Leveraging Connected Vehicle Data to Improve Travel Demand Modeling

Dr. Shanjiang Zhu
Associate Professor
Dept. of Civil, Environmental and Infrastructure Engineering
George Mason University

Dr. Chenfeng Xiong
Assistant Professor
Dept. of Civil and Environmental Engineering
Villanova University

VDOT TPR: Peng Xiao (TMPD), Ying Su (TMPD), Xiao Cui (NOVA District), Lance Dougald (VTRC)

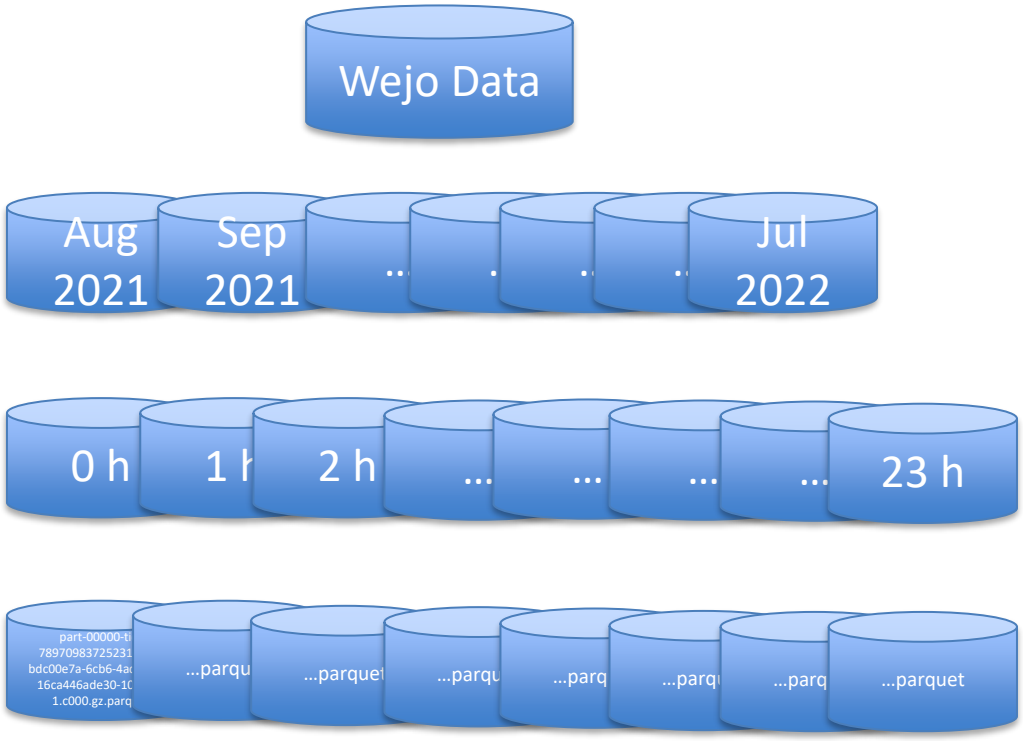
Research Objective

- Background: Analyzing Connected Vehicle (CV) trajectory data (Wejo data) with the objective of improving VDOT travel demand modeling.
- Providing nuanced information about trip time, trip distance, path patterns with fine geographic and temporal resolution.
- Integrating revealed routes based on CV trajectory data into traffic assignment models.

Background

- CV data: Data transmitted by a connected vehicle's onboard systems and devices, including its real-time position, operating performance, and event logs.
- Wejo Group Ltd was a British connected-vehicle data start-up that supplied CV data, including high-frequency vehicle location data (one point every three seconds) and vehicle event data, for planning, operational, and safety applications.
- Data are anonymized, and later are further "fuzzed" to protect privacy.
- Wejo filed for bankruptcy in 2023, but similar data products are now available from other vendors.
- Data is provided in Parquet format, a column-oriented storage format designed for faster query performance and improved compression. Big data analysis techniques are needed to analyze CV data effectively.

CV Data Processing Flow



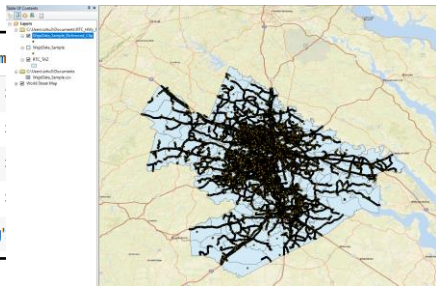
1.79TB

11GB/day

~1GB/hour

23MB/parquet

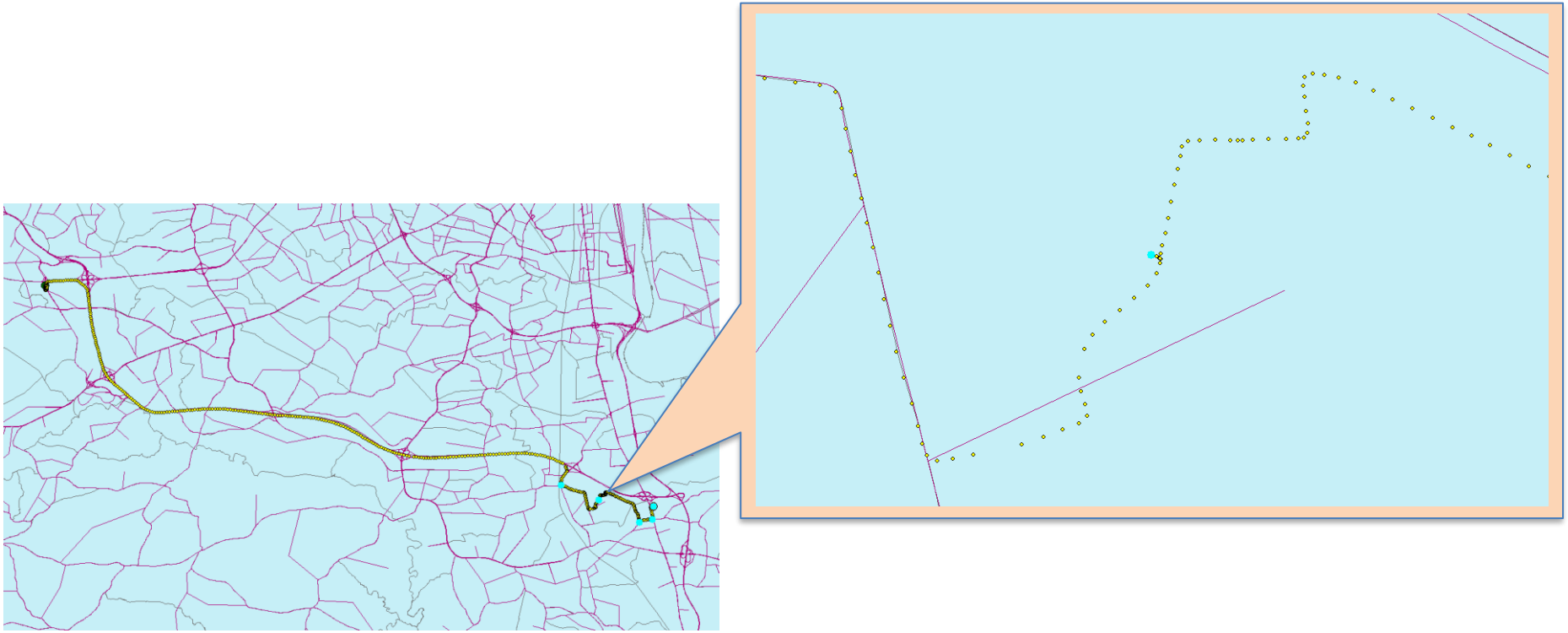
	dataPointId	journeyId	capturedTimestamp	location	m
0	9864d64f-5564-475f-bb2d-edacc77b3796-0-0	01749d4ce482d51796ddc93180ea835f1c39ff7e	2022-06-20T20:21:56.712-0400	{'latitude': 37.3658579, 'longitude': -79.2615...	{'speed': 0.0, 'heading':
2	9864d64f-5564-475f-bb2d-edacc77b3796-1	01749d4ce482d51796ddc93180ea835f1c39ff7e	2022-06-20T20:21:59.712-0400	{'latitude': 37.365860399999995, 'longitude': ...	{'speed': 0.0, 'heading':
4	9864d64f-5564-475f-bb2d-edacc77b3796-2	01749d4ce482d51796ddc93180ea835f1c39ff7e	2022-06-20T20:22:02.712-0400	{'latitude': 37.3658699, 'longitude': -79.2615...	{'speed': 3.22, 'heading':
6	9864d64f-5564-475f-bb2d-edacc77b3796-3	01749d4ce482d51796ddc93180ea835f1c39ff7e	2022-06-20T20:22:05.712-0400	{'latitude': 37.3659141, 'longitude': -79.2615...	{'speed': 9.66, 'heading':
8	9864d64f-5564-475f-bb2d-edacc77b3796-4	01749d4ce482d51796ddc93180ea835f1c39ff7e	2022-06-20T20:22:08.712-0400	{'latitude': 37.3659505, 'longitude': -79.2614...	{'speed': 17.7, 'heading'



