to drive a personal automobile within the same tour because “driving car” is physically unavailable (although the car is parked at the workplace). Thus, the car mode must be excluded from the choice set in this case. However, it is not taken into account within the frame of the trip-based model. As mentioned above, each trip is treated as independent, so a universal choice set is usually given to all travelers without considering mode availability.

This limitation motivated the research team to consider and develop a tour-based mode choice model. To confine mode choice sets to specific modes available to a trip, the overall decision-making process of a tour trip needs to be clarified, and a corresponding modeling approach is required to clearly estimate the chosen mode for each trip within a tour. In this respect, the research team: (1) defines tours and trips within the frame of the tour-based model, (2) collects and manipulates the 2017 NHTS Georgia add-on sample, (3) estimates the tour-based model as an alternative approach for the mode choice modeling in the GSTDM, and (4) discusses its limitations in practical application.



## **DEFINITION OF TOURS**

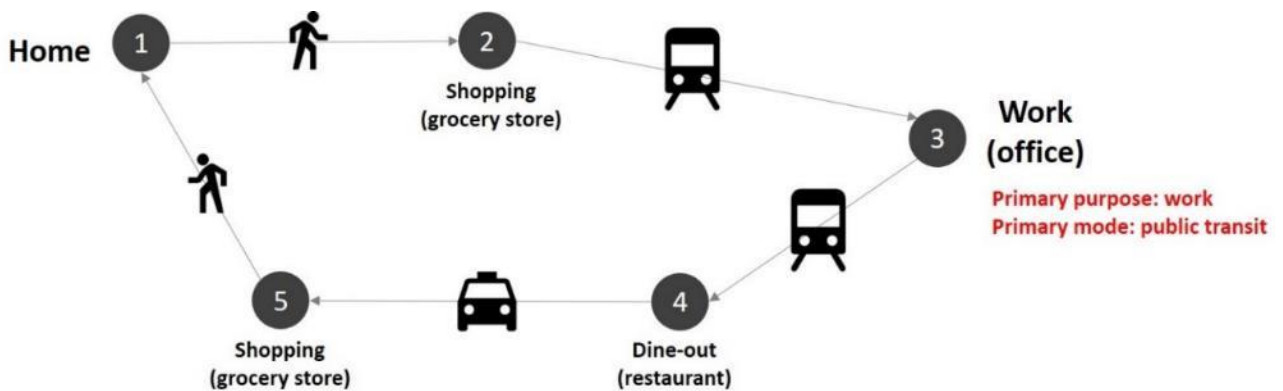
McGuckin and Nakamoto (2004) stated that a tour is the “total travel between two anchor destinations, such as home and work, including both direct trips and chained trips with intervening stops. Note that it is possible to have the two anchor destinations be the same location, as in a home-to-home or work-to-work tour.” A tour is generally defined as a home-to-home loop (i.e., home-based tour), and then it is broken down into each trip between two stops (Krizek 2003). The research team, accordingly, defined a tour in the same way. A tour denotes combined trips with multiple stops that belong to a same home-to-home loop, meaning that the first origin and the last destination are both home, and the remaining origins and destinations are non-home-based.

In order to define a tour, the primary purpose for the tour and the corresponding mode first need to be identified. The primary purpose reflects the most important decision that the traveler made during the trip. The research team hypothesized that travelers first choose a primary purpose and mode, then the remaining trips within the same tour are made conditional on the primary purpose and mode. The latter set of trips is referred to as secondary trips. That is, each tour consists of a single primary trip and multiple secondary trips.

A quick example of a tour is presented in figure 47. The following case is a home-based tour, comprising five dependent trips: (1) shopping from home to the store, (2) commute from the store to the office, (3) dining out from the office to the restaurant, (4) shopping from the restaurant to the store, (5) and returning home from the store. In terms of the traditional trip-based approach, those five trips are treated as independent, but that is not appropriate in the context of the tour-based model. When estimating a mode choice model based on that tour, the

car mode is not expected to be considered as an available alternative except for the first trip leg because the traveler does not drive from home, meaning that the driving option is not physically available. Thus, the second, third, fourth, and fifth trips must not include “driving a car” as an alternative in the choice set. It shows that some (or even all) trips are conditional on a specific mode choice.

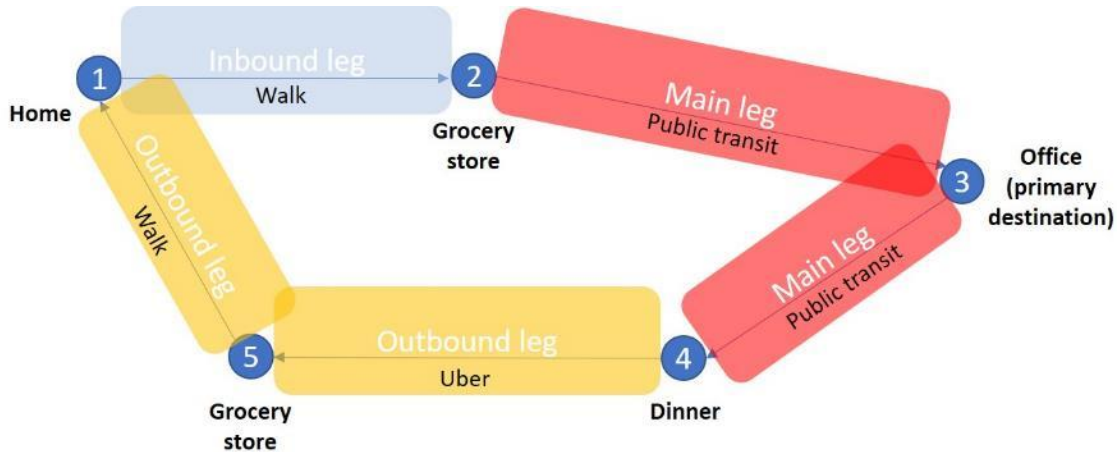
In the case above, the research team conjectured that commute to the office (work purpose trip) is the primary trip, and the rest of the trips are secondary trips. This is based on the assumption that the traveler first decides to use public transit to commute, then selects one of the available modes for the rest of the trips. Therefore, the primary purpose and mode are work and public transit, respectively. Of course, there is a possibility that the mode for dining out can be the primary mode if the traveler plans to drink, so he/she has decided not to drive the car. However, the research team did not consider those exceptional cases in a broad sense to estimate a conventional model in the context of the statewide level.



