TDM Validation Activity: Follow-up OD Study

TETC Traffic Data Marketplace Data Validation

Prepared by: TETC Validation Team

7/30/2025

TDM Validation Activity: Follow-up OD Study

TETC Traffic Data Marketplace Data Validation

Part 2 of inaugural OD validation activity, including a comparison of TDM vendor data with NHTS data in Richmond, VA.

TDM-VAL-8

Data Categories: Origin-Destination

7/30/2025

The Eastern Transportation Coalition is a partnership of 19 states and the District of Columbia focused on connecting public agencies across modes of travel to increase safety and efficiency. Additional information on the Coalition, including other project reports, can be found on the Coalition's website: www.tetcoalition.org

Executive Summary

This validation report represents Phase II of the Eastern Transportation Coalition's first Origin-Destination validation activity. Phase I (link to report) focused on evaluating OD data from four TDM vendors: AirSage, Geotab, INRIX, and Streetlight and consisted of two parts: (I) a vendor questionnaire intended to expose important key aspects of each vendor's product, and (II) an analysis of sample data in Richmond, VA concentrating on county-level OD patterns. Phase II builds off the work in Phase I, but expands the datasets used for analysis, adding Replica (an additional vendor that sells OD data through the TDM) and the National Household Travel Survey (NHTS) 'add-on' dataset. Additionally, whereas the Phase I analysis was focused on the full dataset and high-level OD flows between counties, the analysis in this report explores more granular Traffic Analysis Zone (TAZ) level flows for trips within specific areas.

OD data from October 2022 in the Richmond, VA area (city plus 13 surrounding counties) was obtained at the TAZ level from all five vendors and processed to a common format described later in the report. Sample trips from the 2022 NHTS dataset were assigned TAZ origins and destinations based on reported endpoint latitude/longitude values and aggregated to a similar format to facilitate comparison with vendor data. Due to the large number of TAZs across the entire Richmond area, the analysis was limited to trips within individual counties, with the majority of analysis focusing on trips within Chesterfield county, which was selected based on a mixture of urban and rural characteristics. Results were also generated separately for trips within the city of Richmond (urban) and also for Goochland County (rural).

As with the prior report, the focus of this report is to inform Coalition members about OD data offerings in the TETC Transportation Data Marketplace, understand the extent to which commercial offerings agree with one another (and in this case, also how they compare with the NHTS dataset), and introduce comparative methods and visualizations to utilize in future OD studies. Several key takeaways emerged from the analysis:

- **Major differences between vendor product offerings make direct comparisons challenging.** This finding was highlighted in the Phase I report, but is worth re-emphasizing here. Key differences between vendor products include underlying data sources (connected vehicle data vs smart phone location-based services (LBS) data), the types of vehicles captured (e.g., mixture of all vehicles, freight only), how trips are quantified (e.g., person-trips vs vehicle-trips), trip mode, and whether reported trip counts reflect observed sample probes or population-level estimates. Furthermore, each vendor uses slightly different logic to split GPS waypoints into separate trips, so the same GPS trajectory may result in different trip definitions across vendors. These nuances should be taken into consideration when comparing vendor results to each other or to external data sources. As a result, *even when vendor products disagree with each other – or with an external data source—they may still provide meaningful information.*
    - AirSage reports **person-trips** that reflect population estimates.
    - Geotab reports observed **commercial freight vehicle-trips**
    - INRIX reports observed **vehicle-trips** (mixture of passenger and freight CV data)
    - Streetlight reports **vehicle-trips** that reflect population estimates (CV-data).
    - Replica reports both **vehicle-trips** and **person-trips** that reflect population estimates. For this study, only person-trips are included.

- **The NHTS dataset provides a useful point of comparison, particularly for evaluating travel time and trip length distributions.** Many planners are familiar with the NHTS dataset and its characteristics, so although it does not reflect 'ground truth', it acts as a useful comparative dataset. The ability to consider person-trips and vehicle-trips separately, plus the ability to apply weights to represent population-level behavior, helps facilitate more reasonable comparisons with vendor data.

  However, it appears that the NHTS dataset – particularly when using the weighting factors – is best suited for analyzing travel time and trip distance distributions, rather than identifying OD patterns at the granular TAZ level. Although unweighted NHTS data is used to identify OD patterns in this report, the small number of observations at this detailed spatial scale – particularly in rural areas -- makes it difficult to determine the most important O's and D's.

- **Vendor Travel Time and Trip Length distributions appear reasonable and shifted intuitively across counties**. When viewed individually, vendors' travel time and trip length distributions were generally intuitive, with trip lengths tending to include smaller values in the urban context (city of Richmond) and larger values in rural conditions (Goochland county) – an encouraging result.

  Although results all appeared reasonable, the shape of vendor distributions often differed slightly from each other and with respect to the NHTS dataset. These differences also varied across counties. Given that these datasets represent diverse products (as noted in the first point above), such differences are expected.

- **Top OD patterns are difficult to compare at the TAZ level.** Unlike county-level OD flows explored in Phase I (where the same top Origin and Destination counties were common across vendor datasets), top O's and D's were more challenging to distinguish at the more granular TAZ level. Given that there was often little separation in the top trip-producing or attracting zones, the most effective approach for comparison was to check for the percentage agreement in the top 10% origins and destinations for each data source.

  This analysis showed varying levels of overlap in top origins and destinations between different vendors and NHTS data. Across all datasets, StreetLight and INRIX had the highest degree of overlap with each other and to a slightly lesser degree, with Replica. However, differences in top O and D rankings between datasets are expected given the distinct underlying data products and the small size of TAZ geographies.

In summary, although this analysis does not label vendor products as 'sufficiently accurate' or 'inaccurate' – owing to significant differences in underlying data sources and product characteristics, plus the difficulty of establishing 'ground truth' benchmarks -- the differences observed in the TAZ-level OD analysis appear consistent with self-reported data characteristics described in the Phase I report.

Users of the data are encouraged to (1) explore the value and usability of the data as compared to survey data collection mechanisms, and (2) take note of the approach and characteristics of each vendor, including both the source data (person-trips, vehicle-trips, or freight trips), and the various methods of processing (reporting sample trips only, extrapolating to the population, and how trips are defined based on underlying trajectories). These parameters vary by vendor, and in some cases are configurable, so it is important that agencies understand the data characteristics to make sure they are suitable for intended use cases.

Table of Contents

Introduction

Transportation data sold through the Eastern Transportation Coalition (TETC) Transportation Data Marketplace (TDM) is procured from private industry based on contract specifications. The intent of the Coalition's validation program has evolved from the original Vehicle Probe Project validation which was primarily limited to ensuring that traffic data conforms to contractual standards for speed and travel time accuracy. Although the TDM still includes that essential function, it also has the flexibility to adjust to the needs of the Coalition members as the market evolves and data needs expand. The validation process is overseen by a technical advisory committee that sets general direction and reviews results.

The TDM includes both quantitative and qualitative analysis of datasets available through the marketplace as appropriate for each data type. The marketplace currently contains six core data items: Travel Time/Speed, Volume, Waypoint, Origin-Destination, Freight, and Conflation, with all but one (Travel Time/Speed) being sold through the marketplace for the first time. As such, the validation team, under the guidance of the TETC Validation Technical Advisory Committee (TAC), is establishing benchmarks and methods for effectively evaluating data quality and value applicable to the different data sets.

This validation report represents Phase II of the inaugural OD validation activity. Phase I (link to report), consisted of two parts: (i) a vendor questionnaire intended to expose important key aspects of each vendor's product, and (ii) an analysis of sample data provided by each vendor in Richmond, VA. The latter focused on how to access and interpret vendor datasets, a descriptive analysis of the OD datasets, and preliminary comparative analysis of results across vendors. The study area was chosen to coincide with other complementary datasets owned and managed by Virginia Department of Transportation (VDOT) to support follow-up analysis.

This report builds off the work in Phase I and uses the same vendor OD data for analysis, focusing on TAZ-level geographic zones in certain counties in Virginia during the month of October 2022. The focus of this report is to drill down to a more granular level of analysis and incorporate a supplemental dataset, the Next Generation National Household Travel Survey dataset. Although the NHTS data cannot be considered a true "ground truth" dataset, it is a widely used source of personal and household travel in the US. As such, it is instructive to understand how TDM OD vendor data compares.

As with the prior report, the focus of this report is to inform Coalition members to better understand OD data offerings, understand the extent to which commercial offerings show internal agreement and agreement with the NHTS source, and introduce comparative methods and visualization to utilize in future OD studies.

Data Vendors

All four vendors that were selected through the TDM RFP process in the Origin Destination (OD) category participated in the study: AirSage, Geotab, INRIX, Streetlight. One additional vendor, Replica, was also included, as their OD datasets are sold as ancillary products through prime vendors in the TDM. Each vendor was highly engaged in the process and assisted with obtaining and interpreting the sample data for the analysis.

Geographic Scope

Custom Traffic Analysis Zones (TAZs) were chosen as the spatial geography for the sample data request and were selected to (1) be small enough to investigate granular trip patterns and (2) align with the zonal structure of Virginia DOT's travel demand model in Richmond. Figure 1 shows the 1203 TAZs near Richmond used in the analysis.

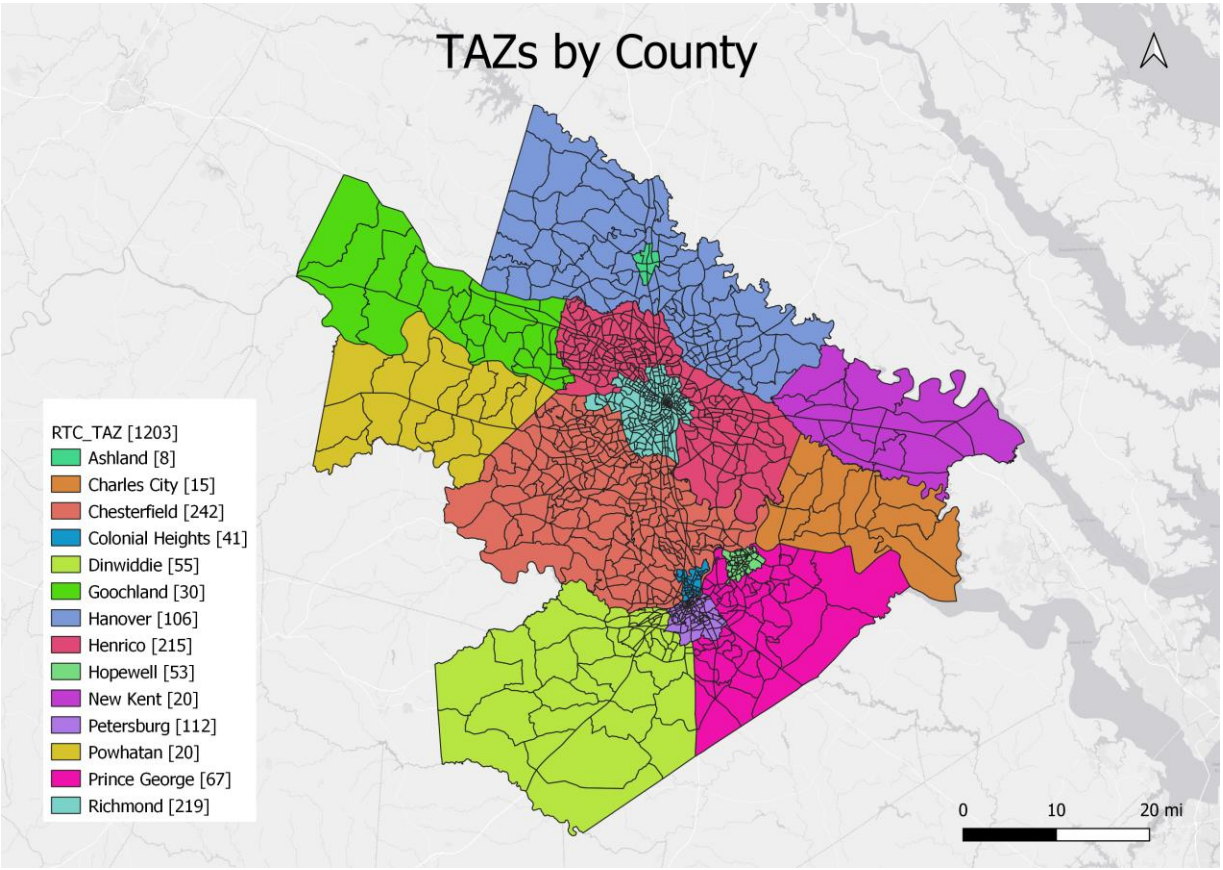

Figure 1. Model TAZs near Richmond, VA, color coded by county.