Example of a Journey with Key_On and Key_Off events

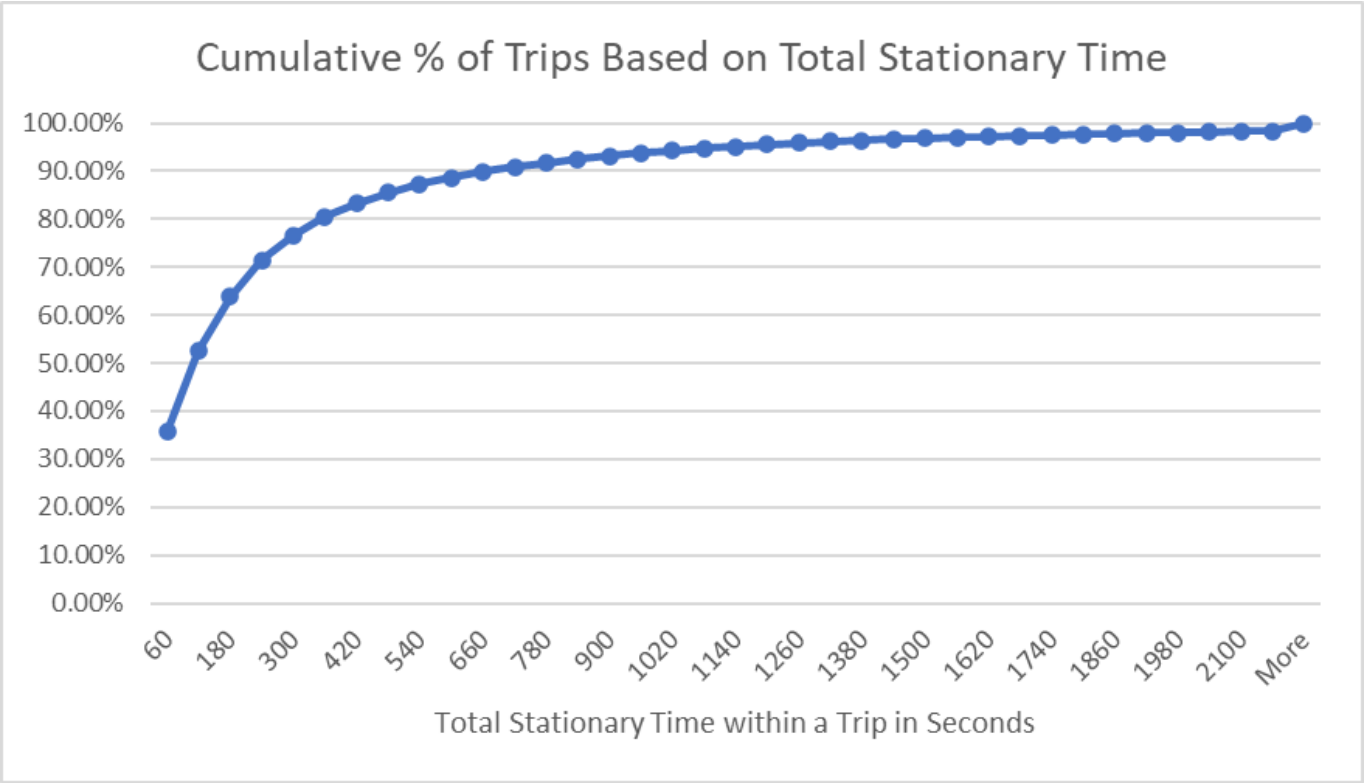
- *A Key_On and Key_Off event marks the start and end of a trip based on the engine status.*
- *Key_On in Green; Key_Off in Red, Mid_journey in Yellow*
- *This trip started on 05/23/2022 16:26:18, lasted 65.7 minutes and traveled 18.61 miles.*
- *JID=4*

Trips with lengthy stops in the middle



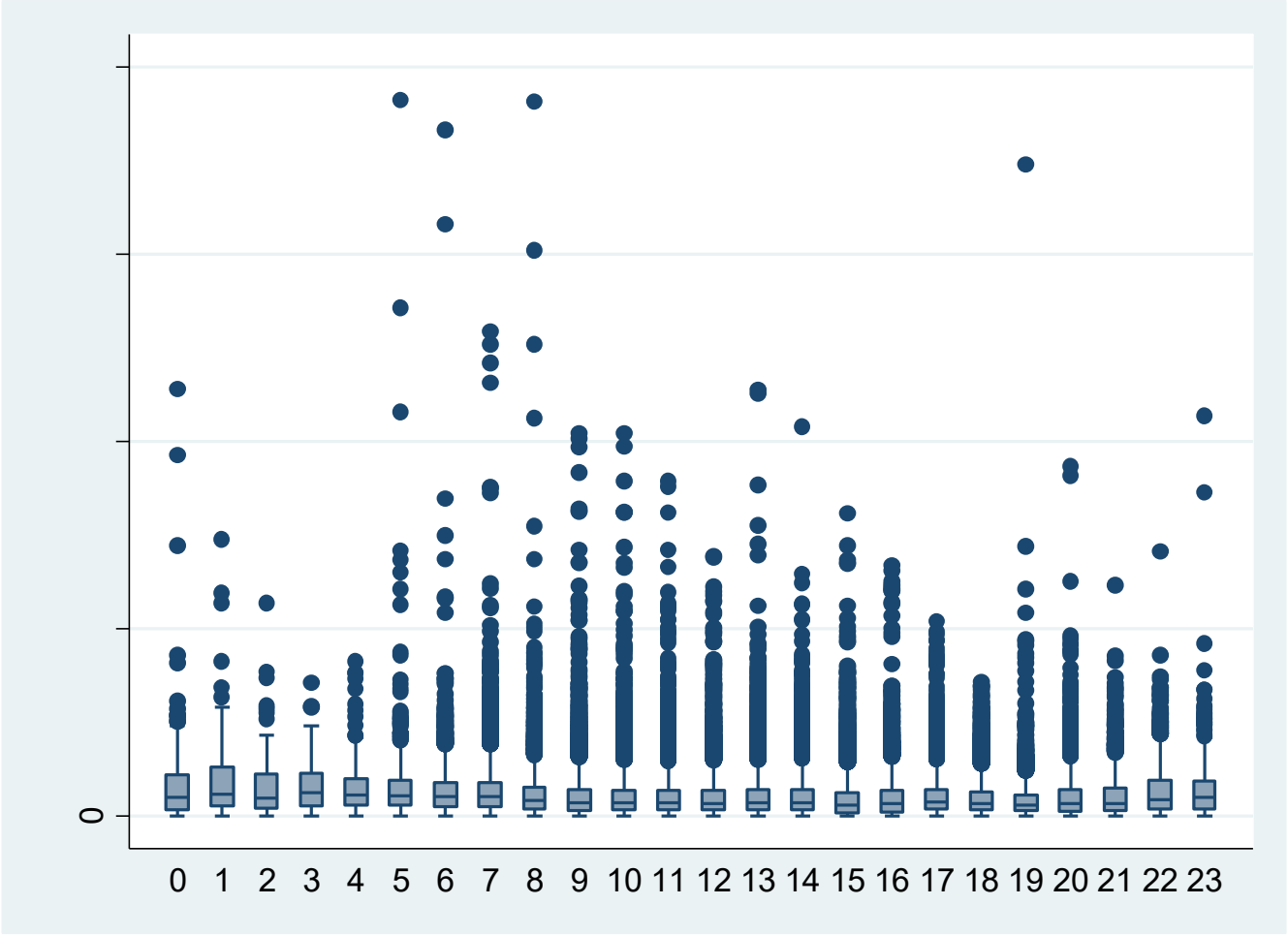
- *By flagging all the stationary intervals in light blue, the vehicle has stopped in five locations. Three of them are typical stops at signals based on the road network. The vehicle clearly pulled at the side of the road for a significant amount of time (26.5 mins) at the second stop, and it did not turn off its engine immediately at the end of the trip (for 12.35 mins).*

Trips with lengthy stops in the middle



- *Over the 109,252 trips identified on 05/23/2022 (UTM), 88.69% have a total stationary interval shorter than 10 minutes within the trip, 4.99% have a total stationary interval longer than 20 minutes.*