**Figure 47. Diagram. Example of a home-based tour.**

Another way to characterize tours and their individual tour legs is embedded in the Atlanta Regional Commission’s model. The tour composition presented in figure 47 may also be

illustrated as shown in figure 48. Two main legs can be specified when their origins or destinations are directly connected to the primary purpose. Then, all trip legs before the main leg are defined as inbound legs, and the remaining tour legs except for the main leg can be labeled as outbound legs.



**Figure 48. Diagram. Composition of a home-based tour.**

## DATA COMPOSITION

The research team used the same dataset as in chapter 5 (i.e., the 2017 NHTS dataset). In order to define tour trips, a set of appropriate variables, which are included in the original dataset and denote tour-related characteristics, were used. Therefore, all independent trips are successfully grouped as tour trips. The following subsections explain how the whole process was conducted.

### Tour-related Variables

There are several variables reflecting tour characteristics in our dataset. One of the most useful variables was the “hometrip” variable, which was generated and shared with us by another GDOT research team (Kash, Mokhtarian, and Circella 2021). The hometrip variable is a categorical variable with four types of trip characteristics identifying each trip as: (1) home-

based start, (2) home-based end, (3) home-based loop, or (4) non-home-based trip (refer to figure 49). By sorting all trips by departure and arrival times, and then grouping trips by the hometrip variable by trip case (*txcaseid* in figure 49), each tour was defined with a unique tour ID and corresponding details of tour, including trip number, number of trip legs, and the presence of home-based work/shopping trip (figure 50).

txcaseid	id	trippurp	choice_code	tdcaseid	personid	hometrip
X300004490101	1	HBW	1	300004490101	3000044901	1. Start Home
X300004490102	2	NHB	4	300004490102	3000044901	0. Not Home-based
X300004490103	3	NHB	4	300004490103	3000044901	0. Not Home-based
X300004490104	4	NHB	1	300004490104	3000044901	0. Not Home-based
X300004490105	5	NHB	1	300004490105	3000044901	0. Not Home-based
X300004490106	6	NHB	1	300004490106	3000044901	0. Not Home-based
X300004490107	7	NHB	1	300004490107	3000044901	0. Not Home-based
X300004490108	8	HBO	1	300004490108	3000044901	2. End Home

**Figure 49. Screenshot. *Hometrip* variable in the dataset (example from R Studio).**

tour_id	trip_number	tour_number	HBSHOP	HBW	tour_legs_count
TX300004490101	1	1	0	1	8
TX300004490101	2	2	0	0	8
TX300004490101	3	3	0	0	8
TX300004490101	4	4	0	0	8
TX300004490101	5	5	0	0	8
TX300004490101	6	6	0	0	8
TX300004490101	7	7	0	0	8

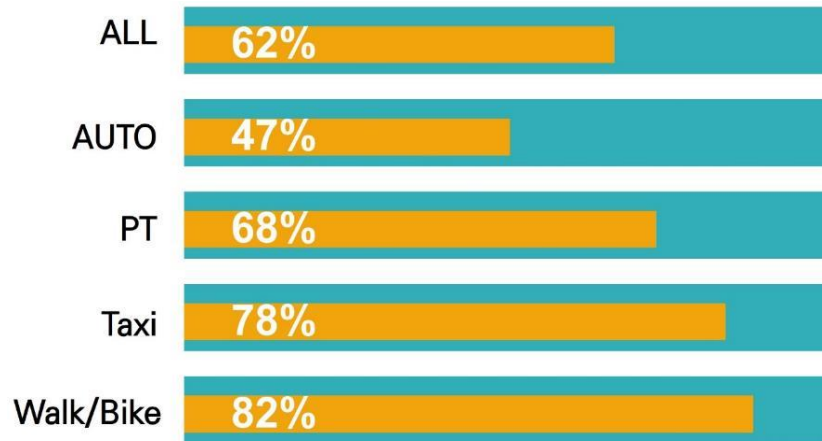
**Figure 50. Screenshot. Data format of the tour-based model (example from R Studio).**

## **Logic of Defining a Primary Trip**

After grouping all independent trips into tour-based trips, several rules are applied to each tour to determine the primary trip. First, the research team assumes that if a tour includes a HBW trip (especially, home to work), that trip is defined as the primary trip, indicating that the primary purpose is HBW, and primary mode is what the traveler used for the HBW trip. If a tour does not include HBW, a secondary trip with the longest travel time is defined as the primary trip.

## **PROFILES OF TOUR-BASED TRIPS**

Data exploration demonstrates that 62 percent of tour trips include only one trip leg (figure 51), meaning that they consist of one primary tour and a return trip. That is, 38 percent of tour trips have multiple trip legs. It tells us that two thirds of tours are quick trips, but there is a nonnegligible portion of complex travel, including multiple stops and corresponding available modes, which strongly support the necessity of the implementation of a tour-based model. The research team further looks into the share of tour trips with a single trip leg by primary mode for each tour. As expected, auto has a relatively lower share of a single leg (47 percent) compared to other primary modes. One possible explanation is that most travelers tend to prefer auto when they plan multiple trips within a tour because auto provides better accessibility and mobility. On the other hand, shares of a single leg for taxi and walk/bike are 78 and 82 percent, respectively, which is a plausible result. It shows trips made by taxi and walk/bike are mostly independent and sole tours with primary and return trips.



**Figure 51. Bar graph. Shares of tour trips with a single trip leg by primary mode.**

Figure 52 also depicts purpose-specific characteristics of tour-based trips. A majority of tours only include a single trip, and shares decrease as the number of trip legs increase, which are plausible results and consistent with the aforementioned result (figure 51).