Whereas the previous report considered this entire study area (focusing on flows between counties), this analysis focuses on the more granular Traffic Analysis Zones for three areas: Chesterfield County, the city of Richmond, and Goochland County. Chesterfield county was chosen for both its importance in the Richmond area and diverse land uses (encompassing both urban and rural locations), while Richmond City and Goochland County were selected to represent urban and rural locations, respectively.

Data Collection

Data used in this analysis includes samples submitted by the five TDM vendors that participated in Part I of the study, plus an additional data source, the 2022 Next Gen National Household Travel Survey (Next Gen NHTS). Although the NHTS dataset is not considered "ground truth", it serves as an important point of comparison.

TDM Vendors

OD data was collected from all five vendors for the study area for the month of October 2022. The interested reader should consult the Phase I report for details about how each vendor's data was accessed and processed to a common format.

Each vendor's dataset was standardized into a common format to support comparative plotting and analysis routines. This standardization process looked slightly different for each vendor; for some vendors it only required renaming columns, while others involved temporal aggregation or turning "total trip" counts into average daily values to make the datasets as comparable as possible. The final standardized data format includes the following fields:

- **vendor_name:** Name of vendor
- **origin_zone:** Origin zone name (corresponds to shapefile model TAZ id)
- **dest_zone:** Destination zone name
- **veh_class**: Vehicle classification (optional, vendor-specific)
- **trip_purpose:** Trip purpose (optional, vendor-specific e.g., HBW, HBO)
- **day_type:** Day type aggregation.
    - *Typical Weekday (Tue-Thu, Non-Holiday)*
    - *Weekend (Sat-Sun)*
    - *All (Mon-Sun)*
- **day_part**: Day part aggregation
    - *AM Peak (6-9am)*
    - *PM Peak (3:30-6:30pm)*
    - *All* day
    - *Off peak (6:30 pm – 6 am)*
- **numtrips_obs:** Number of daily **average observed trips** between zones for specified day_type and day_part. (Not reported by all vendors)
- **numtrips_popest**: Number of daily **average estimated trips** between zones at the population level for specified day_type and day_part. (Not reported by all vendors)
- **avg_tt:** Average travel time in minutes
- **avg_dist**: Average trip distance in miles

Interpretation Considerations

Even after standardizing, it is important to note that the resulting OD patterns and corresponding trip statistics are not directly comparable; each vendor is reporting different types of trips. The following bullets summarize what each vendor is reporting *for this study* (note that they may have additional offerings in general, but this is what was submitted for evaluation).

- AirSage reports **person-trips** that reflect population estimates.

- Geotab reports observed **commercial freight vehicle-trips**, including diverse fleet types (derived from its own fleet management telemetry data).
- INRIX reports observed **vehicle-trips** (mixture of passenger and freight-based CV data)
- Streetlight reports **vehicle-trips** that reflect population estimates (based on CV-data).
- Replica reports both **vehicle-trips** and **person-trips** that reflect population estimates. For this study, only person-trips are included.

NHTS Dataset

The National Household Travel Survey (NHTS), overseen by the U.S. Department of Transportation, collects information on both daily and long-distance travel by U.S. households. Conducted periodically since 1969, typically every five to eight years, the survey tracks shifts in travel behavior, transportation methods, and trip purposes. Across its nine iterations, from 1969 to 2022, the NHTS has provided valuable insights into how travel patterns have changed alongside demographic, economic, and cultural trends in the U.S.

The 2022 NHTS data offers a detailed look at the travel behaviors and demographics of people and households in the United States. It records all trips taken within a 24-hour timeframe by household members aged 5 or older, encompassing a variety of household types and travel activities, while excluding group homes with 10 or more unrelated individuals such as prisons and dormitories. The 24-hour travel window spans from 4 AM on the assigned day to 4 AM the following day. Participants report the purpose of each trip, the mode of transportation, travel duration, day of the week, and vehicle occupancy. This data is integrated with information on household characteristics, vehicles, demographics, and socio-economic factors to provide comprehensive travel records.

Although the NHTS is a national survey, the several 'add-on' studies were contracted by state DOTs/MPOs – including Virginia DOT. This add-on survey resulted in a localized version of the NHTS intended to meet planning requirements. In the context of this validation study, the subset of the dataset that aligns with the Richmond metro areas was used for analysis. *However, it should be emphasized that **the NHTS dataset encompasses all of 2022**, in contrast to the vendor data that focuses on just October 2022.*

Data Schema

The 2022 NHTS data consists of four hierarchical data files: HOUSEHOLD, PERSON, VEHICLE, and TRIP, available in CSV, SAS, and SPSS formats. These files can be integrated using common ID variables. The HOUSEHOLD file, which includes primary details about the household such as size and demographics, serves as the main key for merging all files with its unique identifier, HOUSEID. The VEHICLE file provides details about household vehicles, including make and model. The PERSON file contains information about individual household members, such as age and demographic details. The TRIP file documents trips taken by household members, including start and end times and destinations. Figures 2-5 provide screenshots of data dictionary of each file provided by the 2022 NHTS.

Of particular interest is the Trips file shown in Figure 5, which, among other attributes, contains the start/end location of each trip at several geographical levels such as state, city, zip code, and even latitude and longitude, the travel time and trip distance, start/end time of each trip, amount paid for parking, reason for travel to the destination, transit mode trips used on the trip, number of individuals on the same trips and so on.

NAME	TYPE	LENGTH	LABEL	HH	PER	VEH	TRIP
CNTTDHH	N	8	Count of household trips on travel day	H			
CONFIRMEDHOME_ADDR1	C	46	Confirmed home address line 1	H			
CONFIRMEDHOME_CITY	C	20	Confirmed home address city	H			
CONFIRMEDHOME_FORMATTEDADDR	C	88	Confirmed home formatted address	H			
CONFIRMEDHOME_LATITUDE	N	8	Confirmed home latitude	H			
CONFIRMEDHOME_LOCNAME	C	4	Confirmed home location name	H			
CONFIRMEDHOME_LOCTYPE	C	2	Confirmed home location type	H			
CONFIRMEDHOME_LONGITUDE	N	8	Confirmed home longitude	H			
CONFIRMEDHOME_STATE	C	2	Confirmed home address state	H			
CONFIRMEDHOME_STREETADDR	C	62	Confirmed home address street	H			
CONFIRMEDHOME_ZIP	C	10	Confirmed home zip code	H			
DRVRCNT	N	8	Number of drivers in the household	H			T
FLAG100	C	2	All HH members completed interview?	H			
HHFAMINC	C	2	Household income	H			
HHFAMINC_IMP	C	2	Household income (imputed)	H			
HHRELATD	C	2	Flag indicating at least 2 persons in HH are related	H			
HHRESP	C	2	Person ID of Household respondent	H			
HHSIZE	N	8	Total number of people in household	H	P	V	T
HHVEHCNT	N	8	Total number of vehicles in household	H		V	
HOMEOWN	C	2	Whether home owned or rented	H			
HOMETYPE	C	2	Type of home	H			
HOUSEID	C	10	Unique Identifier- Household	H	P	V	T
IS_MAIL	C	1	IS_MAIL	H	P	V	T
NUMADLT	N	8	Count of adult household members at least 18 years old	H			
PPT517	N	8	Count of household members 5-17 years old	H			
SMPLSRCE	C	2	Sample where the case originated	H	P	V	T
STRATUM_ID	C	4	Stratum ID of the household	H			
TDAYDATE	C	6	Date of travel day (YYYYMM)	H			
TRAVDAY	C	2	Travel day - day of week	H			T
WRKCOUNT	N	8	Count of workers in household	H			
WTHHFIN2D_VA	N	8	2 day VA household weight	H		V	
WTHHFIN5D_VA	N	8	5 day VA household weight	H		V	
WTHHFIN_VA	N	8	7 day VA household weight	H		V	
YOUNGCHILD	N	8	Count of household members under 5 years old	H			
HHSTFIPS	N	2	Household geo-confirmed state FIPS	H			
HHCNTYFP	N	3	Household geo-confirmed county FIPS	H			
HHCT	N	6	Household geo-confirmed census tract FIPS	H			
HHBG	N	2	Household geo-confirmed block group FIPS	H			
HHGEOID	C	15	Household geographic ID (state, county, census tract, block group)	H			
HH_ZCTA5CE20	N	5	Household geo-confirmed zip code	H			

Figure 2. Attributes provided by the 2022 NHTS in the "HOUSEHOLD" file.

NAME	TYPE	LENGTH	LABEL	HH	PER	VEH	TRIP
HHSIZE	N	8	Total number of people in household	H	P	V	T
HOUSEID	C	10	Unique Identifier- Household	H	P	V	T
IS_MAIL	C	1	IS_MAIL	H	P	V	T
LAST30_BIKE	C	2	Used bicycle in last 30 days		P		
LAST30_BKSHR	C	2	Used bike share in last 30 days		P		
LAST30_ESCT	C	2	Used e-scooters in last 30 days		P		
LAST30_MTRC	C	2	Used motorcycle in last 30 days		P		
LAST30_PT	C	2	Used public transit in last 30 days		P		
LAST30_RDSHR	C	2	Used rideshare in last 30 days		P		
LAST30_TAXI	C	2	Used taxi service in last 30 days		P		
LAST30_WALK	C	2	Walked from place to place in last 30 days		P		
MCA8_O	C	111	Uses something else - specify		P		
MCA8_OS	C	2	Uses something else		P		
MCTRANSIT	N	8	Times in past month motorcycle used		P		
MEDCOND	C	2	Condition or disability that makes travel difficult		P		
MEDCOND6	C	2	Length of time respondent has had condition		P		
PARK	C	2	Paid for parking at any time during diary day		P		T
PARKHOME	C	2	Pay for home parking		P		
PARKHOMEAMT	C	4	Whether respondent pays to park at home		P		
PARKHOMEAMT_PAMOUNT	N	8	Cost of parking at home		P		
PARKHOMEAMT_PAYTYPE	C	2	Duration of payment		P		
PAYPROF	C	2	Worked for pay last week		P		
PERSONID	C	2	Person ID within household		P		T
PFLAG	C	2	Indicator of whether person completed survey		P		
PRMACT	C	2	Primary activity for those who did not work for pay last week		P		
PROXY	C	2	Diary completed by self or someone else		P	V	T
PTUSED	N	8	Times in past month public transit used		P		
RACE_1	C	2	Person 5 or older - Race - White		P		
RACE_2	C	2	Person 5 or older -Race - Black or African American		P		
RACE_3	C	2	Person 5 or older - Race - Asian		P		
RACE_4	C	2	Person 5 or older - race - American Indian/Alaska native		P		
RACE_5	C	2	Person 5 or older - Race - Hawaiian Native/Pacific Islander		P		
RACE_O	C	253	Other race for Person 5 or older - other - specify		P		
RACE_SE	C	2	Person 5 or older - race - Other		P		
RET_AMZ	N	8	Number of times returned online purchase at Amazon dropoff center		P		
RET_HOME	N	8	Number of times returned online purchase by home pickup		P		
RET_PUF	N	8	Number of times returned online purchase to post office/UPS/Fed Ex/ similar		P		

Figure 3. A subset of attributes provided by the 2022 NHTS in the "PERSON" file.