Trip Time Distribution



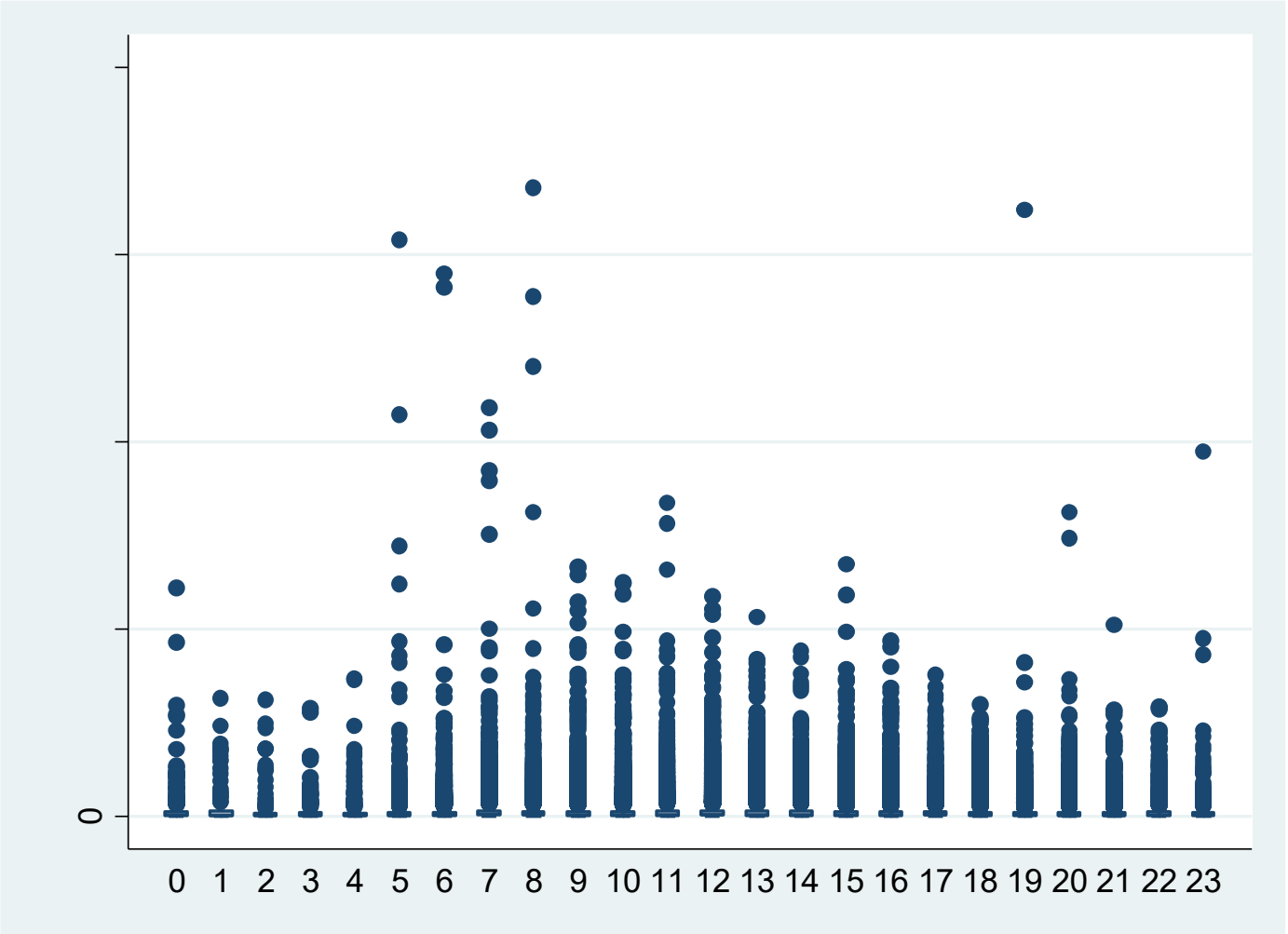
Trip Time Box Plot (y axis in seconds)
Data from 05/23/2022

Super Trips

E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
year	month	day	hour	minute	second	KEY_Statu	newJID	numPoint	eventtime	newPID	Reventtin	RnewPID	triptime	TdistM	Ttimediff	MaxTGap	TTStop
2022	5	23	5	49	58	1	64712	12660	1.97E+12	1	-2E+12	12660	38233	109.2229	0	27	30795
2022	5	23	16	27	11	-1	64712	12660	1.97E+12	12660	-2E+12	1	38233	109.2229	38233	27	30795
2022	5	23	8	1	18	1	111146	12576	1.97E+12	1	-2E+12	12576	38132	40.97071	0	27	33557
2022	5	23	18	36	50	0	111146	12576	1.97E+12	12574	-2E+12	1	38132	40.97071	38132	27	33557
2022	5	23	6	1	10	1	43696	12122	1.97E+12	1	-2E+12	12122	36653	88.75773	0	21	28976
2022	5	23	16	12	3	-1	43696	12122	1.97E+12	12122	-2E+12	1	36653	88.75773	36653	21	28976
2022	5	22	19	59	57	0	9790	11579	1.97E+12	1	-2E+12	11579	34766	7.41105	0	6	32379
2022	5	23	5	39	23	-1	9790	11579	1.97E+12	11579	-2E+12	1	34766	7.41105	34766	6	32379
2022	5	23	6	44	27	1	21816	10452	1.97E+12	1	-2E+12	10452	31559	24.72731	0	42	28268
2022	5	23	15	30	26	0	21816	10452	1.97E+12	10452	-2E+12	1	31559	24.72731	31559	42	28268
2022	5	23	8	6	0	0	95812	10033	1.97E+12	1	-2E+12	10033	30207	15.88133	0	6	27750
2022	5	23	16	29	27	0	95812	10033	1.97E+12	10033	-2E+12	1	30207	15.88133	30207	6	27750
2022	5	23	5	5	43	0	68957	9028	1.97E+12	1	-2E+12	9028	27138	62.27964	0	7	21456
2022	5	23	12	38	1	0	68957	9028	1.97E+12	9028	-2E+12	1	27138	62.27964	27138	7	21456
2022	5	23	7	41	23	1	35175	8603	1.97E+12	1	-2E+12	8603	25848	19.24775	0	7	21841
2022	5	23	14	52	11	-1	35175	8603	1.97E+12	8603	-2E+12	1	25848	19.24775	25848	7	21841
2022	5	23	7	35	7	1	115998	8377	1.97E+12	1	-2E+12	8377	25199	30.75371	0	6	18467
2022	5	23	14	35	6	-1	115998	8377	1.97E+12	8377	-2E+12	1	25199	30.75371	25199	6	18467
2022	5	23	8	33	41	1	57976	8330	1.97E+12	1	-2E+12	8330	25171	13.50814	0	27	24022
2022	5	23	15	33	12	0	57976	8330	1.97E+12	8328	-2E+12	1	25171	13.50814	25171	27	24022
2022	5	23	7	31	39	0	81490	8049	1.97E+12	1	-2E+12	8049	24185	82.21774	0	6	17951
2022	5	23	14	14	44	-1	81490	8049	1.97E+12	8049	-2E+12	1	24185	82.21774	24185	6	17951
2022	5	23	7	36	47	1	77515	7693	1.97E+12	1	-2E+12	7693	23115	16.39346	0	7	20620
2022	5	23	14	2	2	-1	77515	7693	1.97E+12	7693	-2E+12	1	23115	16.39346	23115	7	20620
2022	5	23	0	54	43	1	67332	7602	1.97E+12	1	-2E+12	7602	22800	122.9316	0	6	9294
2022	5	23	7	14	43	-1	67332	7602	1.97E+12	7602	-2E+12	1	22800	122.9316	22800	6	9294
2022	5	23	13	40	33	1	71385	7558	1.97E+12	1	-2E+12	7558	22749	115.0251	0	9	8382

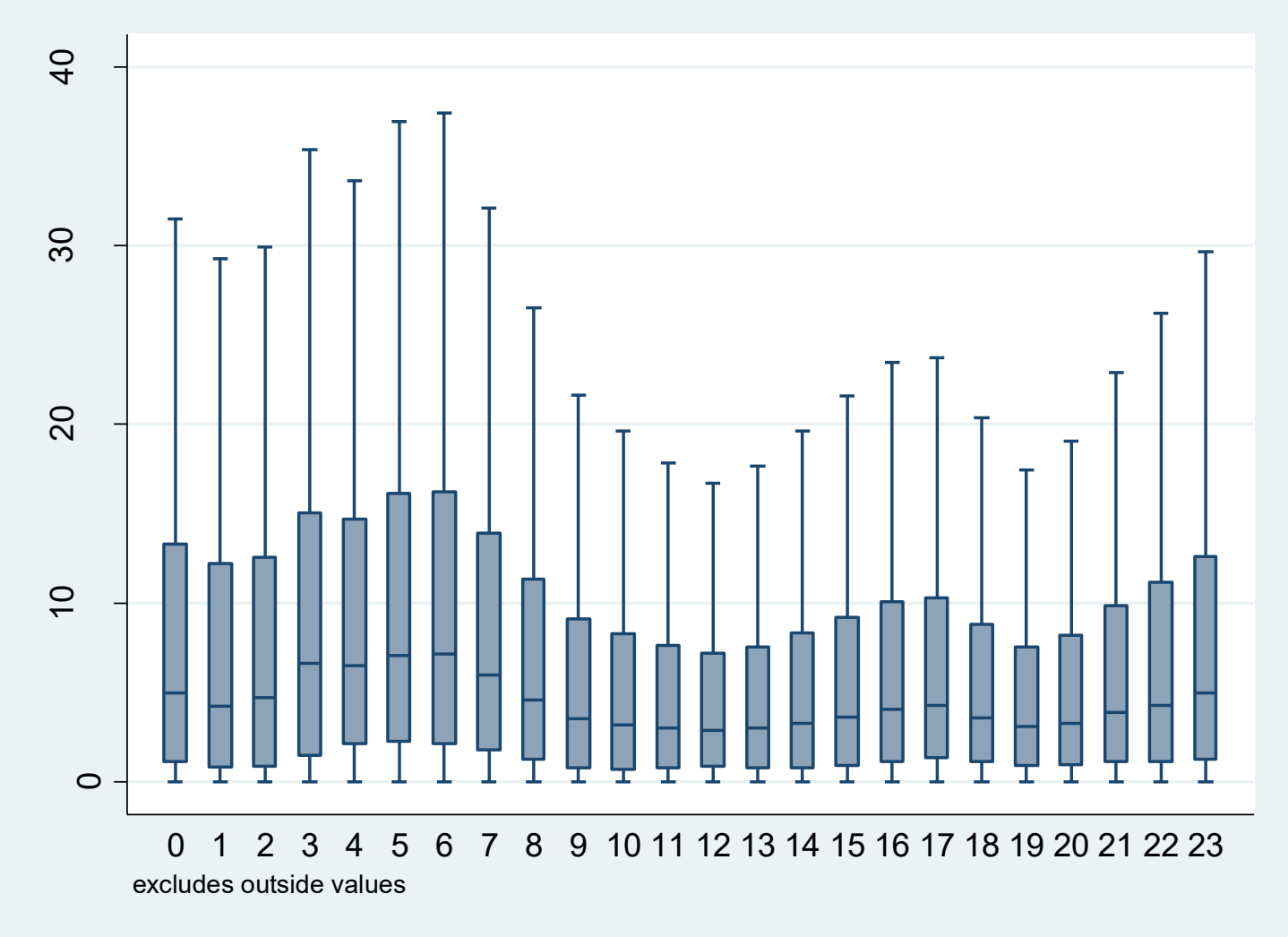
- *The maximum time gap for these trips is low, showing they have continuous trajectory.*
- *The distance is long, but possible.*
- *The total stop time is long. Excluding such time, the trip time is reasonable (about 2h and 4min)*
- *Possible delivery worker/uber driver?*