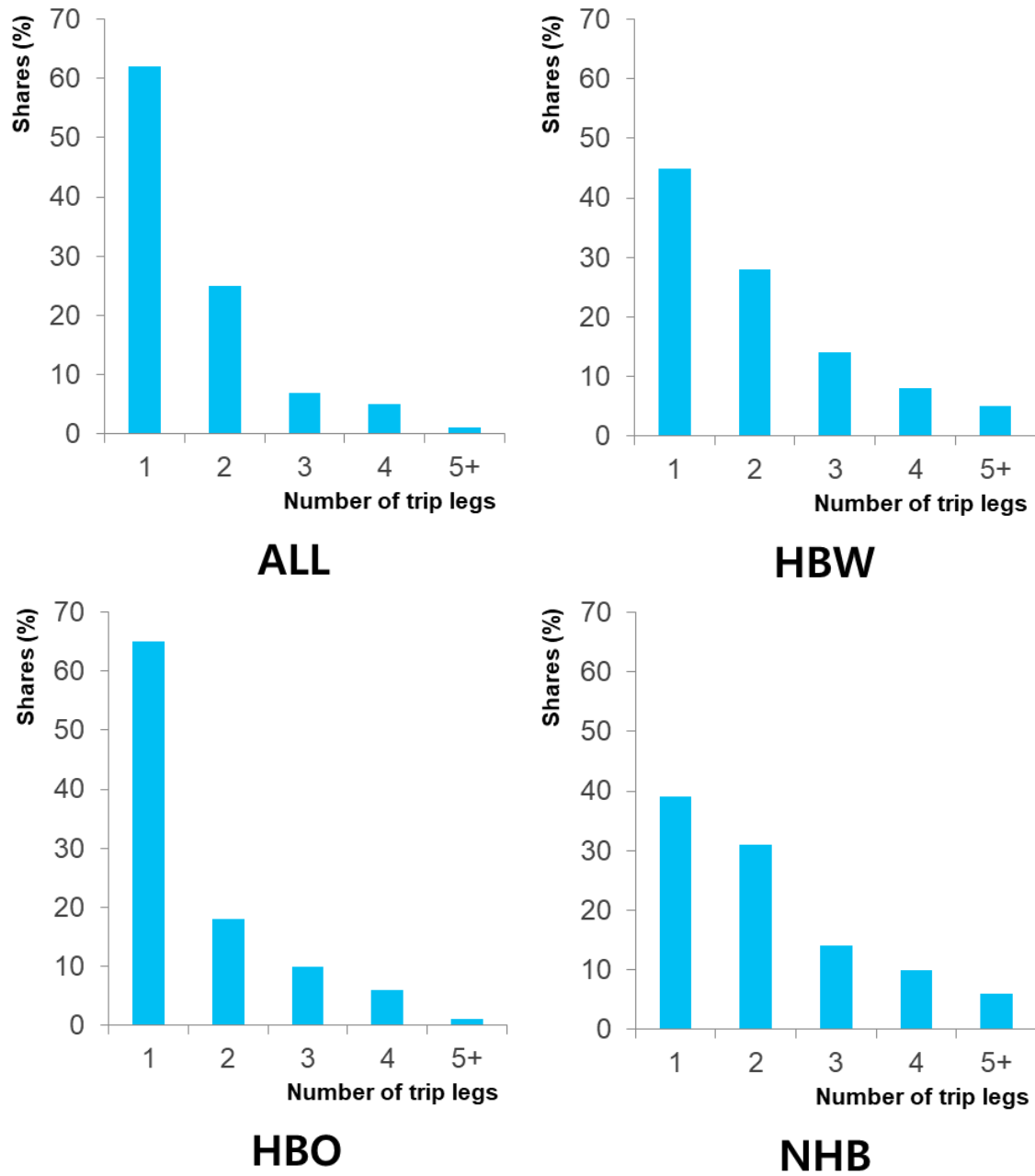


Figure 52. Histograms. Distributions of the number of trip legs by trip purpose.

## MODEL SPECIFICATIONS

Similar to the trip-based mode choice model (chapter 6), the tour-based model is estimated only for the “all purpose” including HBW, HBO, and NHB. For future estimation, purpose-specific

mode choice models may need to be developed because utility functions and magnitude of explanatory variables differ by trip purpose. As the research team clarified, however, the ultimate goal of this research is to develop an improved mode choice model in the regional behavior context. Thus, a single tour-based model covering all purposes is proposed in this research. Also, as shown in chapter 6, it is further classified into two parts based on travel distance (i.e., a threshold of 50 miles) since general travel behavior and mode choice sets are dissimilar between short- and long-distance trips. However, estimation of the tour-based model is confined to short-distance trips since: (1) a vast majority of trips in the NHTS are short-distance trips within 50 miles, and (2) estimating the mode choice model in chapter 6 also is implemented only for the short-distance model.

### **Overall Structure**

The tour-based model consists of two specific models: tour and trip models. The theoretical hypothesis in organizing such structure is that, as mentioned, travelers first determine a primary purpose and mode, which is called the tour mode, and then the rest of trips are made conditional on the primary trip. That is, the tour model only includes a single primary trip (i.e., one primary trip per tour), and the trip model includes the remaining trips within a tour (multiple secondary trips per tour). In estimation of both models, unequal choice sets are constructed based on the same theory and approach described in chapter 5.

### **Tour Model**

The research team regrouped the two-level mode choice set of the tour model, including four alternatives at the upper nest (i.e., auto, bike/walk, taxi, and transit). The bike/walk mode was separated in the trip-based model; however, that specification does not work for the tour-based



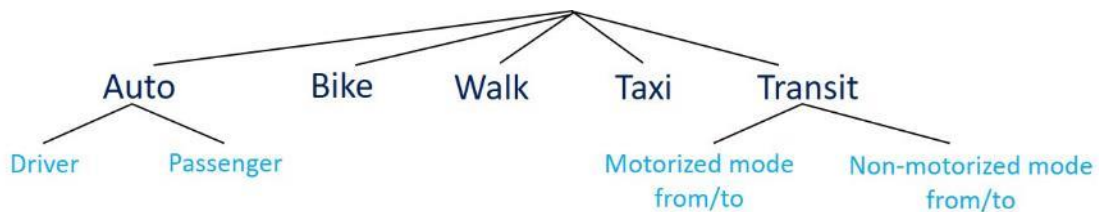
model. At the lower level, auto is divided into driver and passenger, whereas transit is not split into two specific modes (i.e., the nesting parameter for transit was not significant). Thus, the final model choice set includes six specific modes as illustrated in figure 53.



**Figure 53. Model diagram. Mode choice set with a nested structure (tour model).**

### Trip Model

The structure of the mode choice set for the trip model is identical to that of the conventional trip-based model. It consists of the two-level nested structure; five alternatives are at the upper level (i.e., auto, bike, walk, taxi, and transit). At the lower level, auto is divided into driver, and transit is split into two specific alternatives based on whether access/egress modes are motorized or not. Consequently, the final model choice set includes seven specific modes, as illustrated in figure 54.



**Figure 54. Model diagram. Mode choice set with a nested structure (trip model).**

## **Explanatory Variables**

The research team uses the same set of explanatory variables that was used to estimate the conventional trip-based model in chapter 6. The final set of explanatory variables comprises three categories, including mode attributes, SED traits, and accessibility. Table 40 shows the mode attribute variables. Travel time is classified into IVTT and OVTT. All modes have IVTT, but car and transit only have OVTT. Concerning public transit, three specific OVTT indicators, including terminal, waiting, and transfer times are merged altogether to calculate total OVTT. Travel cost is a single variable combining fuel cost, parking fee (only for car), and transit fare (only for transit).