NAME	TYPE	LENGTH	LABEL	HH	PER	VEH	TRIP
ANNMILES	N	8	Self-reported annualized mile estimate			V	
COMMERCIALFREQ	C	2	Over past 30 days, how many days was vehicle used for business purposes?			V	
FUELTYPE_O	C	35	Other fuel type used - specify			V	
HHSIZE	N	8	Total number of people in household	H	P	V	T
HHVEHCNT	N	8	Total number of vehicles in household	H		V	
HHVEHUSETIME_DEL	C	2	Over past 30 days, how many days was vehicle used for deliveries?			V	
HHVEHUSETIME_OTH	C	2	Over past 30 days, how many days was vehicle used for other business?			V	
HHVEHUSETIME_RS	C	2	Over past 30 days, how many days was vehicle used for rideshare?			V	
HOUSEID	C	10	Unique Identifier- Household	H	P	V	T
HYBRID	C	2	Hybrid vehicle			V	
IS_MAIL	C	1	IS_MAIL	H	P	V	T
MAKE	C	2	Vehicle make ID			V	
MAKENAME	C	26	Vehicle make			V	
MODEL	C	5	Vehicle model ID			V	
MODLNAME	C	91	Vehicle model			V	
PROXY	C	2	Diary completed by self or someone else		P	V	T
SMPLSRCE	C	2	Sample where the case originated	H	P	V	T
VEHAGE	N	8	Age of vehicle, based on model year			V	
VEHCOMMERCIAL	C	2	Vehicle used for business purposes			V	
VEHCOM_DEL	C	2	Vehicle used for delivery service			V	
VEHCOM_OTH	C	2	Vehicile used for other business purposes			V	
VEHCOM_RS	C	2	Vehicle used for rideshare			V	
VEHFUEL	C	2	Type of fuel vehicle runs on			V	
VEHID	C	2	Vehicle ID of vehicle used from household roster			V	T
VEHMILES	N	8	Miles put on vehicle in past year			V	
VEHOWNED	C	2	Vehicle owned for 1 year or more			V	
VEHOWNMO	C	2	Vehicles owned less than 1 year - months owned			V	
VEHTYPE	C	2	Vehicle type			V	
VEHTYPE_O	C	53	Other vehicle type - specify			V	
VEHYEAR	N	8	Vehicle year			V	
WHOMAIN	C	2	Main driver of vehicle			V	
WTHHFIN2D_VA	N	8	2 day VA household weight	H		V	
WTHHFIN5D_VA	N	8	5 day VA household weight	H		V	
WTHHFIN_VA	N	8	7 day VA household weight	H		V	

Figure 4. Attributes provided by the 2022 NHTS in the "VEHICLE" file.

NAME	TYPE	LENGTH	LABEL	HH	PER	VEH	TRIP	V
ORIGIN_CITY	C	22	Trip origin city				T	
ORIGIN_COUNTRY	C	24	Country where trip originated				T	
ORIGIN_FORMATTEDADDR	C	168	Formatted origin address				T	
ORIGIN_LAT	N	8	Origin latitude				T	
ORIGIN_LOC	C	144	Origin address				T	
ORIGIN_LOCNAME	C	6	Origin location name				T	
ORIGIN_LOCTYPE	C	2	Origin location type				T	
ORIGIN_LON	N	8	Origin longitude				T	
ORIGIN_MAPPINDROP	C	2	Marker indicating whether pin drop was used				T	
ORIGIN_STATE	C	26	Trip origin state				T	
ORIGIN_STREETADDR	C	146	Trip origin street address				T	
ORIGIN_ZIP	C	5	Trip origin Zip				T	
PARK	C	2	Paid for parking at any time during diary day		P		T	
PARK2	C	4	Paid for parking				T	
PARK2_PAMOUNT	N	8	Amount paid for parking				T	
PARK2_PAYTYPE	C	2	Periodicity of parking payment				T	
PERSONID	C	2	Person ID within household		P		T	
PROXY	C	2	Diary completed by self or someone else		P	V	T	
PSGR_FLG	C	2	Flag for passenger on trip				T	
PUBTRANS	C	2	Used public transit on trip				T	
R_AGE	N	8	Respondent age		P		T	
SMPLSRCE	C	2	Sample where the case originated	H	P	V	T	
STRTTIME	C	4	24 hour local start time of trip				T	S
TRAVDAY	C	2	Travel day - day of week	H			T	
TRIPID	C	2	Trip ID for each trip a person took				T	
TRPACCMP	N	8	Number of people on trip				T	
TRPHHACC_98	C	2	No household members on trip				T	
TRPHHVEH	C	2	Household vehicle used for trip				T	
TRPMILES	N	8	Calculated Trip distance converted into miles				T	
TRPTRANS	C	2	Transit mode used on trip				T	
TRPTRANS_O	C	64	Other transit mode used on trip - specify				T	
TRVLCMIN	N	8	Trip Duration in Minutes				T	TI
VEHID	C	2	Vehicle ID of vehicle used from household roster			V	T	
WALK	C	2	Minutes walked from parking to destination				T	
WHODROVE	C	2	Person who drove on trip				T	
WHYFROM	C	2	Reason for previous trip				T	
WHYTO	C	2	Reason for travel to destination				T	
WHYTO_O	C	545	Reason for travel to destination: Other				T	
WTTRDFIN2D_VA	N	8	2 day VA trip weight				T	
WTTRDFIN5D_VA	N	8	5 day VA trip weight				T	
WTTRDFIN_VA	N	8	7 day VA trip weight				T	
ORIG_ST	N	2	Trip origin geo-confirmed state FIPS				T	
ORIG_CNTY	N	3	Trip origin geo-confirmed county FIPS				T	
ORIG_CT	N	6	Trip origin geo-confirmed census tract FIPS				T	
ORIG_BG	N	2	Trip origin geo-confirmed block group FIPS				T	
ORG_GEOID	C	15	Trip origin geo-confirmed ID (state, county, census tract, block group)				T	
ORIG_ZCTA5CE20	N	5	Trip origin geo-confirmed zip code				T	
ORIG_COUNTRY	C	24	Trip origin geo-confirmed country				T	
DEST_ST	N	2	Trip destination geo-confirmed state FIPS				T	
DEST_CNTY	N	3	Trip destination geo-confirmed county FIPS				T	
DEST_CT	N	6	Trip destination geo-confirmed census tract FIPS				T	
DEST_BG	N	2	Trip destination geo-confirmed block group FIPS				T	
DEST_GEOID	C	15	Trip destination geo-confirmed ID (state, county, census tract, block group)				T	
DEST_ZCTA5CE20	N	5	Trip destination geo-confirmed zip code				T	
DEST_COUNTRY	C	24	Trip destination geo-confirmed country				T	

Figure 5. Some of the attributes provided by the 2022 NHTS in the "TRIP" file.

Weighting

The 2022 NHTS data includes various weights to ensure accurate analysis across different contexts. These weights are divided into categories for households, individuals, and trips, and are available in Seven-day, Five-day, and Two-day versions. Seven-day weights are used for annual travel estimates, while Five-day and Two-day weights are intended for weekday and weekend travel estimates, respectively. Household weights are applied to household-level data, such as vehicle ownership and the number of drivers. Vehicle weights match household weights and are used for vehicle-specific estimates. Trip weights are used to estimate trip numbers and travel distances, especially for vehicle trips categorized by purpose. Person weights are used for individual-level data, including non-household and non-travel day items like the number of workers by gender or drivers by annual mileage.

In the context of this validation study, weights were applied to the Trips file to create several different OD datasets. Figure 6 shows a sample of trips with start/end location, an indicator of being a driver on the trip or not, transit mode used on the trip, and 2-day, 5-day, and 7-day trip weights.

ORIGIN_LAT	ORIGIN_LON	DESTINATION_LAT	DESTINATION_LON	DRVR_FLG	TRPTRANS	WTTRDFIN_VA	WTTRDFIN5D_VA	WTTRDFIN2D_VA
37.71369	-75.7443	37.71202	-75.75538	-1	20	21648.43283	26357.16768	0
37.71202	-75.75538	37.71369	-75.7443	-1	20	21648.43283	26357.16768	0
37.71369	-75.7443	37.70141	-75.71498	1	3	21648.43283	26357.16768	0
37.70141	-75.71498	37.69423	-75.72118	1	3	21648.43283	26357.16768	0
37.69423	-75.72118	37.71369	-75.7443	1	3	21648.43283	26357.16768	0
37.76094	-75.69768	38.06269	-75.54993	1	1	27995.85738	24352.04016	0
38.06269	-75.54993	37.76094	-75.69768	1	1	27995.85738	24352.04016	0
37.71615	-75.77052	37.71615	-75.77052	-1	20	34093.79458	0	43988.75959
37.9337	-75.37366	37.93308	-75.36613	1	1	32834.07641	0	24058.42548
37.93308	-75.36613	37.92793	-75.35547	1	1	32834.07641	0	24058.42548
37.92793	-75.35547	37.9337	-75.37366	1	1	32834.07641	0	24058.42548
37.65279	-75.81152	37.68274	-75.74853	1	4	73989.04342	0	62484.32608
37.68274	-75.74853	37.65279	-75.81152	1	4	73989.04342	0	62484.32608
38.0376	-78.44911	38.0356	-78.44806	-1	20	125487.8777	146021.6873	0
38.0376	-78.44911	38.0356	-78.44806	-1	20	102202.4715	131171.4406	0
38.0356	-78.44806	38.11189	-78.36947	-1	9	102202.4715	131171.4406	0
38.11189	-78.36947	38.0356	-78.44806	-1	9	102202.4715	131171.4406	0
38.0356	-78.44806	38.03705	-78.44919	-1	20	102202.4715	131171.4406	0
38.0376	-78.44911	38.0356	-78.44806	-1	20	102202.4715	116616.3285	0
38.0356	-78.44806	38.11189	-78.36947	-1	9	102202.4715	116616.3285	0
38.11189	-78.36947	38.0356	-78.44806	-1	9	102202.4715	116616.3285	0
38.0356	-78.44806	38.03705	-78.44919	-1	20	102202.4715	116616.3285	0
37.99762	-78.38682	38.04854	-78.54141	1	1	10044.92227	10486.41744	0
38.04854	-78.54141	37.99762	-78.38682	1	1	10044.92227	10486.41744	0
37.99762	-78.38682	38.00876	-78.50059	1	3	17816.42442	19858.56251	0
38.00876	-78.50059	37.99762	-78.38682	1	3	17816.42442	19858.56251	0
38.13575	-78.52307	38.15598	-78.41541	1	4	21784.23296	19636.86136	0
38.15598	-78.41541	38.12743	-78.44375	1	4	21784.23296	19636.86136	0
38.12743	-78.44375	38.13575	-78.52307	1	4	21784.23296	19636.86136	0
38.13575	-78.52307	38.04723	-78.70723	1	3	21639.38775	19683.66076	0
38.04723	-78.70723	38.13575	-78.52307	1	3	21639.38775	19683.66076	0
38.13575	-78.52307	38.07692	-78.50662	2	3	25678.91253	21824.36772	0

Figure 6. Screenshot of a sample of trips with weights.

For this validation study, three levels of trips are explored: without trip-level weighting, with person-trip level weighting, and with vehicle-trip level weighting. In person-trip level weighting, all trips are considered regardless of whether the traveler is the driver or passenger and irrespective of the travel mode. In contrast, vehicle-trip level weighting considers only trips made by drivers in passenger vehicles, including cars, vans, SUVs/crossovers, pickup trucks, and recreational vehicles.

The following example illustrates how the weighting procedure works, with Table 1 showing five trips from zone A to B. The total number of trips without weighting, with person trip level weighting and with vehicle trip level weighting are calculated as follows:

Table 1. Example trips for weighting procedure illustration.