Total Stop Time within a Trip

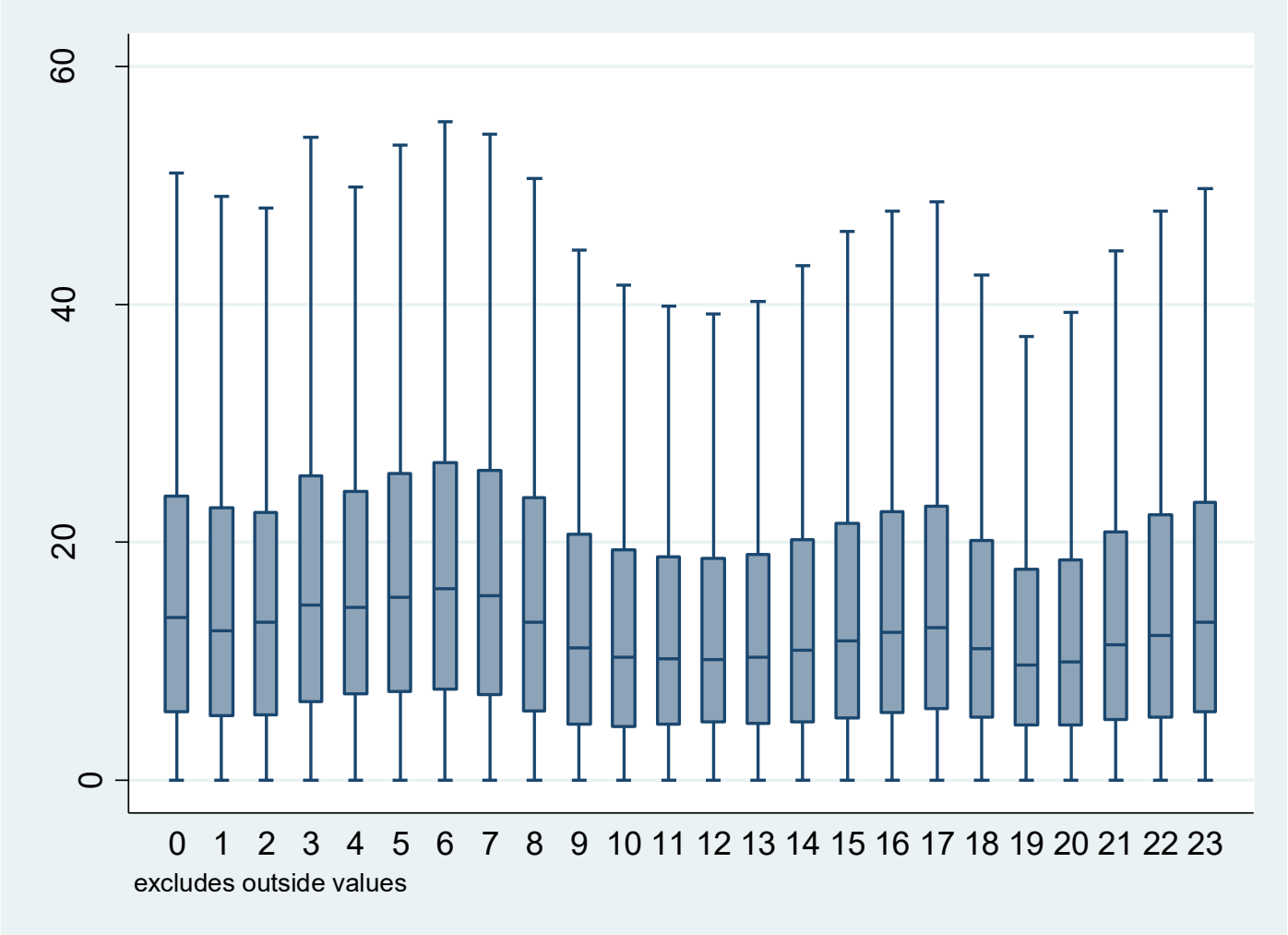


Total Stop Time within a Trip Box Plot (y axis in seconds)
Data from 05/23/2022

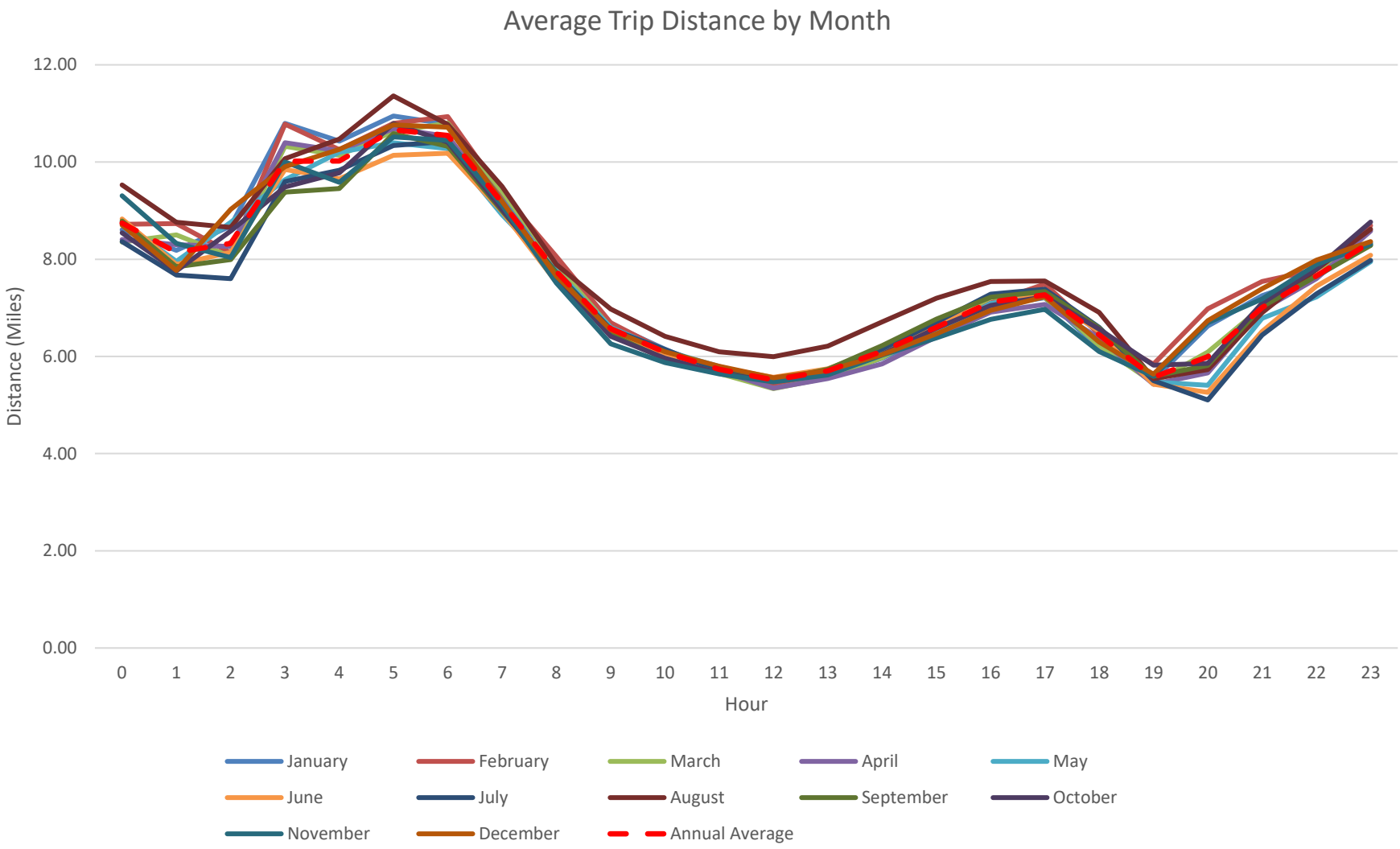
Annual Workday Internal-Internal Trip Distance Distribution