Table 41 shows the SED variables. They are grouped into individual and household levels, and two additional variables—*active driver* and *vehicle sufficiency*—are created to explain whether individuals are able to drive and own available personal automobiles in their household.

**Table 40. Mode attribute variables of the tour-based mode choice model.**

Variable	Mode	Description	
In-vehicle Travel Time (IVTT) <sup>a</sup>	Car	Driving time	
	Taxi		
	Transit <sup>g</sup>	On-board time	
	Bike	Biking time	
	Walk	Walking time	
Out-of-vehicle Travel Time (OVTT)	Car <sup>b</sup>	Urban = 5 minutes	
		Suburban = 3 minutes	
	Rural = 1 minute		
Terminal time	Transit <sup>c,g</sup>	(1) Walking time from/to a station or	
		(2) driving time from/to a station if trip length is longer than 1 mile	
Waiting time, Transfer time	Transit <sup>g</sup>	High frequency = half of headway	
		Low frequency = 5 minutes	
Travel Cost (US dollars)	Fuel <sup>d</sup>	$\frac{\text{Trip distance (mile)}}{\text{Gas mileage (mpg)}} \times \text{gas price (\$/g)}$	
	Toll fee	Car Zero (not considered)	
	Parking fee (monthly) <sup>e</sup>	Car	Average daily parking fee for
			(1) those who are full-time workers and work in urban areas, or
	(2) those who are university (or graduate) students		
	Parking fee (one-time)	Car	Urban = \$4/hr (\$24/day more than 6 hrs) Second city = \$1 Other = \$0
	Fare (monthly pass) <sup>f</sup>	Transit <sup>g</sup>	Average daily transit fare (e.g., \$4/day) for those who are full-time workers or students
Fare (one-time)	Transit <sup>g</sup>	Assign transit fares depending on the transit line	

<sup>a</sup> IVTT was obtained from Google API.

<sup>b</sup> Car depends on locations of origin and destination.

<sup>c</sup> Transit was obtained from Google API. In particular, the threshold of 1 mile is determined based on the TCQSM.

<sup>d</sup> Official fuel economy data, including EVs and hybrid cars, were used.

<sup>e</sup> Parking fee (monthly) was applied to specific trip purposes (commute to work or attend school).

<sup>f</sup> Full-time workers or students were assumed to purchase the monthly pass for all trips.

<sup>g</sup> Information about available transit lines from Google API and relevant transit fare information, which were manually collected from MPO agencies, are combined.

**Table 41. Socioeconomic variables of the tour-based mode choice model.**

Category	Variable	Description
Individual	Female	1 = Yes, 0 = No
	Age	Continuous
	Under 16 years	1 = Yes, 0 = No
	Education	Categorical
	Employment	1 = worker, 0 = non-worker
Household	Active driver	1 = Yes, 0 = No
	Household size	Continuous
	Vehicle ownership	Continuous
	Income	Categorical
	Number of drivers	Continuous
Attitudes	Vehicle sufficiency	1 = Yes, 0 = No
	Perceived health status	1 (excellent) to 5 (poor)

Note: All socioeconomic variables are from the 2017 NHTS dataset.

Regarding transit accessibility, seven specific variables are initially chosen to estimate the final model, indicating to what extent public transit is accessible to travelers (table 42 and table 35) as introduced in tasks 1 and 4.

**Table 42. Accessibility variables of the mode choice model.**

Category	Variable	Description
AllTransit Score	Performance score	1 (poor) to 10 (excellent)
	Walkable neighborhood	Continuous
Transit Accessibility	# of jobs (workers) accessible in 30 min transit ride	Continuous
	Transit connectivity index	0 (poor) to 35 (excellent)
	Transit trip per week	Continuous
	# of transit stops within ½ miles	Continuous
	# of high frequency transit routes within ½ miles	Continuous

Note: Accessibility variables were defined using AllTransit data provided by the Center for Neighborhood Technology.

### Combination Rules for Tour and Trip Modes

In order to take associations between tour and trip mode choices in reality into account, thereby being conditioned to narrow modes down to specific available alternatives, the research team develops combination rules (table 43). Basically, all trip modes are constrained by the primary

tour mode via this rule, and each trip mode is confined based on what the tour mode is. The final correspondence rules are determined based on the following principles:

- Taxi and transit trips are not allowed when the tour mode is Driver within the same tour.
- When the tour mode is Passenger, all trip modes other than Driver are available.
- Bike/Walk and Taxi tours do not include auto trips for particular trip legs within the same tour.
- With respect to Transit tours, only the Driver alternative is not allowed for the trip mode since driving in a transit tour is illogical behavior.

**Table 43. Combination rules for tour and trip modes.**

Tour Mode	Trip Mode				
	Driver	Passenger	Bike/Walk	Taxi	Transit
Driver	●	●	●		
Passenger		●	●	●	●
Bike/Walk			●	●	●
Taxi			●	●	●
Transit		●	●	●	●

Note: ● indicates that the trip mode is available given the corresponding tour mode.

## ESTIMATION RESULT

The final estimation result is presented in table 36. As described above, the research team develops two specific models: tour and trip models. The tour model is first estimated, then the trip model is developed sequentially (please note that both models are all-purpose models based on the ultimate goal of this research to propose an improved method for the Georgia Statewide Model, as discussed in chapter 5). While initial attempts include multinomial logit models with various explanatory variables, those models did not exhibit reasonable goodness of fit. As a result, nested logit forms with two hierarchies are specified in the final model.

## **Tour Model**

As shown in table 44 and equation 9, the final tour model exhibits an acceptable model fit (adjusted rho squared = 0.328). All explanatory variables except for two taxi-related variables are statistically significant at the 5 percent level with the expected signs. The estimation result is similar to the traditional mode choice model presented in chapter 5: (1) travelers are much more likely to prefer private modes (auto) over other modes (walk/bike or public transit); (2) travel time variables are alternative-specific, whereas travel cost is a generic variable with all negative impacts on travelers' utility; (3) coefficients for taxi tend to be insignificant (constant and OVTT are not significant at the 10 percent level); (4) driving availability (i.e., active driver and vehicle sufficiency) is still a critical indicator explaining mode choice behavior, and (5) transit accessibility has a positive influence on choosing public transit. Regarding VOTTS, the tour model exhibits reasonable values; VOTTS for driver, passenger, taxi, and public transit are \$47/hr, \$37/hr, \$54/hr, and \$10/hr, respectively, which are significantly lower than previous values obtained from the conventional mode choice model. In particular, VOTTS for driving is decreased by almost half of the former one (it was \$82/hr), which strongly supports the research hypothesis and the motivation for proposing the tour-based model.