Index	Origin	Destination	Driver	Transit Mode	Weight
1	A	B	1	car	10
2	A	B	-1	school bus	15
3	A	B	1	van	10
4	A	B	1	motorcycle	5
5	A	B	-1	car	20

- **Number of trips from A to B without any weighting**:
  All the trips are counted, so in this case, 5 unweighted trips are observed from A to B

- **Number of trips from A to B with person trip level weighting:**
  All the trips with their associated weight are counted:
  $$1 \times 10 + 1 \times 15 + 1 \times 10 + 1 \times 5 + 1 \times 20 = 60\ trips$$

- **Number of trips from A to B with vehicle trip level weighting:**
  Only the trips for which the respondent is the driver and is driving a passenger car are counted (trips having index of 1 and 3).
  $$1\text{x}10 + 1\text{x}10 = 20\ trips$$

NHTS Dataset Caveats

There are several aspects of the NHTS dataset that are worth noting:

- As mentioned previously, the dataset encompasses all of 2022 (in contrast to vendor datasets that focus on October).
- When considering 'Typical Weekday' trips, all trips taken Monday-Friday are included (in contrast to vendor datasets that report trips taken Tuesday-Thursday). This approach was selected to correspond to the period used by NHTS to determine the 5-day weights (described above for scaling to the population level).
- The weights can be used to scale individual trip observations to the population level to characterize travel time or trip distance distributions for vehicle-trips or person-trips. However, based on our understanding, these weights cannot be applied to OD matrices of arbitrary geographies. For example, a weight of 5 does not mean that 5x the number of trips were observed at the exact same origin and destination. *Accordingly, in the following analysis, weighted results will be used primarily for travel time / trip length distributions, but OD patterns at the TAZ level will focus on un-weighted samples.*

- The 2022 NHTS represents the first survey year for which respondents were not asked to complete a travel diary. According to official documentation (link), the USDOT estimated the impact of this change as lowing trip rates by about 20% (due to the methodology rather than changes in travel behavior). Although this primarily impacts the validity of comparisons with prior years, it also may indicate that the 2022 dataset --which relies on respondents remembering trips they take rather than recording details with a trip diary -- is underreporting the true number of trips.

Analysis Methods

The intent of this analysis is to focus on OD patterns and trip length/travel time distributions at the TAZ level and quantify the extent to which the vendors show internal agreement and/or agreement with the Next Gen NHTS dataset.

It is important to re-emphasize that although comparing OD patterns and trip distributions between vendors can be instructive, all five vendors' OD products are slightly different (vehicle vs people movements, sample vs population estimates, personal vs commercial vehicle movements), and thus, differences are expected. *This is especially true of Geotab, which reports commercial vehicle movements, and thus is fundamentally different than the other vendors and the Next Gen NHTS dataset.* Additionally, it should be noted that NHTS – which is not a 'ground truth' dataset but serves as an important point of comparison -- differs slightly in terms of temporal period and aggregation; it uses all of 2022 (rather than just the month of October) and uses all weekdays in 'Typical weekday' calculations (rather than Tu-Th). Furthermore, the 2022 NHTS survey uses a new data gathering methodology relative to prior years, which likely undercounts the true trip rate.

For this study, both trip statistics and TAZ-level OD analysis are performed. This is similar to the Phase 1 report analysis, but trips are limited to a specific county, and analysis zones for OD patterns are at the TAZ, rather than county level. Trip Statistics focus on daily trip counts, travel time distributions, and distance distributions across all trips within the target county, while the OD Analysis includes comparison of top origins, top destinations and top OD pairs for TAZ zones.

Results and Discussion

In this study, the primary analysis focuses on trips within Chesterfield County – an area that contains a blend of urban and rural characteristics, and according to the US Census Bureau had a 2023 population of 383,876 (link). Subsequently, a subset of analyses is performed for two additional counties: Richmond City (population 229,247), serving as a case study for an urban area, and Goochland County (27,197), representing a rural area. These comparisons are intended to determine whether the results are consistent in both urban and rural areas.

Chesterfield County

Trip Summary Statistics

Tables 2 and 3 report descriptive statistics for all vendors and the NHTS, respectively, including trip count (average trips per day or day part), average trip distance, and average travel time for trips within Chesterfield County. Results are broken down by scenario, including *Day Type* & *Day Part,* where possible. Afterwards, Figure 7 and Figure 8 compare histograms for travel time and trip distance, providing additional detail into how each source's data is distributed.

Two caveats are worth noting upfront regarding travel time and trip distances. First, two vendors (AirSage and StreetLight) did not report travel time values, but this speaks more towards how data was queried at the time of data collection than vendor capabilities. While not available for this analysis, both vendors' OD products include this information, and it is expected that travel time data could be obtained in future efforts. Additionally, AirSage's trip distances were not collected in the query, so the validation team used centroid-to-centroid distances between OD pairs as an approximation. This approximation does not represent AirSage's native trip distance methodology, which is based on full path reconstruction using location-based services (LBS) data, not centroid to centroid calculations.

Table 2. Descriptive statistics of ODs by data vendors in Chesterfield County

AirSage				
Day Type	**Day Part**	**Count (Pop Est)**	**Avg Dist* (mi)** **Centroid-to-Centroid*	**Avg TT (min)**
Typical Weekday	AM Peak	98,243	3.9	N/A
Typical Weekday	PM Peak	70,408	3.2	
Typical Weekday	Off Peak	330,564	3.0	
Typical Weekday	All	499,215	3.2	
Weekend	All	531,384	2.8	
INRIX				
Day Type	**Day Part**	**Count (Sample)**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	5,209	6.6	15.3
Typical Weekday	PM Peak	8,570	5.5	13.6
Typical Weekday	Off Peak	20,593	5.2	12.5
Typical Weekday	All	34,372	5.5	13.2
Weekend	All	23,472	5.4	12.7
Streetlight				
Day Type	**Day Part**	**Count (Pop Est)**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	85,925	7.2	N/A
Typical Weekday	PM Peak	178,918	5.9	
Typical Weekday	Off Peak	402,549	5.6	
Typical Weekday	All	677,602	5.9	
Weekend	All	548,660	5.8	
Geotab				
Day Type	**Day Part**	**Count (Sample)**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday **15 min trip chain param*	All	2,025	5.2	19.3
Replica				
Day Type	**Day Part**	**Count (Pop Est)**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	180,320	5.6	18.0
Typical Weekday	PM Peak	244,803	5.7	17.5
Typical Weekday	Off Peak	530,010	5.6	16.9
Typical Weekday	All	955,133	5.6	17.3
Weekend	All	864,713	5.6	17.0

Table 3. Descriptive statistics of ODs by NHTS in Chesterfield County

NHTS – No Weighting				
Day Type	**Day Part**	**Count**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	0.67	5.6	19.8
Typical Weekday	PM Peak	1.33	6.2	18.1
Typical Weekday	Off Peak	2.73	5.2	14.9
Typical Weekday	All	4.73	5.6	16.5
Weekend	All	2.55	5.4	16.0
NHTS – Person level weighting				
Day Type	**Day Part**	**Count**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	86,763	5.8	22.1
Typical Weekday	PM Peak	155,144	6.2	18.0
Typical Weekday	Off Peak	260,781	5.3	15.2
Typical Weekday	All	502,688	5.7	17.2
Weekend	All	475,079	4.0	13.8
NHTS – Vehicle-level weighting				
Day Type	**Day Part**	**Count**	**Avg Dist (mi)**	**Avg TT (min)**
Typical Weekday	AM Peak	49,156	6.1	15.1
Typical Weekday	PM Peak	92,782	6.9	16.4
Typical Weekday	Off Peak	176,958	5.4	13.8
Typical Weekday	All	318,897	5.9	14.8
Weekend	All	252,407	4.4	13.6

The comparison of travel distances and travel times across the different datasets—AirSage, INRIX, Streetlight, GeoTab, Replica, and NHTS—reveals some insights. First, consider the Count column, which summarizes the average trips per day (or day part) for each vendor and the NHTS datasets. While it is not particularly meaningful to compare count values that represent observed samples (e.g., INRIX's connected vehicle sample count to NHTS's survey sample count), comparing counts that reflect population estimates can provide more insight. For example, for scenario '*Typical Weekday, All Hours'*, AirSage estimates about 499.2k trips/weekday, while StreetLight and Replica estimate 677.6k and 955.1k, respectively. Given the population of Chesterfield county, this implies about 1.3, 1.8, and 2.5 trips/weekday per person. For comparison, the NHTS dataset shows about 502.7k person-trips/weekday (1.3 per person in the county) for person-trips and 318.9 vehicle-trips/weekday (0.8 per person in the county). These results are in the same ballpark, with AirSage's counts matching the NHTS weighted person-trip counts quite closely. Again, however, note that NHTS's new data gathering technique introduced in 2022 may have underestimated the true trip rate.

Average trip distances show general agreement across most vendors and the NHTS datasets, typically ranging from 4-7 miles for all vendors except AirSage, which was a bit lower. However, this is likely an artifact of how distance is calculated for AirSage (centroid-to-centroid), which can have zero-distance trips when a trip starts and ends in the same TAZ. Weekday trips showed particularly strong agreement (ranging from 5.7 to 5.9 miles for NHTS datasets and 5.2 to 5.9 miles for INRIX, StreetLight, Geotab, and Replica).

Average travel times from vendors were also in the same general range as NHTS values, though details vary by day type and part. The NHTS vehicle-trip dataset showed an average travel time of 15.1 and 16.2-minutes during weekday AM/PM peak periods, with Off-Peak weekday travel times around 13.8 minutes. As a point of comparison, INRIX showed AM/PM peak travel times of 15.3 and 13.6 minutes respectively, with Off-Peak around 12.8 minutes. Their AM peak and Off-peak travel times were very close to NHTS, although the PM peak differed by over 2 minutes. The NHTS person-trip dataset showed higher average travel times, with weekday AM Peak, PM Peak, and Off-Peak travel times of 22.1, 18.0, and 15.2 minutes, respectively. Replica's person-trip dataset showed travel times of 18.0, 17.5, and 16.9 for these same aggregations. Finally, Geotab – whose results are summarized by Typical Weekday without breaking down by day part, reported an average travel time of 19.3 minutes.

Travel Time Distributions

Figure 7 shows the distribution of weekday travel time across the range of values for NHTS and the vendor datasets, enabling a more detailed view of the underlying values that underpin the average values reported in Tables 2-3. As is the case in Tables 2-3, two vendors (AirSage and StreetLight) were not shown in Figure 7 because travel time data was not available.

The bar plots in Figure 7 illustrate a general skewness towards shorter trips. In particular, most of the NHTS trips (weighed and unweighted) have a travel time between 5 to 20 minutes, with the highest frequency in the 10-15-minute category. Vendor trips also have the majority of trips between 5-20 minutes, with Replica showing similarly shaped distribution to the weighted NHTS datasets (highest frequency in 10–15-minute category), and INRIX and Geotab having the most trips in the 5–10-minute category. Only a small portion of trips have travel times above 35 minutes, which is expected considering these trips only represent trips starting and ending in Chesterfield VA (although some Geotab freight trips do, as could be expected). In general, the trips documented by all vendors exhibit a similar trend to those recorded in the NHTS. It is worth mentioning that NHTS trips show a greater proportion falling within the 30 to 35-minute range, possibly attributed to respondents rounding off their reported travel times to a half hour.

Travel Distance Distributions

Figure 8 demonstrates the weekday travel distance distributions of the trips reported by AirSage, INRIX, StreetLight, Geotab, Replica, and the NHTS survey. Similar to the distribution of travel times, the vendors demonstrate travel distance distribution patterns that are generally comparable to NHTS trips, with INRIX and StreetLight distributions visually matching NHTS distributions the most closely. Across datasets, most trips have a travel distance of less than 10 miles, with vendors reporting the highest frequency in either the 0-2 mi or 2–4-mile bins – consistent with NHTS survey results (highest frequency in 2-4 mi bin, followed by the 0-2 mi bin). As a reminder, AirSage did not directly provide travel distance; it is computed by the validation team as the centroid-to-centroid distance between zones at the TAZ level. When trips are internal to a TAZ, they have distance = 0, which is bucketed in a separate histogram bin for clarity.