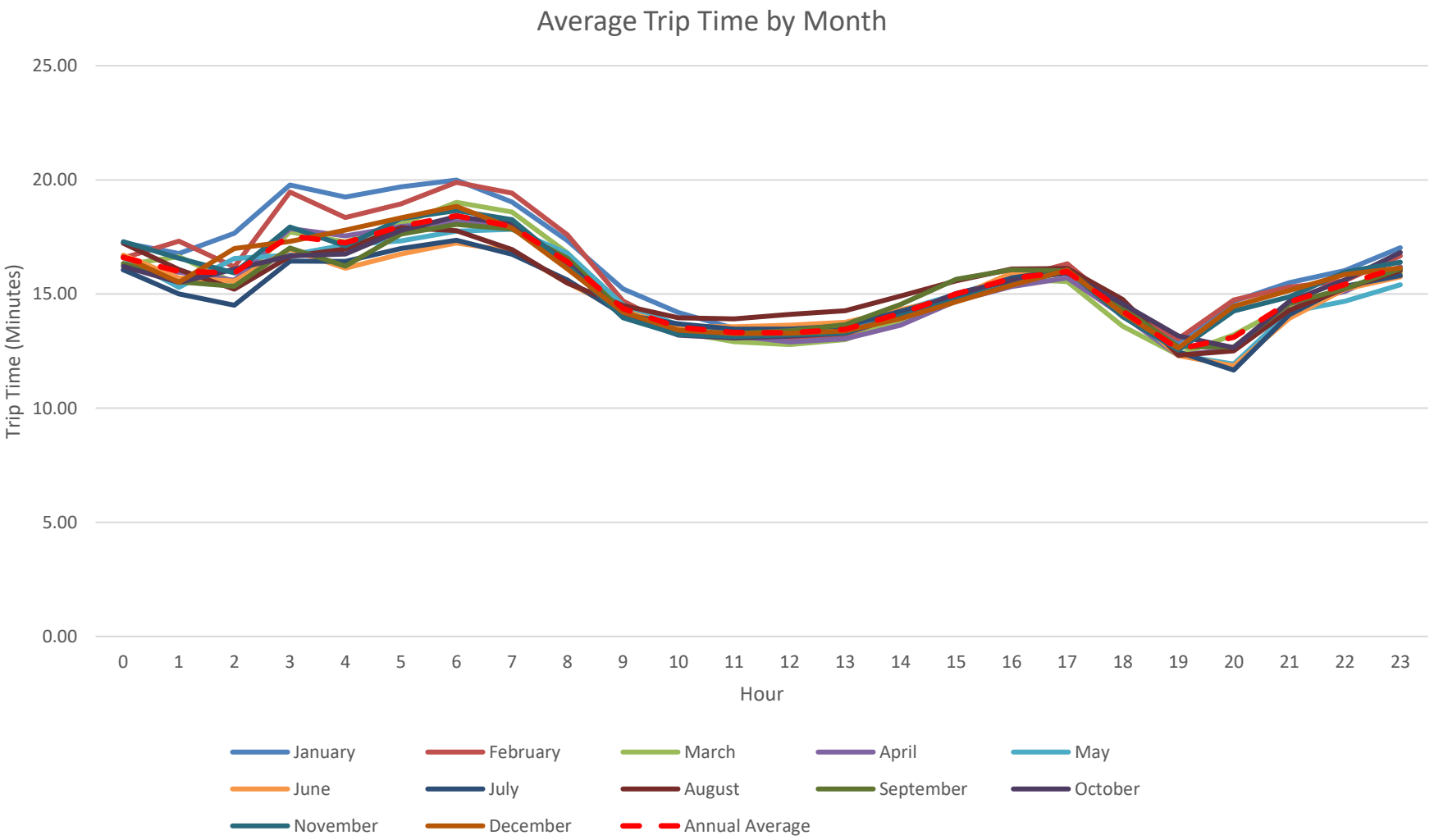
Annual Workday Internal-Internal Trip Time Distribution



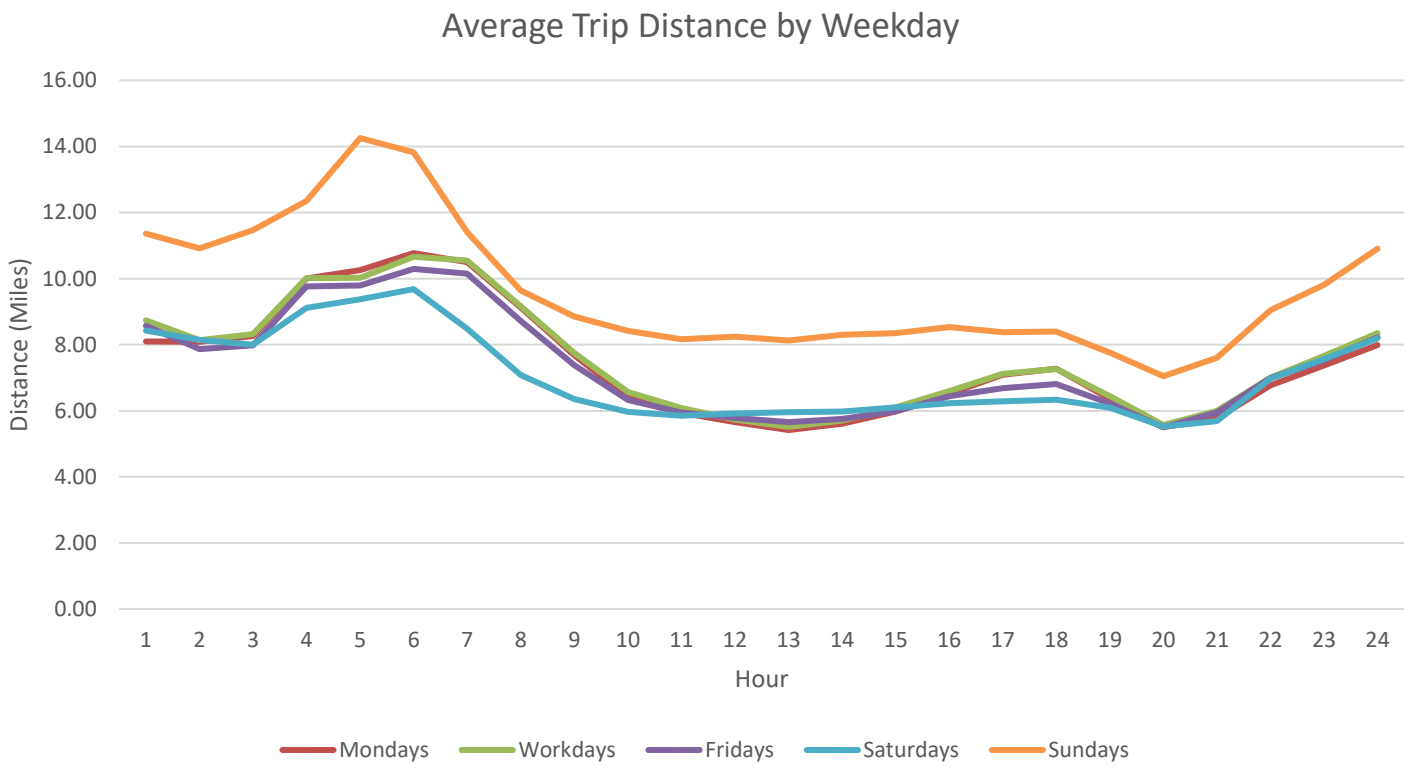
Average Trip Distance- Workdays (Tuesday, Wednesday, Thursday)



Average Trip Time - Workdays (Tuesday, Wednesday, Thursday)