**Table 44. Tour-based mode choice model (tour model).**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>b</sup></b>	<b>Z-value</b>	<b>P-value</b>
<i>Alternative Specific Constant</i>				
Drive	1.762	0.074	23.928	<0.001
Passenger	0.667	0.041	16.313	<0.001
Eco <sup>a</sup>	-	-	-	-
Taxi	-0.615	0.456	-1.350	0.177
Public transit	-1.963	0.102	-19.311	<0.001
<i>Mode Attributes</i>				
IVTT_auto	-0.098	0.015	-6.430	<0.001
IVTT_passenger	-0.078	0.010	-8.097	<0.001
IVTT_taxi	-0.113	0.038	-2.981	0.003
IVTT_PT	-0.021	0.002	-10.355	<0.001
OVTT_passenger	-0.762	0.140	-5.433	<0.001
OVTT_taxi	-0.183	0.215	-0.851	0.395
OVTT_public transit	-0.112	0.024	-4.690	
Total time_eco	-0.193	0.060	-3.196	0.001
Cost	-0.126	0.015	-8.301	<0.001
<i>SED</i>				
Female	-0.151	0.032	-4.777	0.005
Age	-0.087	0.028	-3.090	0.002
Active driver	0.261	0.010	25.213	<0.001
Vehicle sufficiency	0.836	0.040	20.961	<0.001
<i>Accessibility</i>				
Average performance score	0.136	0.013	10.321	<0.001
<i>Tour Characteristics</i>				
Shopping trips within a tour (dummy)	0.341	0.041	8.412	<0.001
Number of stops within a tour_auto	0.265	0.024	10.994	<0.001
Number of stops within a tour_PT	-0.108	0.006	-18.004	<0.001
<i>Nesting Parameters</i>				
Auto	0.902	0.048	18.831	<0.001

N = 19,318

LL(c) = -14,897.83

LL( $\beta$ ) = -9,993.34

Adjusted  $\rho^2$  (MS) = 0.328

<sup>a</sup> Eco includes walk and bike modes. It is defined as the reference mode in the final model.

<sup>b</sup> SE is the standard error.

## Trip Model

Table 45 and equation 10 demonstrate the estimation result of the trip model. It produces plausible results and a respectable model fit with adjusted rho-squared (market shared) of 0.368.

A vast majority of coefficients are statistically significant at the 5 percent level, except the taxi-related coefficients. It tells us that the current variables cannot explain travelers' utility for taxi, calling for additional investigation on potential factors associated with the taxi mode.

$$\begin{aligned}
U(\text{Driver}) &= 1.762 - 0.098(\text{IVTT}_{\text{auto}}) - 0.126(\text{cost}) \\
&\quad + 0.836(\text{vehicle sufficiency}) + 0.261(\text{active driver}) \\
&\quad + 0.479(\text{worker}) + 0.341(\text{shop}) \\
&\quad + 0.265(\text{\# of stops within a tour}_{\text{auto}}) \\
U(\text{Passenger}) &= 0.667 - 0.078(\text{IVTT}_{\text{pax}}) - 0.762(\text{OVTT}_{\text{pax}}) - 0.126(\text{cost}) \\
U(\text{Bike, Walk}) &= 0 - 0.193(\text{TT}) - 0.151(\text{female}) - 0.087(\text{age}) \\
&\quad + 0.136(\text{avg. ps}) \\
U(\text{Taxi}) &= -0.615 - 0.113(\text{IVTT}_{\text{taxi}}) - 0.183(\text{OVTT}_{\text{taxi}}) - 0.126(\text{cost}) \\
U(\text{Transit}) &= -1.963 - 0.021(\text{IVTT}_{\text{PT}}) - 0.112(\text{OVTT}_{\text{PT}}) - 0.126(\text{cost}) \\
&\quad - 0.087(\text{age}) + 0.136(\text{avg. ps}) \\
&\quad - 0.108(\text{\# of stops within a tour}_{\text{PT}})
\end{aligned}
\tag{10}$$

The estimation result is consistent with the previous models presented in this research (i.e., the conventional mode choice models in chapter 5 and the tour model). Three types of explanatory variables (i.e., mode attributes, SED, and accessibility) are incorporated in the final model. It shows that: (1) travelers tend to prefer private modes to public transit; (2) travel time and cost have negative impacts on travelers' utility, as expected; (3) female travelers are more likely to use automobiles as a driver or passenger, while younger people tend to prefer public transit; and (4) transit accessibility is directly associated with preference for public transit and bike/walk (it suggests that proximity to transit stations can encourage people to use both public transit and nonmotorized modes, including walking and biking).

Most notably, two additional explanatory variables indicating the number of shopping trips and stops (i.e., number of destinations) within each tour are selected in the trip model. It offers clear evidence that tour characteristics need to be considered in the mode choice model, pointing to the



necessity of adopting a tour-based model to fully describe travelers' mode choice behavior. With respect to VOTTS, the trip model also demonstrates plausible outcomes. VOTTS for driver, passenger, taxi, and public transit are \$37/hr, \$34/hr, \$40/hr, and \$12/hr, respectively, which are similar to those in the tour model and significantly lower than the trip-based mode choice model.

**Table 45. Tour-based mode choice model (trip model).**

<b>Explanatory Variable</b>	<b>Coefficient</b>	<b>SE<sup>d</sup></b>	<b>Z-value</b>	<b>P-value</b>
<i>Alternative Specific Constant</i>				
Drive	1.361	0.179	7.610	<0.001
Passenger	0.807	0.127	6.350	<0.001
Eco <sup>a</sup>	-	-	-	-
Taxi	0.366	0.574	0.638	0.177
Public transit (motorized AE)	-1.462	0.142	-10.311	<0.001
Public transit (nonmotorized AE)	-1.787	0.113	-15.867	<0.001
<i>Mode Attributes</i>				
IVTT_auto <sup>b</sup>	-0.086	0.002	-43.93	<0.001
IVTT_passenger	-0.078	0.003	-24.441	<0.001
IVTT_taxi	-0.093	0.033	-2.785	0.005
IVTT_PT	-0.028	0.001	-38.915	<0.001
OVTT_taxi	-0.126	0.004	-29.121	0.395
Total time_eco <sup>c</sup>	-0.131	0.015	-8.511	<0.001
Cost	-0.138	0.004	-34.058	<0.001
<i>SED</i>				
Female	0.162	0.021	7.857	<0.001
Age	-0.031	0.010	-3.104	0.001
Active driver	0.430	0.034	12.534	<0.001
Vehicle sufficiency	0.677	0.022	31.097	<0.001
<i>Accessibility</i>				
Average performance score	0.074	0.026	2.831	0.005
<i>Tour Characteristics</i>				
Number of shopping trips within a tour	0.203	0.004	48.012	<0.001
Number of stops within a tour	-0.167	0.005	-30.802	<0.001
<i>Nesting Parameters</i>				
Auto	0.847	0.068	12.514	<0.001
Public transit	0.285	0.091	3.121	0.002