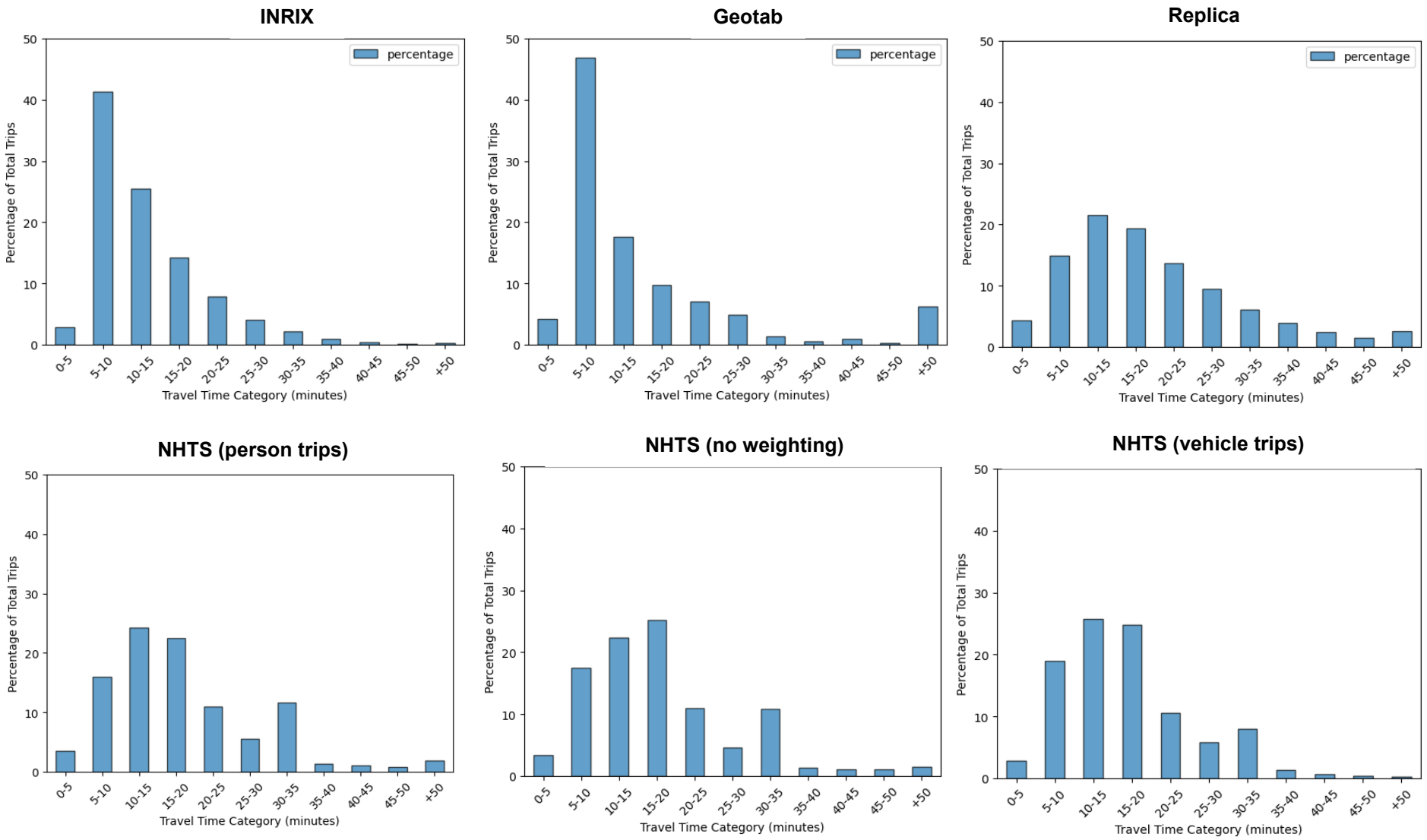

Figure 7 – Travel time distributions for trips in Chesterfield County

[1] AirSage and StreetLight travel times are not included due to how their data was queried at the time of data collection. However, travel time information is available for both vendors' OD products.

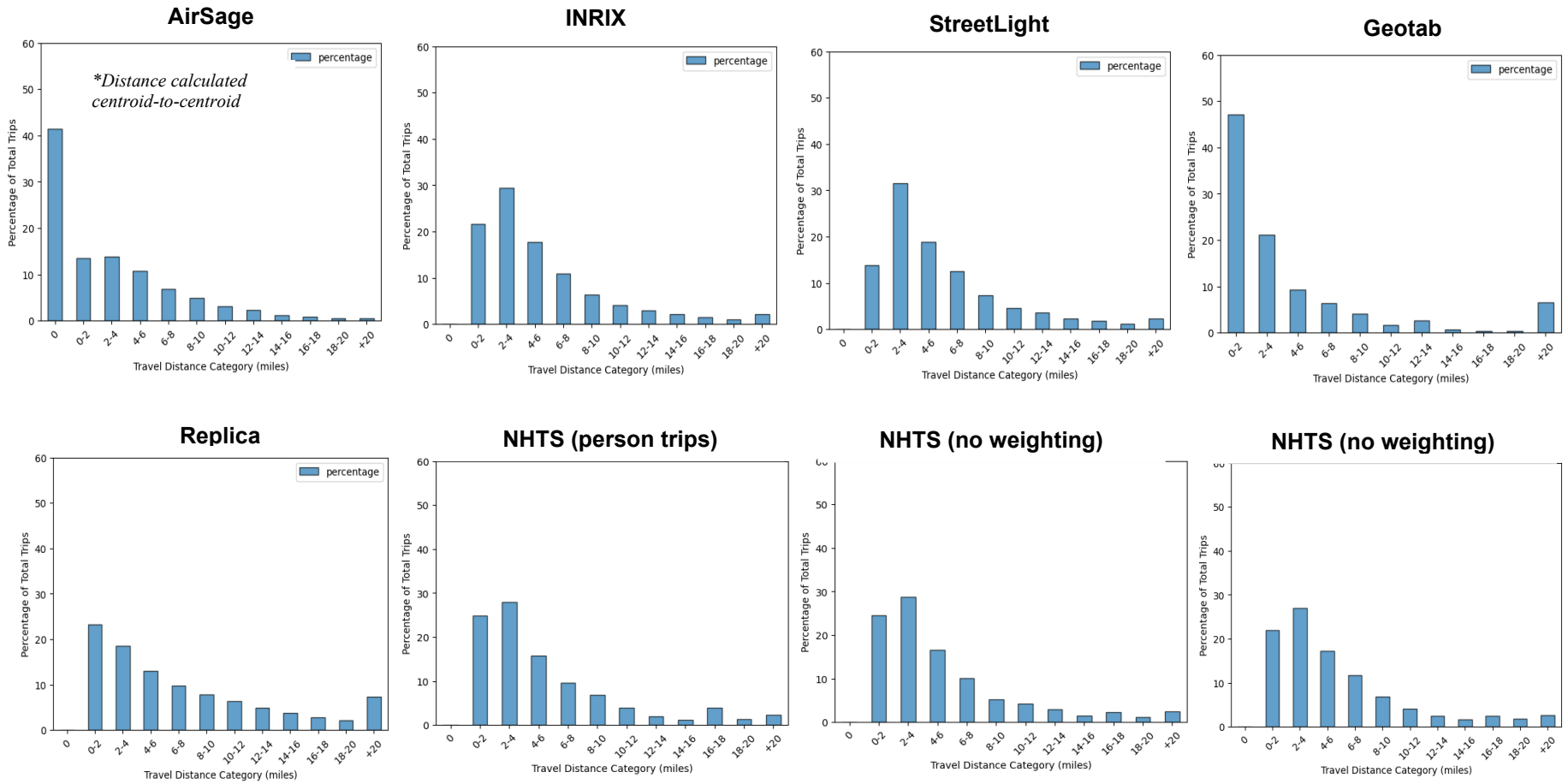

Figure 8 – Travel distance histograms for trips in Chesterfield County

Travel Distance CDF plots by time-of-day

The travel time and trip distance histograms above focus on weekday travel time across all time periods. Rather than repeating these plots for different times of day, we use empirical Cumulative Distribution Functions (CDF) to illustrate time-of-day specific travel distances and travel times (i.e., AM Peak, PM Peak, Off-peak, and All hours) for each vendor on the same plot – shown below in Figures 9 and 10. To interpret travel time and distance CDF plots, note that the x axis represents travel time or distance, while the y axis shows the cumulative portion of trips. Each point on the plot indicates the proportion of trips with travel times or distances less than or equal to a specific value on the x axis. A steep slope signifies that many trips are clustered around a particular travel time or distance, whereas a gradual slope suggests a more even distribution. When comparing different CDFs, a curve that reaches higher percentages more quickly indicates shorter travel times or distances. The median travel time or distance can be identified where the CDF reaches 50% on the y axis and other percentiles can be similarly interpreted from the plot.

Based on Figures 9 and 10, there is a general agreement between different data vendors and NHTS vehicle-level weighted results, with a slight difference observed between the CDF plots of AM peaks and all other periods in data vendor products. Furthermore, the CDF plots of NHTS results show higher percentages at slightly shorter travel times and distances.

As mentioned earlier, it is important to note that although AirSage does not directly supply travel distance data, the validation team calculates it as the distance from centroid to centroid between zones at the TAZ level. As a result, numerous trips are recorded with a distance of zero, representing intrazonal travel, which is why the CDF plot does not begin at zero. Moreover, the CDF of NHTS demonstrates the discrete nature of surveys compared to the continuous nature of travel time and distance reported by vendors. This distinction is evident as NHTS data points typically represent specific, separate intervals, whereas vendor-reported data tends to provide a continuous range of travel times and distances.