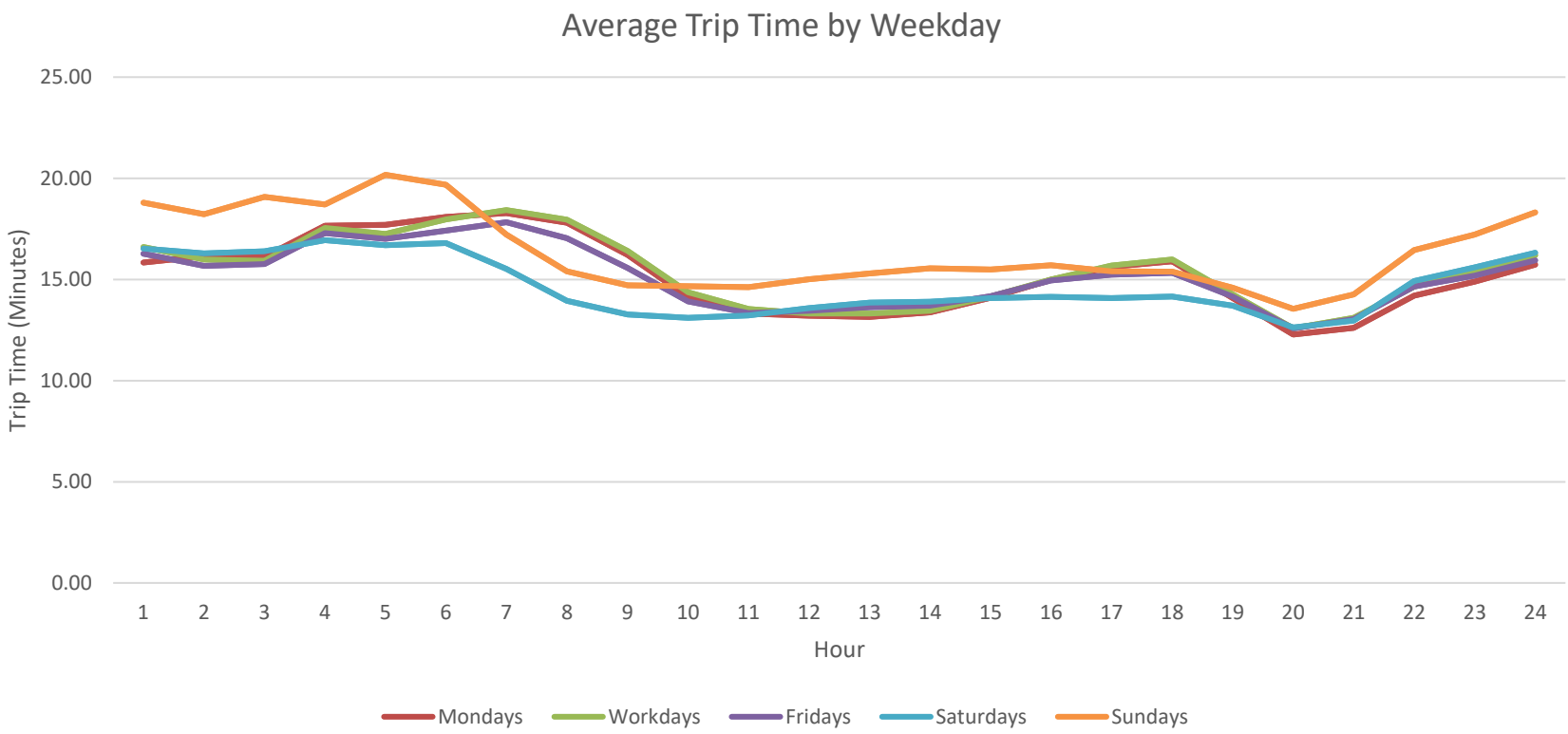


Average Trip Distance Over Different Days of the Week



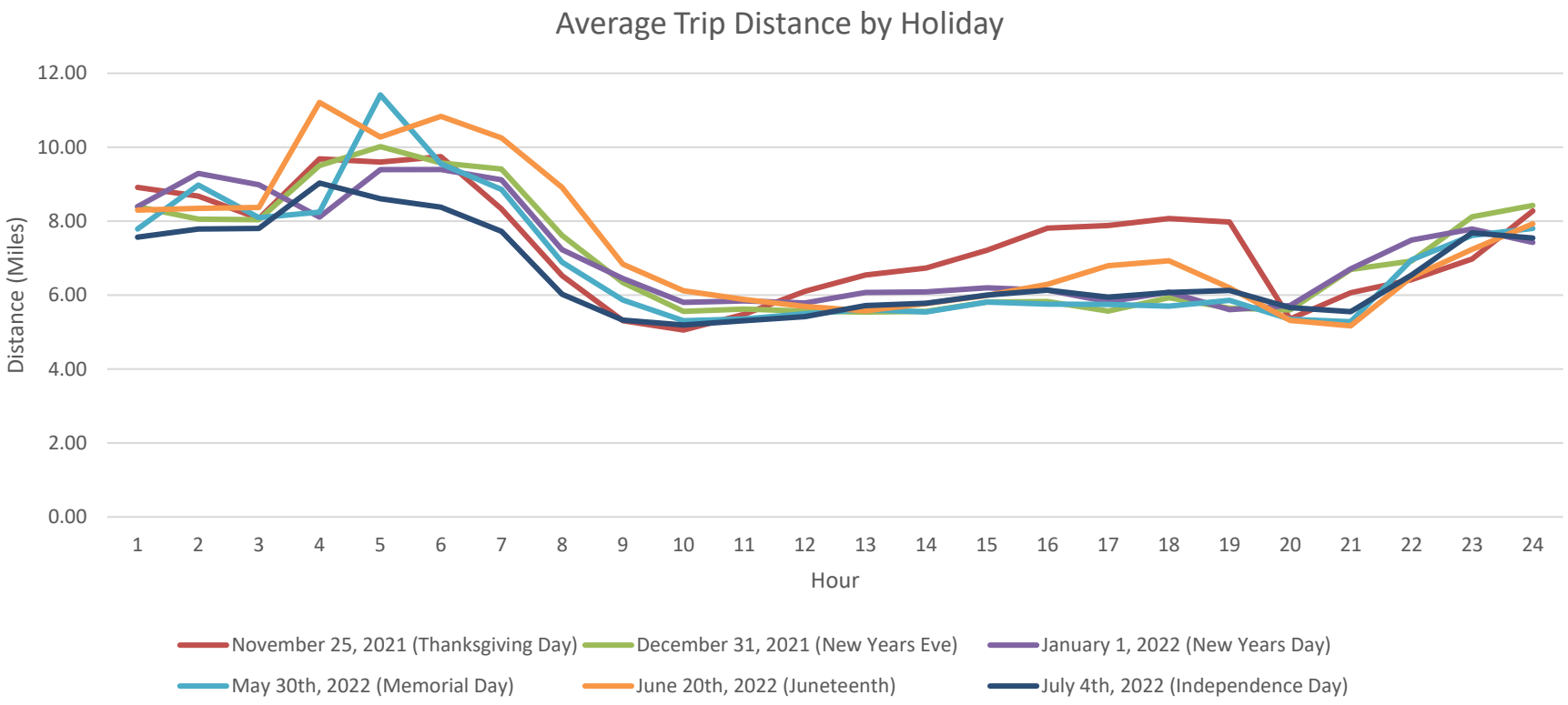
- *Holidays are excluded, but not days adjacent to holidays.*
- *Saturdays and Sundays have different traffic patterns, with Sundays more so than Saturdays.*

Average Trip Time Over Different Days of the Week

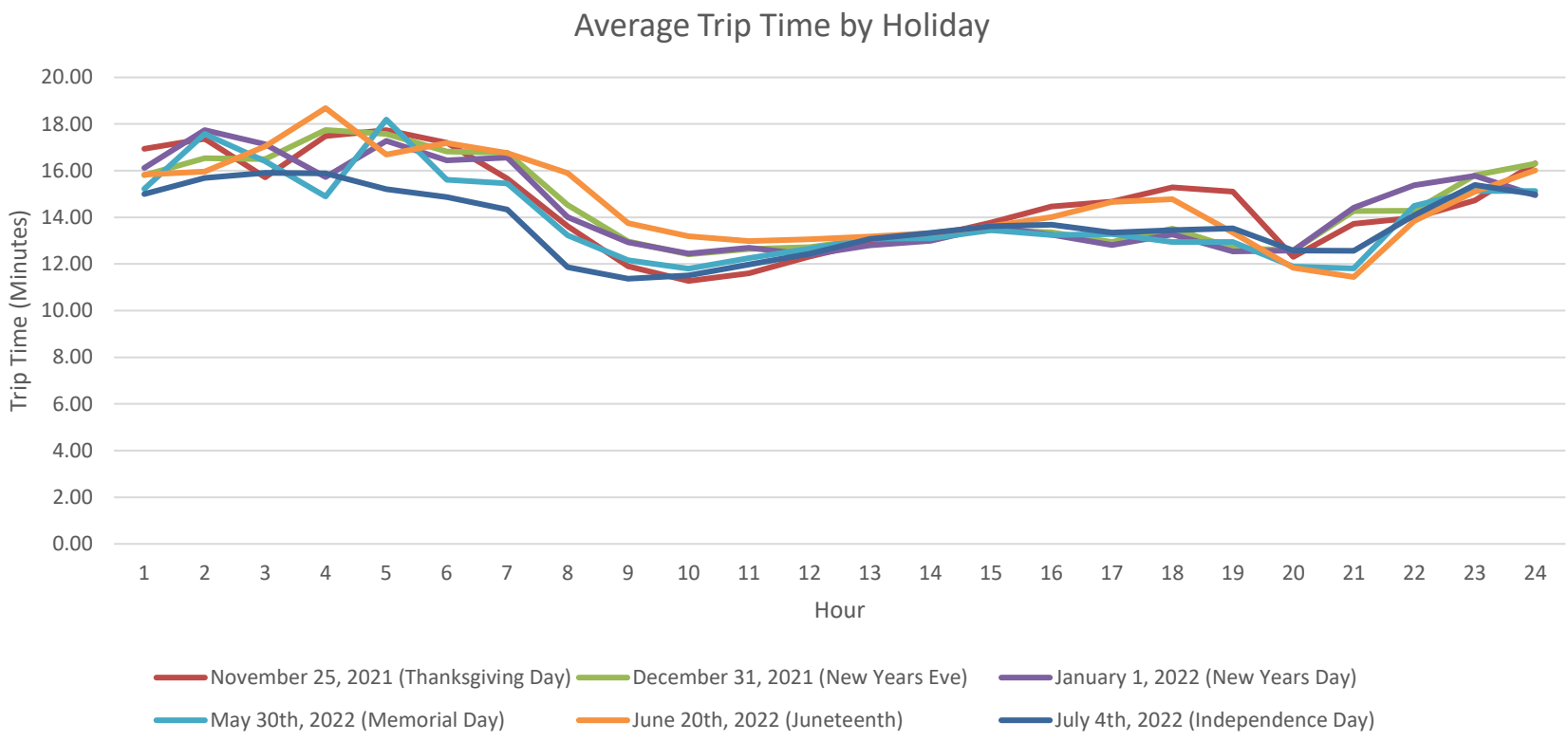


- *Holidays are excluded, but not days adjacent to holidays.*
- *Saturdays and Sundays have different traffic patterns, with Sundays more so than Saturdays.*