N = 34,197

LL(c) = -26,561.04

LL( $\beta$ ) = -16,767.57

Adjusted  $\rho^2$  (MS) = 0.368

<sup>a</sup> Walk is the reference mode.

<sup>b</sup> This coefficient is also used for OVTT for transit with motorized AE.

<sup>c</sup> This coefficient is also used for OVTT for transit with nonmotorized AE.

<sup>d</sup> SE is the standard error.

$$\begin{aligned}
U(\text{Driver}) &= 1.361 - 0.086(\text{IVTT}_{\text{auto}}) - 0.138(\text{cost}) \\
&\quad + 0.677(\text{vehicle sufficiency}) + 0.430(\text{active driver}) \\
&\quad + 0.162(\text{female}) + 0.320(\# \text{ of stops within a tour}_{\text{auto}}) \\
&\quad + 0.203(\# \text{ of shopping trips within a tour}) \\
U(\text{Passenger}) &= 0.807 - 0.078(\text{IVTT}_{\text{pax}}) - 0.138(\text{cost}) + 0.162(\text{female}) \\
U(\text{Bike, Walk}) &= 0 - 0.131(\text{TT}) - 0.031(\text{age}) + 0.074(\text{avg. ps}) \\
&\quad - 0.167(\# \text{ of stops within a tour}_{\text{non-auto}}) \\
U(\text{Taxi}) &= 0.366 - 0.093(\text{IVTT}_{\text{taxi}}) - 0.126(\text{OVTT}_{\text{taxi}}) - 0.138(\text{cost}) \\
U(\text{PT}_{\text{motor}}) &= -1.462 - 0.028(\text{IVTT}_{\text{PT}}) - 0.086(\text{OVTT}_{\text{PT}}) - 0.138(\text{cost}) \\
&\quad - 0.031(\text{age}) + 0.074(\text{avg. ps}) \\
&\quad - 0.167(\# \text{ of stops within a tour}_{\text{non-auto}}) \\
U(\text{PT}_{\text{non-motor}}) &= -1.787 - 0.028(\text{IVTT}_{\text{PT}}) - 0.131(\text{OVTT}_{\text{PT}}) - 0.138(\text{cost}) \\
&\quad - 0.031(\text{age}) + 0.074(\text{avg. ps}) \\
&\quad - 0.167(\# \text{ of stops within a tour}_{\text{non-auto}})
\end{aligned} \tag{11}$$

## LIMITATIONS ON APPLICATION OF THE TOUR-BASED APPROACH TO THE GSTDM

At present, there are several limitations on application of the tour-based mode choice model to the GSTDM, although the newly developed model well describes the current travelers' choice behavior. First and foremost, the proposed model specification and the definition of trips is not compatible with the current GSTDM, which is the trip-based model. Specifically, O-D tables obtained after the trip distribution stage denote zone-to-zone trips for each zone pair. In the mode choice stage, those trips are divided into several mode-specific trips via a choice model. In this process, all trips are treated as independent, and trip chaining and relationship between trips are not considered. To take tour-based characteristics into account, those independent trips need to be grouped into a tour that belongs to the same home-to-home loop for each traveler. However, this framework cannot be applied in the middle of the four-step stage; a completely different modeling framework for the tour-based approach needs to be established. Thus, additional exploration is called for in order to propose a completed modeling framework.

Second, some variables indicating tour characteristics incorporated in the tour-based model will not be available when estimating future choices because those variables (e.g., the number of stops made within a tour) cannot be determined in the current trip-based modeling framework, meaning that additional methodologies need to be implemented to define them. That is, separate models are also required to account for characteristics of trip chaining since the traditional choice model does not incorporate tour-related indicators. For example, ARC implements an activity-based model, which is called Coordinated Travel-Regional Activity Based Modeling Platform (CT-RAMP), based on the same modeling approach. Interestingly, CT-RAMP separately estimates several choice and frequency models, including time-of-day choice, intermediate stop destination choice, and tour stop frequency models. These are necessary to identify tour characteristics so that critical tour-based variables can be defined and utilized in the tour-based choice model as explanatory variables.

## CHAPTER 8. SUMMARY AND CONCLUSIONS

In this report, the research team outlined a number of areas for improvement in the Georgia Statewide Travel Demand Model. We focused on the development of a vehicle ownership model, a time-of-day segmentation, and destination choice and mode choice models for the GSTDM. By incorporating these updates, the statewide model can have a more realistic and accurate representation of travel through the state of Georgia.

The main data source used to develop these analyses was the 2017 National Household Travel Survey, which contains data on over 8000 households and 56,000 trips for the state of Georgia. We worked closely with the team members of another GDOT project, “Analysis of the Georgia Add-on to the 2016–2017 National Household Travel Survey” who were already working on a 2017 NHTS analysis and provided this project with a more enriched 2017 NHTS dataset on which to build our analysis. We further augmented the 2017 NHTS data with multiple other sources based on our needs, including adding AllTransit data with variables on transit access, Census data to add extra geographical features, distance skim files to add distances among TAZs, and Google API to add travel information on multiple mode/trip characteristics. The details on data augmentations are discussed more closely in each of the chapters of this report.

In the first task of this project, we started with investigating the vehicle ownership choice of Georgian households. We used various disaggregate models to gain insight into how Georgian households choose the number of vehicles they own, and compared these models on their insights and prediction accuracy. These results are especially useful to the GSTDM should GDOT decide to upgrade the current trip-based model to a disaggregate activity-based model in the future. For the current aggregate trip-based GSTDM, however, we estimated another linear

model whose results could have been correctly aggregated to the TAZ level from the household level. The output of this task, the average vehicle ownership for all TAZs, is ready to be used in the other steps of the statewide model.