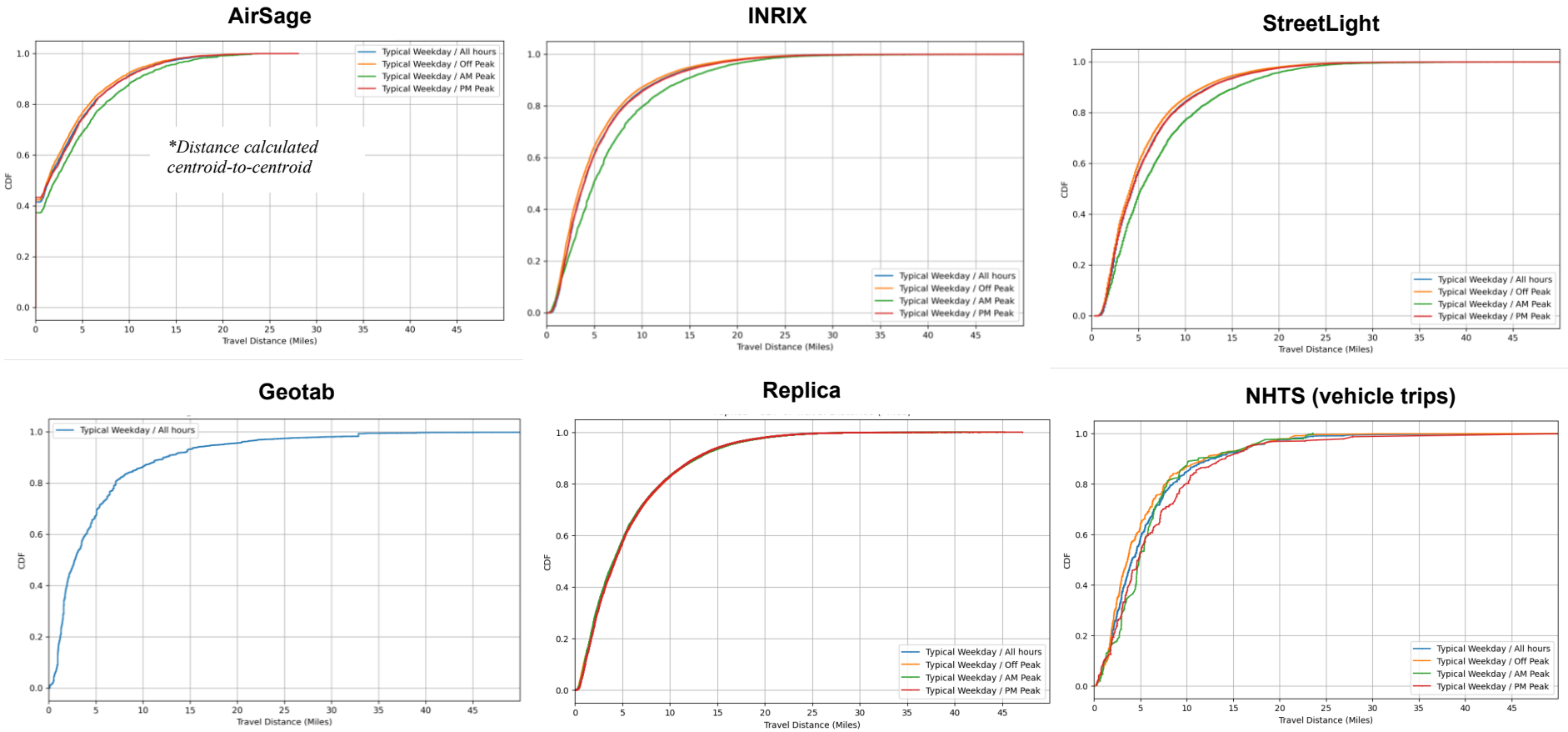

Figure 9 – Travel Distance CDF plots for trips in Chesterfield County

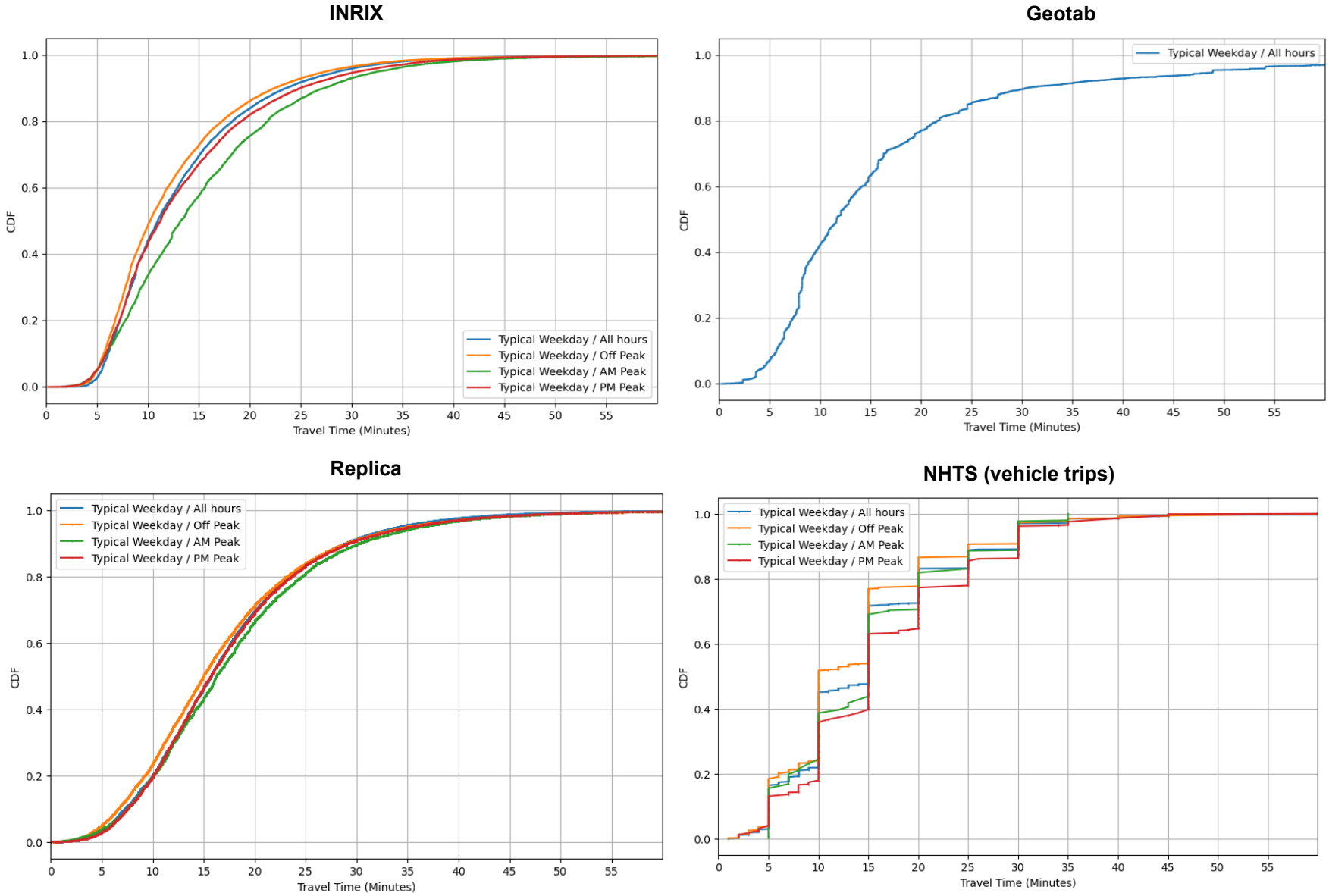

Figure 10 – Travel Time CDF plots for trips in Chesterfield County

OD Analysis: Top Origins and Destinations

Before comparing Origin-Destination patterns at the TAZ level, it is helpful to first summarize each dataset's coverage in terms of number (or percentage) of unique origin zones, destination zones, and OD zone pairs, as shown in Table 4. Of the total 242 TAZs in Chesterfield County, most vendors' datasets contain 242 (100%) of the possible unique origin and destination TAZs – both during typical weekday and weekend periods -- with the exception being Geotab, which has trips originating in about 76% of TAZs and ending in about 77% of TAZs on typical weekdays. The NHTS is similar during typical weekdays (about 70% and 71% of unique origin and destination TAZs, respectively), but drops to about 33% for O's and D's during weekend periods.

The last column in Table 4 gives an indication of the sparseness of each OD matrix; there are over 58k possible OD pairs (242 * 242 TAZs) in Chesterfield County, and vendor datasets show a wide range of observed pairs. Replica's dataset included the most unique OD pairs (70% and 66% for typical weekdays and weekends, respectively), followed by INRIX (50% and 40%), StreetLight (40% and 35%), AirSage (16% and 14%), and Geotab (2% for typical weekdays). For the NHTS dataset, only 0-1% of OD pairs were captured. ***It should be noted that these percentages do not imply better or worse accuracy,*** as there are not necessarily observed trips between every possible – or even the majority of -- OD pairs. Furthermore, nuances in how vendors define trips, their underlying data sources (e.g., LBS vs Connected Vehicle), and scaling policies (e.g., AirSage provides scaled outputs specifically for pairs of zones observed in underlying samples) may contribute to differences in OD matrix sparseness.

Table 1 – Summary of unique O's, D's, and O-D pairs for each data source in Chesterfield County

Data Source	Day Type	Unique Origin TAZs*	Unique Destination TAZs*	Unique OD Pair TAZs**
AirSage	Typical Weekday	242 (100%)	242 (100%)	9548 (16%)
	Weekend	242 (100%)	242 (100%)	8003 (14%)
INRIX	Typical Weekday	242 (100%)	242 (100%)	29068 (50%)
	Weekend	242 (100%)	242 (100%)	23450 (40%)
StreetLight	Typical Weekday	242 (100%)	242 (100%)	23281 (40%)
	Weekend	242 (100%)	242 (100%)	20369 (35%)
Geotab	Typical Weekday	183 (76%)	186 (77%)	1104 (2%)
Replica	Typical Weekday	242 (100%)	242 (100%)	41006 (70%)
	Weekend	242 (100%)	242 (100%)	38742 (66%)
NHTS (Unweighted)	Typical Weekday	170 (70%)	171 (71%)	785 (1%)
	Weekend	80 (33%)	81 (33%)	159 (0.3%)

* 242 total TAZs in Chesterfield County
* 58,564 total OD Pairs (=242*242)

Limiting our analysis to Typical Weekdays, Table 4 shows that all datasets – including NHTS – have at least 170 unique trip origin and destination TAZs – with most containing the full 242. Given this high level of coverage for trip Origins and Destinations (separately, if not at the OD Pair level), top origins and destinations can be meaningfully identified and compared between data sources.

Tables 5-6 summarize the top origins and destinations for the unweighted NHTS dataset. These tables highlight that beyond the first ranked O and D (TAZ 733), all others are bunched close together without meaningful separation – a finding that is consistent when inspecting vendor results. Because origins/destinations #2-5 are not much different than #6-10 and geographies are so small, simply identifying and comparing the top few origins/destinations from each vendor is likely not a meaningful way to compare OD patterns. As such, when comparing top origins/destinations, we focus the top 10% of O's and D's (in this case, the top 24 of 242 TAZs).

Table 5 – Top 10 Origins in Chesterfield County for NHTS (unweighted)

Rank	Origin TAZ	# Trips
1	733	54
2 (Tie)	726	32
2 (Tie)	795	32
4	810	30
5	724	29
6	740	27
7 (Tie)	822	26
7 (Tie)	722	26
9	782	25
10	796	23

Table 6 – Top 10 Origins in Chesterfield County for NHTS (unweighted)

Rank	Origin TAZ	# Trips
1	733	56
2	726	33
3	792	30
4	724	29
5	795	28
6 (Tie)	810	27
6 (Tie)	722	27
6 (Tie)	740	27
9	806	24
10	822	23

Figure 11 shows the top 10% of origin zones in terms of the number of trips reported by vendors and the unweighted NHTS dataset, while Figure 12 shows the top 10% of destinations. The maps indicate that visually, many of the vendors appear to share similar top-ranked Origins and Destinations (with the exception of Geotab, which is recording commercial freight trips). It is difficult to see whether the exact TAZs are the same in the various maps, but minimally, many of the top O's and D's appear in similar locations.