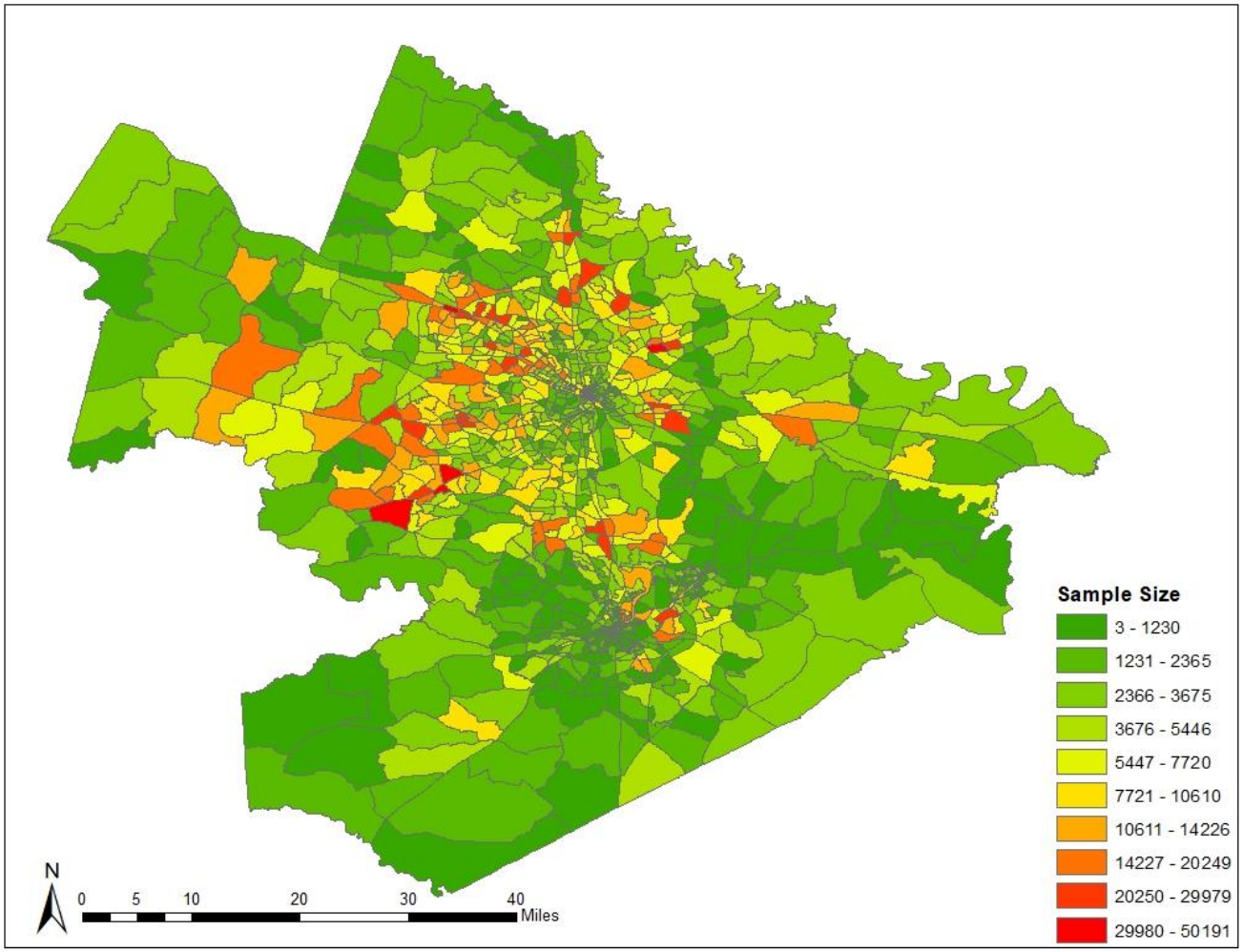
Average Trip Distance – Per Day (Holidays)



Average Trip Time – Per Day (Holidays)

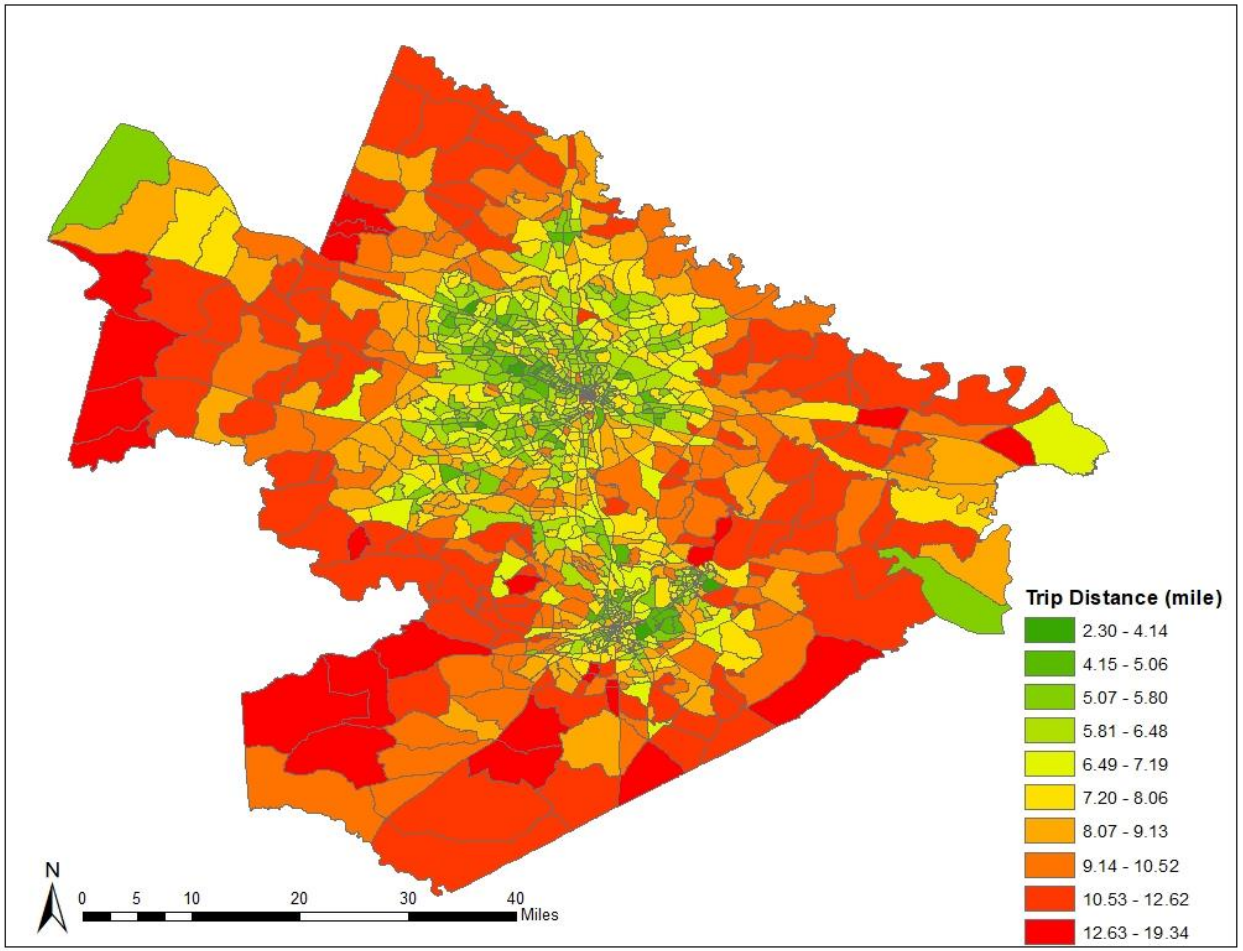


Workday Trips Sample Size by Traffic Analysis Zones (TAZs)