In the second task of this project, we investigated methods to introduce time of day in the GSTDM. After reviewing the different time-of-day methodologies in the literature, and in consultation with the Office of Planning at GDOT, we adopted a “trips-in-motion” approach. We discussed the details of this approach and how it compares with the other traditional TOD methods, and concluded that this approach produces a more accurate estimate of the proportion of trips that occur in each time period. Analyzing the temporal distribution of the 2017 NHTS data, we proposed four TOD periods, namely AM peak, Midday, PM peak, and Night periods, and presented shares of trips, or time-of-day factors, by each period and trip purpose, along with TOD factors for through trips. Those newly estimated factors can be utilized in the current GSTDM when predicting peak-time or hourly traffic volumes. Notably, the research team proposed applying the TOD factors right after the trip-generation step, not after the trip-assignment step (postprocessing), which will lead to more accurate travel forecasts after accounting for time-specific travel time and traffic congestion.

In the third task of this project, we focused on improving the trip-distribution step in the GSTDM. Reviewing the literature and the practice in other states, we proposed using a destination choice model in lieu of the gravity model currently used by the GSTDM for the short-distance trips. We used multiple data sources to further complement the 2017 NHTS dataset, and estimated 12 destination choice models, one for each time of day (AM peak, Midday, PM peak, Night) and trip purpose (HBW, HBO, NHB). The models showed promising

results, and provide more flexibility in including socioeconomic characteristics in trip distribution.

In the fourth task of this project, the team aimed to improve the modal split step of the GSTDM. We used the 2017 NHTS Georgia trip file and used the Google API to add data on alternative mode characteristics. To better understand mode choice, the research team separated short-distance trips and long-distance trips. We tested multiple model structures to get the best short-distance mode choice results, including MNL and nested MNL. The final model was the nested logit model with seven specific modes (driving, passenger, walk, bike, taxi, public transit with motorized/nonmotorized first- and last-mile modes). It demonstrated the decent goodness of fit with the expected signs, but value of travel time savings for auto (driving and passenger) is estimated substantially higher than expected, which may result from the fundamental limitation of the traditional trip-based approach. To address this issue, the research team further proposed a tour-based model as an additional task (fifth task). Regarding the long-distance trips, however, the team could not estimate a satisfactory model because of the relatively small number of long-distance cases in the NHTS dataset. We, however, provided exploratory insights into long-distance trips, and discussed recommendations on remedying the lack of long-distance data such as collecting and analyzing data from other states that are most similar to Georgia.

In the fifth task of this project, the team proposed a tour-based mode choice model in the context of the GSTDM. It improves the fundamental limitation of the trip-based approach that all trips are treated as independent and separated travel (resulting in counterintuitive VOTTS for auto) by taking *tours* into account. A tour includes combined trips with multiple stops that belong to the same home-to-home loop. Similar to the traditional mode choice model that the research team developed in the fourth task, the tour-based mode choice model is divided into two specific

models (short- and long-distance models) based on a threshold of 50 miles. The research team developed a single short-distance model including all purposes, given that a vast majority of trips in NHTS are shorter than 50 miles and mode-specific models are currently not necessary within the frame of the Georgia statewide model. The same set of explanatory variables used in the trip-based model (fourth task) were incorporated, and the same mode choice structure of a nested form were considered in the final model. The estimation model demonstrated a respectable model fit, and critical determinants such as mode attributes (travel time and cost) were mostly significant at the 5 percent level, with the expected signs. Most importantly, it provided a plausible range of VOTTS for auto, which are more consistent with typical VOTTS for auto proposed by USDOT and other previous empirical studies.



## APPENDIX

## ACKNOWLEDGMENTS

This project builds on the work done by a concurrent Georgia Department of Transportation project (Kash, Mokhtarian, and Circella 2021). We express our gratitude to Dr. Gwen Kash and Professor Patricia Mokhtarian for sharing their work and helping develop the initial analyses for this project. The latest version of the GSTDM was also shared with us by GDOT and their former consultants (HNTB, Atkins) who initially developed and then further updated the previous versions of the Statewide Travel Demand Model used by GDOT. We also thank South Carolina's Department of Transportation in their cooperation in sharing their NHTS add-on dataset. Finally, our team extends special thanks to Mr. Habte Kassa and Ms. Sarah Lamothe from the GDOT Office of Planning, who proactively followed the activities of this project, provided valuable feedback and guidance throughout the project, and helped us access datasets and information needed to carry out our work. Mr. Kassa not only served as the technical/implementation manager for this project, but he also acted as an active member of our research team and considerably contributed to the successful completion of the project.

## REFERENCES

- Anowar, S., Yasmin, S., Eluru, N., and Miranda-Moreno, L.F. (2014). “Analyzing Car Ownership in Quebec City: A Comparison of Traditional and Latent Class Ordered and Unordered Models.” *Transportation* 41(5), pp. 1013–1039. Available online: <https://doi.org/10.1007/s11116-014-9522-9>.
- Beck, M.J., Rose, J.M., and Hensher, D.A. (2013). “Environmental Attitudes and Emissions Charging: An Example of Policy Implications for Vehicle Choice.” *Transportation Research Part A: Policy and Practice* 50, pp. 171–182. Available online: <https://doi.org/10.1016/j.tra.2013.01.015>.
- Ben-Akiva, M.E. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Transportation Studies series, Vol. 9, MIT Press, Cambridge, MA.
- Bhat, C.R. and Pulugurta, V. (1998). “A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions.” *Transportation Research Part B: Methodological* 32(1), pp. 61–75. Available online: [https://doi.org/10.1016/S0191-2615\(97\)00014-3](https://doi.org/10.1016/S0191-2615(97)00014-3).
- Bowman, J.L. and Ben-Akiva, M.E. (2001). “Activity-based Disaggregate Travel Demand Model System with Activity Schedules.” *Transportation Research Part A: Policy and Practice* 35(1), pp. 1–28. Available online: [https://doi.org/10.1016/S0965-8564\(99\)00043-9](https://doi.org/10.1016/S0965-8564(99)00043-9).
- Breiman, L. (2001). “Random Forests.” *Machine Learning* 45(1), pp. 5–32. Available online: <https://doi.org/10.1023/A:1010933404324>.
- Cutler, A., Cutler, D.R., and Stevens, J.R. (2011). “Random Forests.” In *Ensemble Machine Learning: Methods and Applications* (Zhang, C. and Ma, Y.Q., eds.), Springer, New York, pp. 157–175. Available online: [http://dx.doi.org/10.1007/978-1-4419-9326-7\\_5](http://dx.doi.org/10.1007/978-1-4419-9326-7_5).
- Efraimidis, P. and Spirakis, P. (2008). “Weighted Random Sampling.” In *Encyclopedia of Algorithms* (Kao, M.-Y., ed.), Springer, Boston, MA, pp. 1024–1027. Available online: <https://doi.org/10.1007/978-0-387-30162-4>.
- Ermagun, A., Rashidi, T.H., and Lari, Z.A. (2015). “Mode Choice for School Trips: Long-Term Planning and Impact of Modal Specification on Policy Assessments.” *Transportation Research Record: Journal of the Transportation Research Board* 2513(1), pp. 97–105. Available online: <https://doi.org/10.3141/2513-12>.
- Claritus. (2020). Claritas Prizm® Premier Methodology. Accessed in July 2021 at: <https://enviroanalytics.com/docs/default-source/us---data-product-support-documents/claritas-prizm-premier-methodology-ea.pdf>
- Federal Highway Administration (FHWA). (2013). *Georgia Department of Transportation (GDOT) Statewide Travel Model Peer Review Report*. Report FHWA-HEP-13-031, Travel Model Improvement Program, U.S. Department of Transportation, FHWA, Washington, DC.