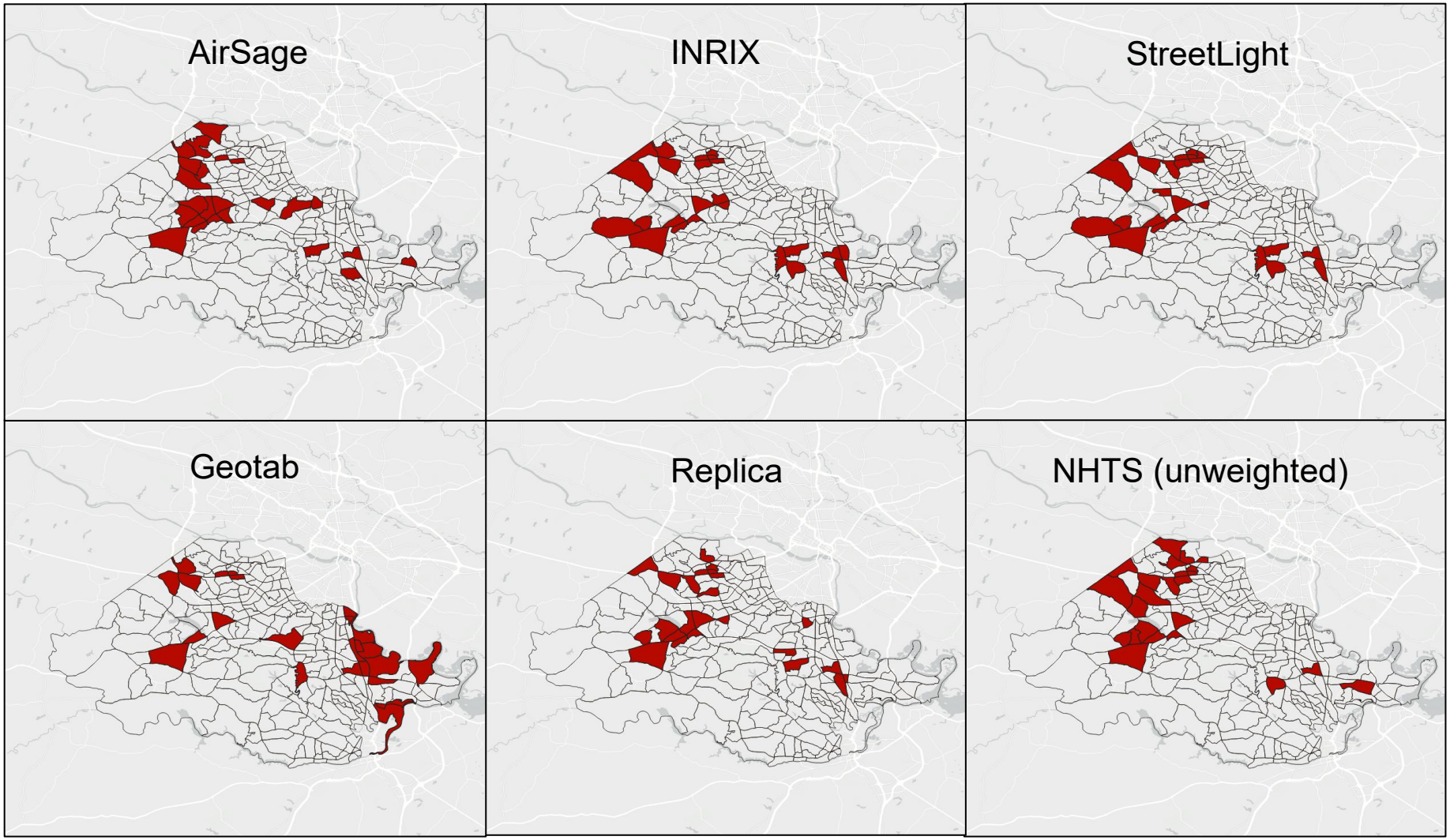

Figure 11 – Map view of top 10% of origins for trips in Chesterfield County.

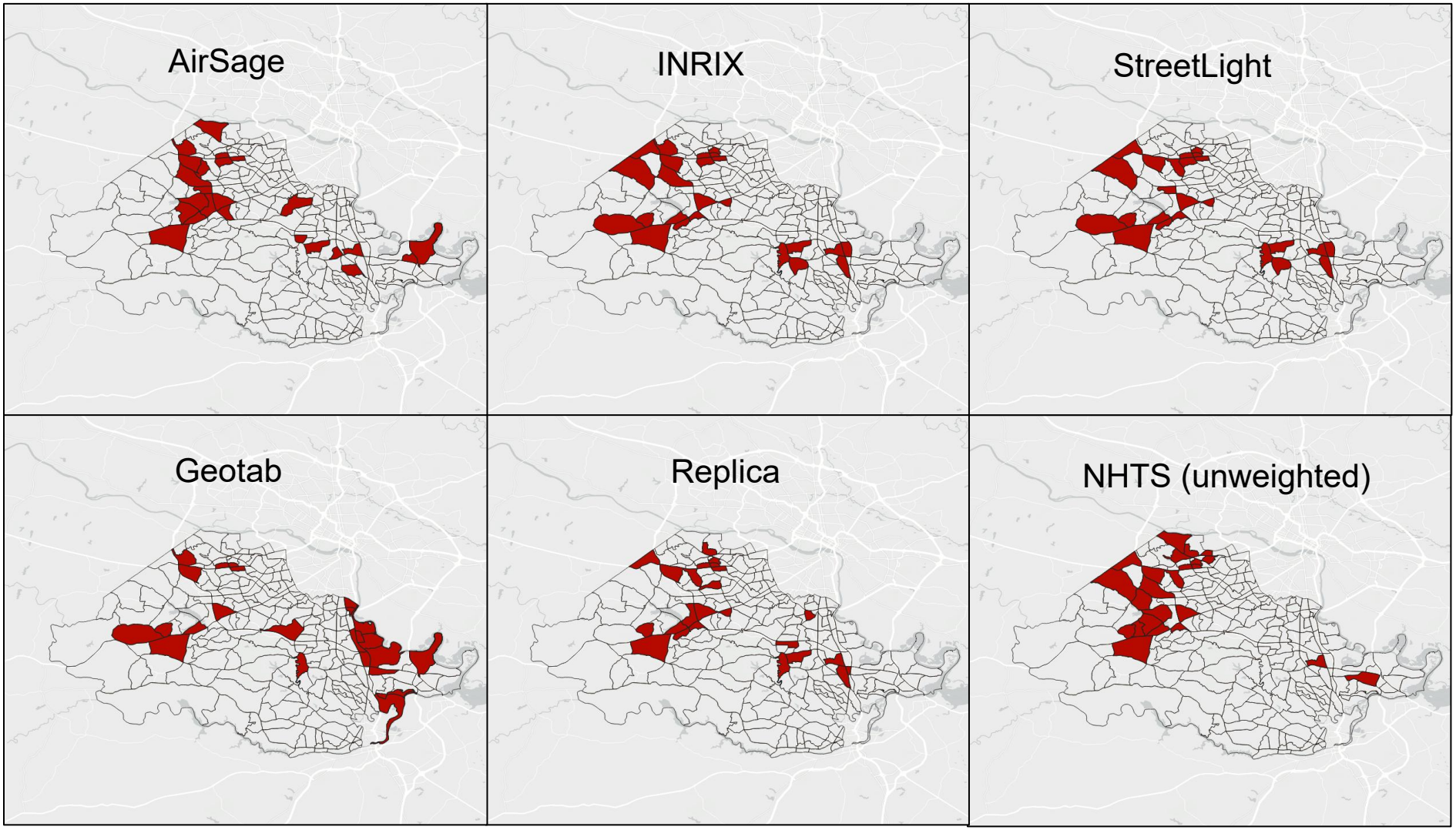

Figure 12 – Map view of top 10% of destinations for trips in Chesterfield County

Figures 14 and 15 illustrate a similar, but slightly different approach for quantifying how well the top origins and destinations match, with Figure 14 focusing on origins and Figure 15 on destinations. Each figure contains a heatmap whose cell values represent the percentage of common origins (or destinations) ranked in the top 10% by trip count. All datasets are represented in this matrix, including INRIX, AirSage, Geotab, StreetLight, Replica, and the unweighted NHTS data. The idea, then, is to see how much overlap there is between top origins (Fig 14) and top destinations (Fig 15) – an approach that is less sensitive to slight differences in ranking. More concretely, this involves finding the top 24 origin or destination TAZs (i.e., 10% of the 242 TAZs in Chesterfield County) for each data source and calculating percentage overlap between each pair. For example, if a pair of data sources share 10 of the same top 24 origins, their overlap percentage would be 10/24*100 = 41.7%.

The highest overlaps are observed between INRIX and StreetLight in both origin and destination heatmaps (88% and 92%, respectively). The lowest overlaps are often between Geotab and other data vendors, which makes sense given the nature of freight trips. The NHTS data (unweighted) has the highest percentage overlap with StreetLight, INRIX, and Replica datasets for top Origins (62%, 58%, and 58%, respectively) and top Destinations (all 54%). However, it is worth emphasizing that (1) we are comparing un-weighted and weighted products together (i.e., ones that represent observed samples and others that have been scaled to estimate population trends) and (2) we are including a freight vendor that represents different types of trips.

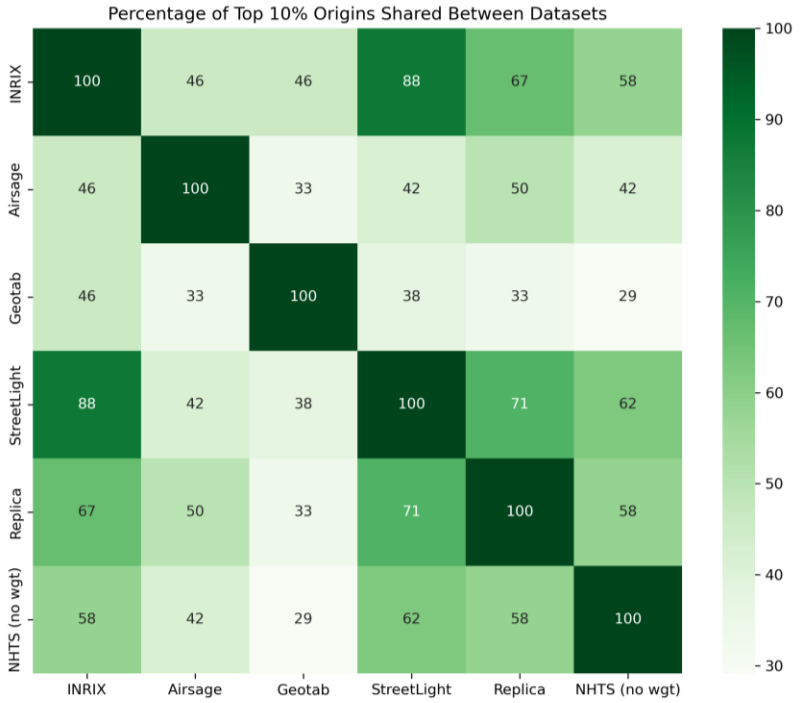

Figure 14 – Percentage of top 10% origin zones shared between sources in Chesterfield

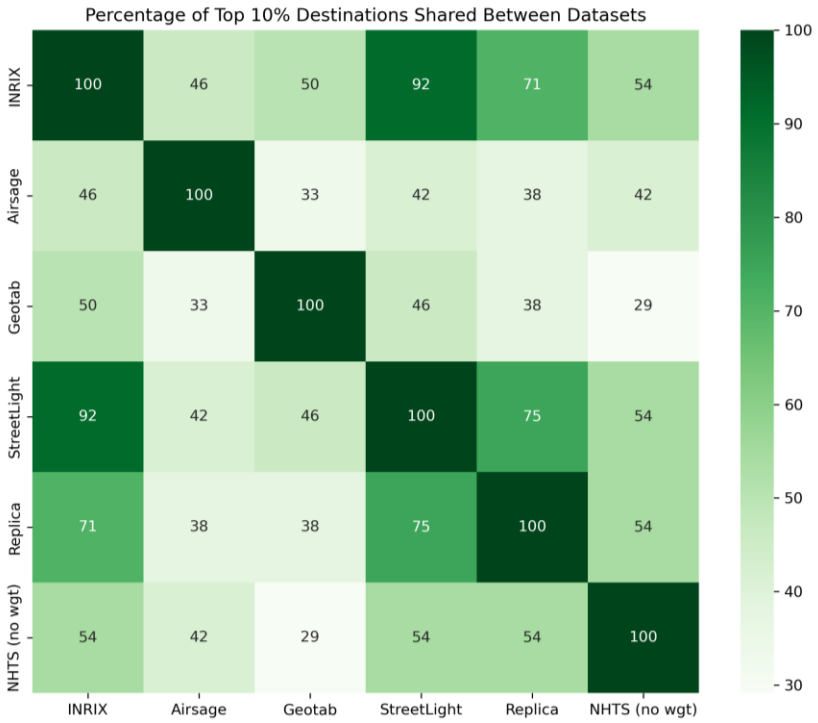

Figure 7 – Percentage of top 10% destination zones shared between sources in Chesterfield

Additional Analysis in Urban and Rural Counties

All results produced so far have focused on Chesterfield County, which was selected as a representative county that contains both urban and rural areas. After sharing preliminary results with VDOT, it was recommended to check observed findings with at least one other county with different characteristics. Two other counties with differing characteristics, Richmond City (urban) and Goochland (rural), were selected for subsequent analysis to determine if similar patterns would be observed. For each county, a subset of the analysis -- travel time histograms and top O's and D's – were produced as a point of comparison.