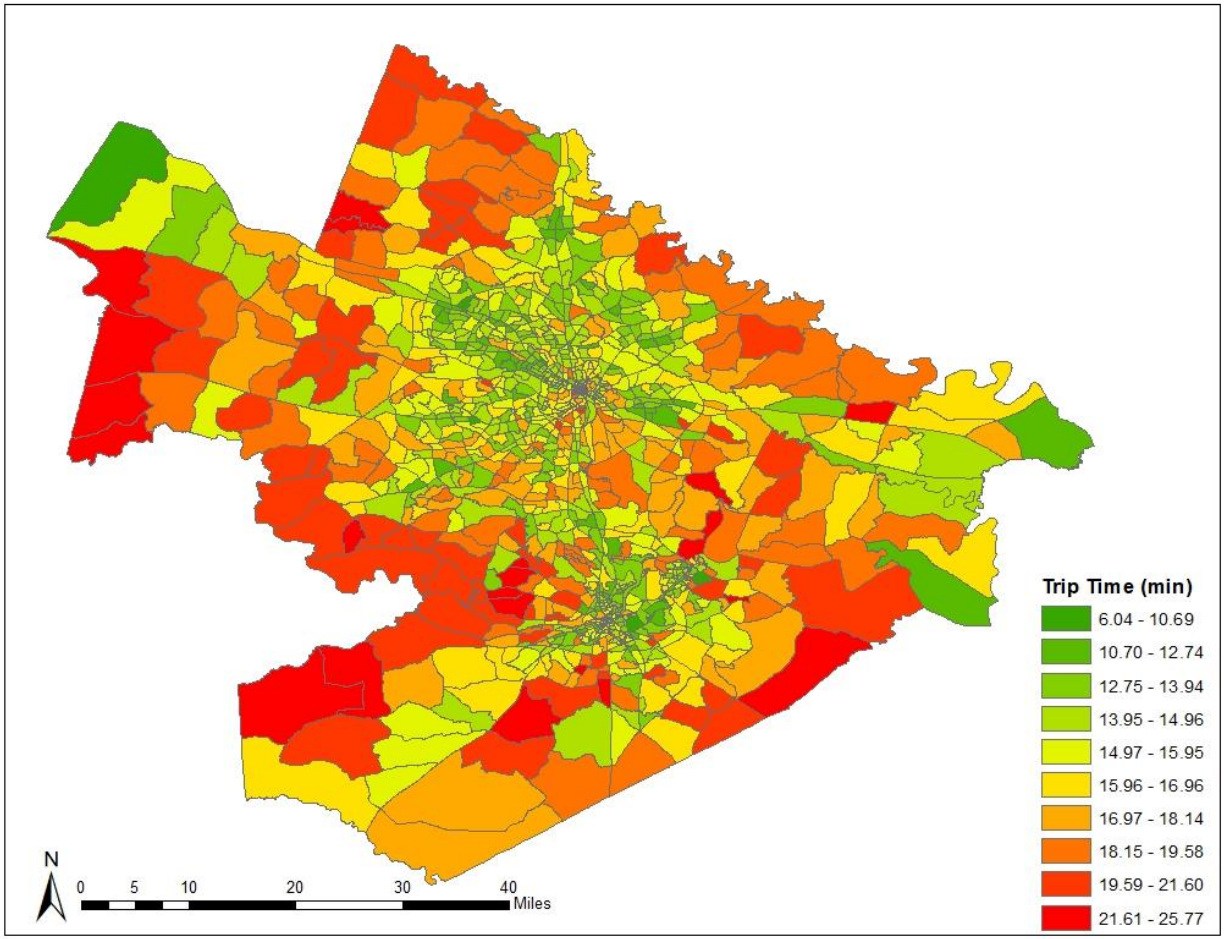
- *Total number of observations by TAZs over 71 Tuesdays, Wednesday, and Thursdays (excluding holidays) during the study period (08/2021-07/2022) based on the trip origin.*

Average Workday Trip Distance by TAZs



- *Average trip distance in miles for internal-internal trips by TAZs over 71 Tuesdays, Wednesday, and Thursdays (excluding holidays) during the study period (08/2021-07/2022)*

Average Workday Trip Time by TAZs



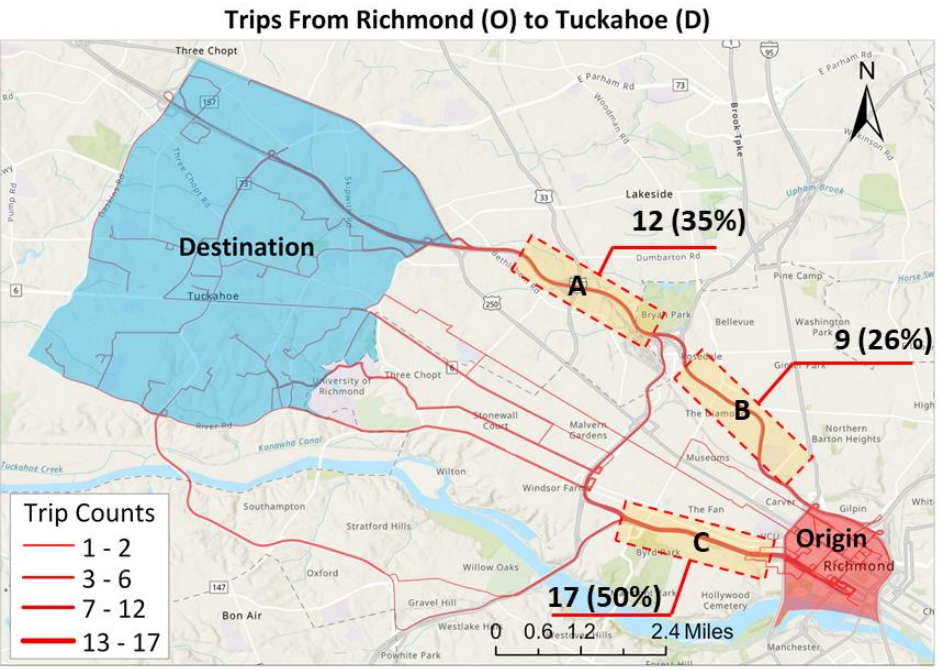
- *Average trip time in minutes for internal-internal trips by TAZs over 71 Tuesdays, Wednesday, and Thursdays (excluding holidays) during the study period (08/2021-07/2022)*

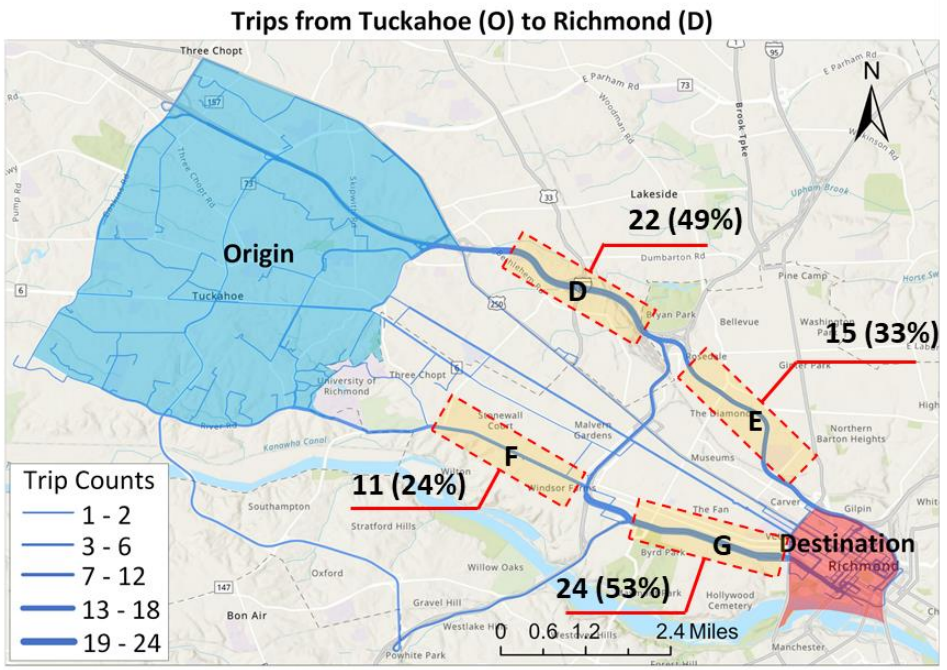
Wejo OD Table using 05/23/2022 Data

Data Set	Processing Steps Applied	Num of Points
OD Points	List of OD points exported from previous projects by eliminating all flawed trips	245,206 points
OD Points Matched to TAZs	Eliminating ODs missing either Engine-On or Engine-Off events, or with either end not matched with a TAZ	191,124 points, or 95,562 trips
OD tables based on all trips	OD tables of Wejo trips by TAZ pairs	43,924 unique OD pairs, or 3.06% of the total
OD tables with diagonal cells removed	OD tables of Wejo trips by TAZ pairs without intra-zonal trips	42,793 unique OD pairs with interzonal trips, or 2.96% of the total

OD matrices are sparse and we do not have to analyze all OD pairs!

Demonstration of Path Identification (Cluster of TAZs)





Total observed trips on this example OD Pair (One Day, May 23rd, 2022):
Richmond to Tuckahoe: 34
Tuckahoe to Richmond: 45

Trip Group	A	B	C
Avg travel time (min.)	18.40	17.58	21.52
T – Shortest T (min.)	+0.82	--	+3.94
Number of Trips	12	9	17
Variance	20.06	21.78	23.36
Welch's t Test		--	*

Trip Group	D	E	F	G
Avg travel time (min.)	28.67	25.87	23.23	26.78
T – Shortest T (min.)	+5.44	+2.64	--	+3.55
Number of Trips	22	15	11	24
Variance	48.00	40.50	47.58	15.59
Welch's t Test	**		--	

Preliminary Conclusions and Ongoing Work

- Analysis in this study based on the CV data provides a rich set of information about the temporal and geographic distribution of vehicle trips. Such information includes the ***OD locations, actual trajectory, non-moving time during a trip, number of stops, total trip distance, and total trip time***. It provides travel demand modelers with updated and nuanced details of traffic patterns, which could complement other data sources and improve traffic demand modeling practice.
- Wejo data captures ***time of day, day of the week, and monthly variation*** in traffic patterns and travel behavior. Such temporal nuance is not available from other data sources like the NHTS data. The richness of the data and the ***up-to-date features*** could help VDOT stakeholders better develop and calibrate travel demand models.
- Based on this study and the literature, Wejo data covers about 2-5% of trips. Wejo data does not provide the social demographic information about the drivers.
- Ongoing Work: Develop a methodology for incorporating revealed path patterns based on CV GPS trajectory data into the trip assignment process as a new path file.
- Potential use case: Calibration of destination choice model; time of day model; route utility function, etc.

Comments and Suggestions

Thank You!

Shanjiang Zhu, Ph.D., Associate Professor

George Mason University

703-993-1797, szhu3@gmu.edu

http://civil.gmu.edu/people/shanjiang-zhu