Available online:

[http://www.dot.ga.gov/InvestSmart/TravelDemandModels/PeerReview2012\\_FullReport.pdf](http://www.dot.ga.gov/InvestSmart/TravelDemandModels/PeerReview2012_FullReport.pdf).

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009). “Sampling of Alternatives for Route Choice Modeling.” *Transportation Research Part B: Methodological* 43(10), pp. 984–994. Available online: <https://doi.org/10.1016/j.trb.2009.03.001>.

Georgia Department of Transportation (GDOT). (2018). *Georgia’s Traffic Monitoring Guide*. Report, GDOT, Office of Transportation Data, Atlanta, GA. Available online: [http://www.dot.ga.gov/DriveSmart/Data/Documents/Guides/2018\\_Georgia\\_Traffic\\_Monitoring\\_Program.pdf](http://www.dot.ga.gov/DriveSmart/Data/Documents/Guides/2018_Georgia_Traffic_Monitoring_Program.pdf).

Google. (2021). *Distance Matrix API documentation*. Accessed in July 2021 at: <https://developers.google.com/maps/documentation/distance-matrix/overview>.

Greene, W.H. and Hensher, D.A. (2003). “A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit.” *Transportation Research Part B: Methodological* 37(8), pp. 681–698. Available online: [https://doi.org/10.1016/S0191-2615\(02\)00046-2](https://doi.org/10.1016/S0191-2615(02)00046-2).

HNTB. (2019). *2015/2050 Georgia Statewide Travel Demand Model*. Report, Georgia Department of Transportation, Office of Planning, Atlanta, GA. Available online: <http://www.dot.ga.gov/InvestSmart/TravelDemandModels/StatewideModelFactsheet.pdf>.

Kash, G., Mokhtarian, P.L., and Circella, G. (2021). *Analysis of the Georgia Add-on to the 2016–2017 National Household Travel Survey*. FHWA-GA-21-18-24, Georgia Department of Transportation, Atlanta, GA.

Kim, J. and Lee, S. (2017). “Comparative Analysis of Traveler Destination Choice Models by Method of Sampling Alternatives.” *Transportation Planning and Technology* 40(4), pp. 465–478. Available online: <https://doi.org/10.1080/03081060.2017.1300242>.

Krizek, K.J. (2003). “Neighborhood Services, Trip Purpose, and Tour-based Travel.” *Transportation* 30(4), pp. 387–410. Available online: <https://doi.org/10.1023/A:1024768007730>.

Ma, J. and Demetsky, M.J. (2013). *Integration of Travel Demand Models with Operational Analysis Tools*. Final Report VCTIR 14-R5, Virginia Center for Transportation Innovation and Research, Charlottesville, VA. Available online: [http://www.virginiadot.org/vtrc/main/online\\_reports/pdf/14-r5.pdf](http://www.virginiadot.org/vtrc/main/online_reports/pdf/14-r5.pdf).

McGuckin, N. and Nakamoto, Y. (2004). *Trips, Chains and Tours—Using an Operational Definition*. Presented at the 2004 National Household Travel Survey Conference, November 1–2, Washington, DC. Available online: <http://onlinepubs.trb.org/onlinepubs/archive/conferences/nhts/McGuckin.pdf>.

Mishra, S., Wang, Y., Zhu, X., Moeckel, R., and Mahapatra, S. (2013). *Comparison Between Gravity and Destination Choice Models for Trip Distribution in Maryland*. Paper presented at the Transportation Research Board 92<sup>nd</sup> Annual Meeting, January 13–17, Washington, DC.

- Moeckel, R., Donnelly, R., and Ji, J. (2019). *Statewide Transportation Models in the U.S.: A Review of the State of Practice*. Paper presented at the Transportation Research Board 98<sup>th</sup> Annual Meeting, January 13–17, Washington, DC.
- Mohammadian, A. and Miller, E.J. (2003). “Dynamic Modeling of Household Automobile Transactions.” *Transportation Research Record: Journal of the Transportation Research Board* 1831(1), pp. 98–105. Available online: <https://doi.org/10.3141/1831-11>.
- Peevy, P. and Kassa, H. (2012). *Statewide Travel Demand Model: GDOT*. Presentation to the Atlanta Regional Commission Model Users Group.
- U.S. Department of Transportation (USDOT). (2016). *The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations Revision 2 (2016 Update)*. USDOT, Office of the Secretary of Transportation, Washington, DC. Available online: <https://www.transportation.gov/sites/dot.gov/files/docs/2016%20Revised%20Value%20of%20Travel%20Time%20Guidance.pdf>.
- WSP/Parsons Brinckerhoff. (2015). North Carolina Statewide Transportation Model Generation 2.0 (NCSTMGen2), prepared for North Carolina Department of Transportation.
- Zhang, Y. and Xie, Y. (2008). “Travel Mode Choice Modeling with Support Vector Machines.” *Transportation Research Record: Journal of the Transportation Research Board* 2076(1), pp. 141–150. Available online: <https://doi.org/10.3141/2076-16>.
- Zhao, Y. and Kockelman, K.M. (2002). “The Propagation of Uncertainty Through Travel Demand Models: An Exploratory Analysis.” *The Annals of Regional Science*, 36(1), pp. 145–163. Available online: <https://doi.org/10.1007/s001680200072>.