Richmond City (Urban)

Figure 16 shows the travel time distribution of trips in Richmond City. As was the case in Chesterfield, the histogram in Figure 16 shows a skewness towards shorter trips, with most NHTS trips having travel times between 5-20 minutes, with the highest frequency in the 15-20-minute category. Vendor trips also have the majority of trips between 5-20 minutes, with Replica showing the highest frequency in 10–15-minute category, and INRIX and Geotab having the most trips in the 5–10-minute category. Although travel time distributions were similar, the distributions appear to be slightly more concentrated at lower values.

Trip length distributions, shown in Figure 17, were more noticeably distinguishable from Chesterfield, with values shifted towards lower values – in particular a spike in the 0–2-mile trip distance range. Whereas NHTS and several vendors had the most trips in the 2-4 mile category in Chesterfield, NHTS and vendor results showed the highest concentration in the 0–2-mile category in Richmond City. These observations -- shorter-distance trips, with comparable but slightly lower travel times – are intuitive for a more urban environment.

Top Origins and Destinations

Table 5 summarizes each dataset's coverage in terms of number (or percentage) of unique origin zones, destination zones, and OD zone pairs in Richmond County. Of the total 219 TAZs in Richmond County, most vendors' datasets contain close to the maximum possible unique origin and destination TAZs – both during typical weekday and weekend periods -- with the exception being Geotab, which has trips originating in about 84% of TAZs and ending in about 86% of TAZs on typical weekdays. The NHTS is a bit lower during typical weekdays (about 69% and 68% of unique origin and destination TAZs, respectively), but drops to about 26-27% for O's and D's during weekend periods.

Additionally, there are about 48k possible OD pairs (219 * 219 TAZs) in Richmond County, and vendor datasets show a wide range of observed pairs. Replica's dataset included the most unique OD pairs (78% and 71% for typical weekdays and weekends, respectively), followed by INRIX (48% and 35%), StreetLight (35% and 28%), AirSage (7% and 6%), and Geotab (1% for typical weekdays). For the NHTS dataset, only 0-1% of OD pairs were captured. *These percentages are a bit lower than Chesterfield but are meant to be a descriptive characterization of the datasets and not a measure of accuracy.* There are not necessarily trips between every possible – or even the majority of -- OD pairs.

Table 5 – Summary of unique O's, D's, and O-D pairs for each data source in Richmond

Data Source	Day Type	Unique Origin TAZs*	Unique Destination TAZs*	Unique OD Pair TAZs**
AirSage	Typical Weekday	219 (100%)	219 (100%)	3589 (7%)
	Weekend	217 (99%)	215 (98%)	2701 (6%)
INRIX	Typical Weekday	219 (100%)	219 (100%)	23105 (48%)
	Weekend	219 (100%)	219 (100%)	16739 (35%)
StreetLight	Typical Weekday	219 (100%)	219 (100%)	16843 (35%)
	Weekend	219 (100%)	219 (100%)	13405 (28%)
Geotab	Typical Weekday	184 (84%)	189 (86%)	564 (1%)
Replica	Typical Weekday	219 (100%)	219 (100%)	37229 (78%)
	Weekend	219 (100%)	219 (100%)	34212 (71%)
NHTS (Unweighted)	Typical Weekday	152 (69%)	149 (68%)	429 (1%)
	Weekend	59 (27%)	56 (26%)	110 (0.2%)

* 219 total TAZs in Richmond County
* 47,961 total OD Pairs (=219*219)

Figures 18 and 19 illustrate the top 10% of origin and destination TAZs in terms of the number of trips reported by vendors and the 2022 Next Gen NHTS in Richmond County. These maps show that while there is some slight overlap between top O's or D's between sources (e.g., StreetLight and INRIX), there is not broad agreement. However, as noted previously in the section focusing on Chesterfield County, trip counts are bunched closely together, and minor differences can alter the ranking of top origins or destinations.

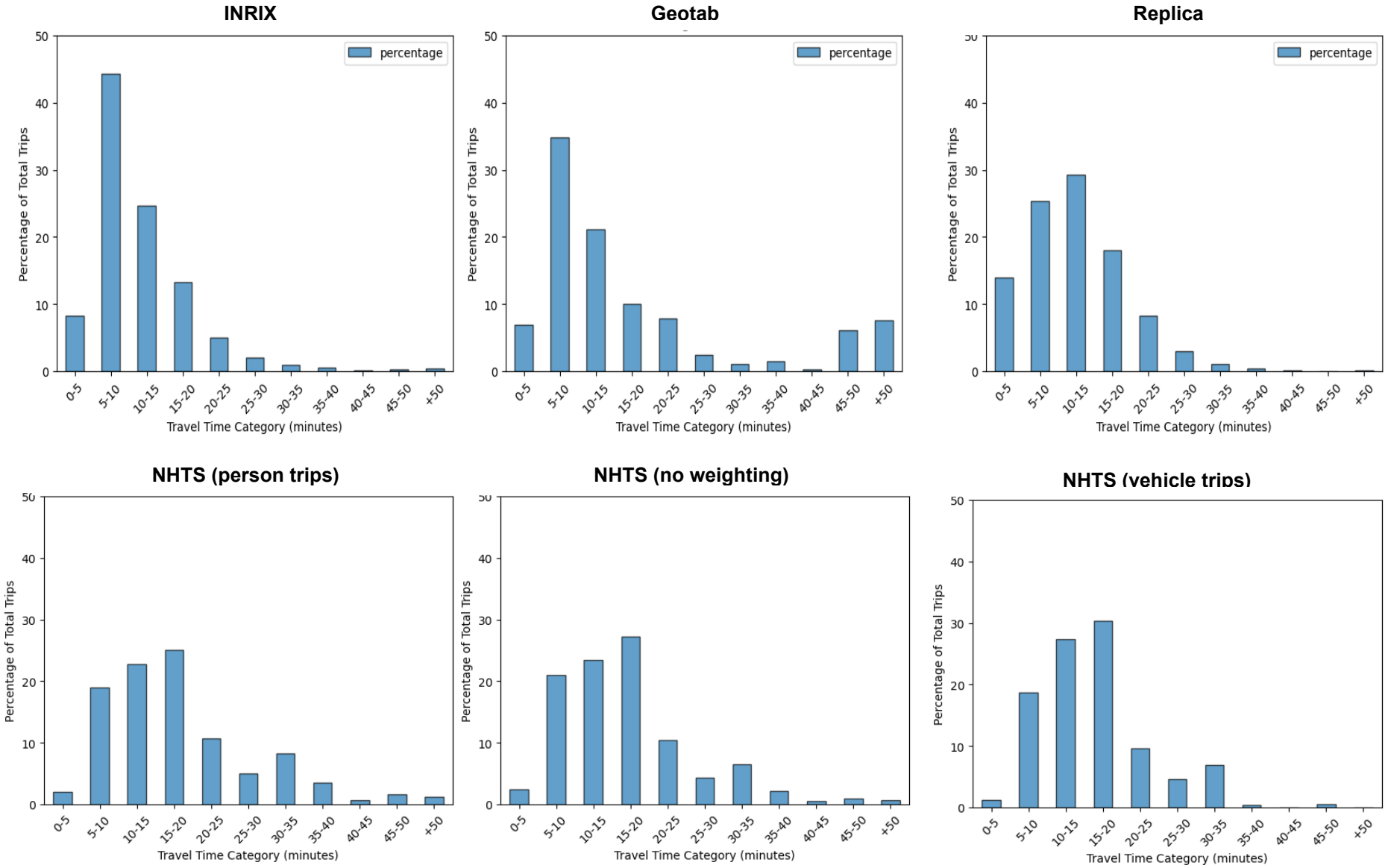

Figure 16 – Travel Time distributions for trips in Richmond

[2] AirSage and StreetLight travel times are not included due to how their data was queried at the time of data collection. However, travel time information is available for both vendors' OD products.

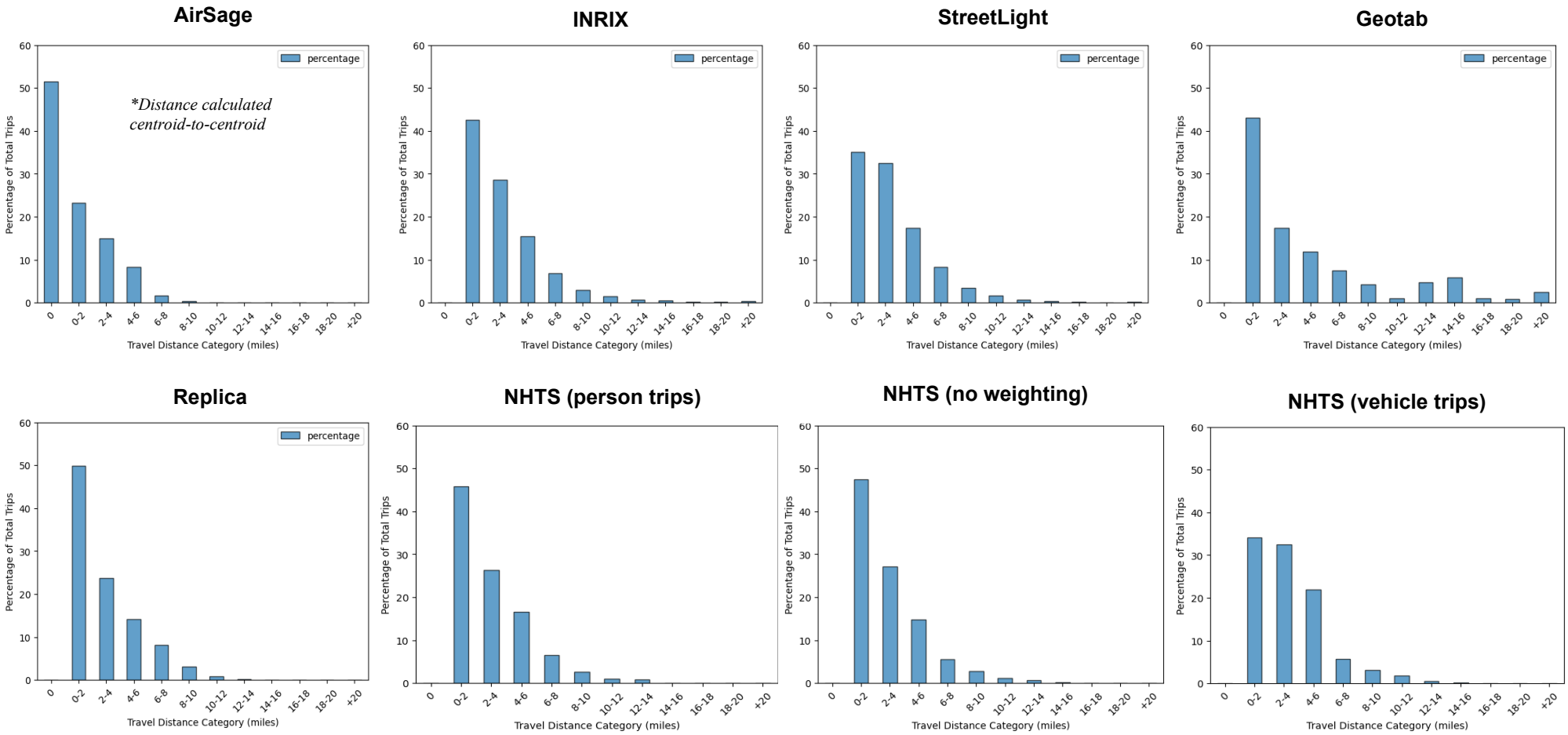

Figure 17 – Travel Distance distributions for trips in Richmond

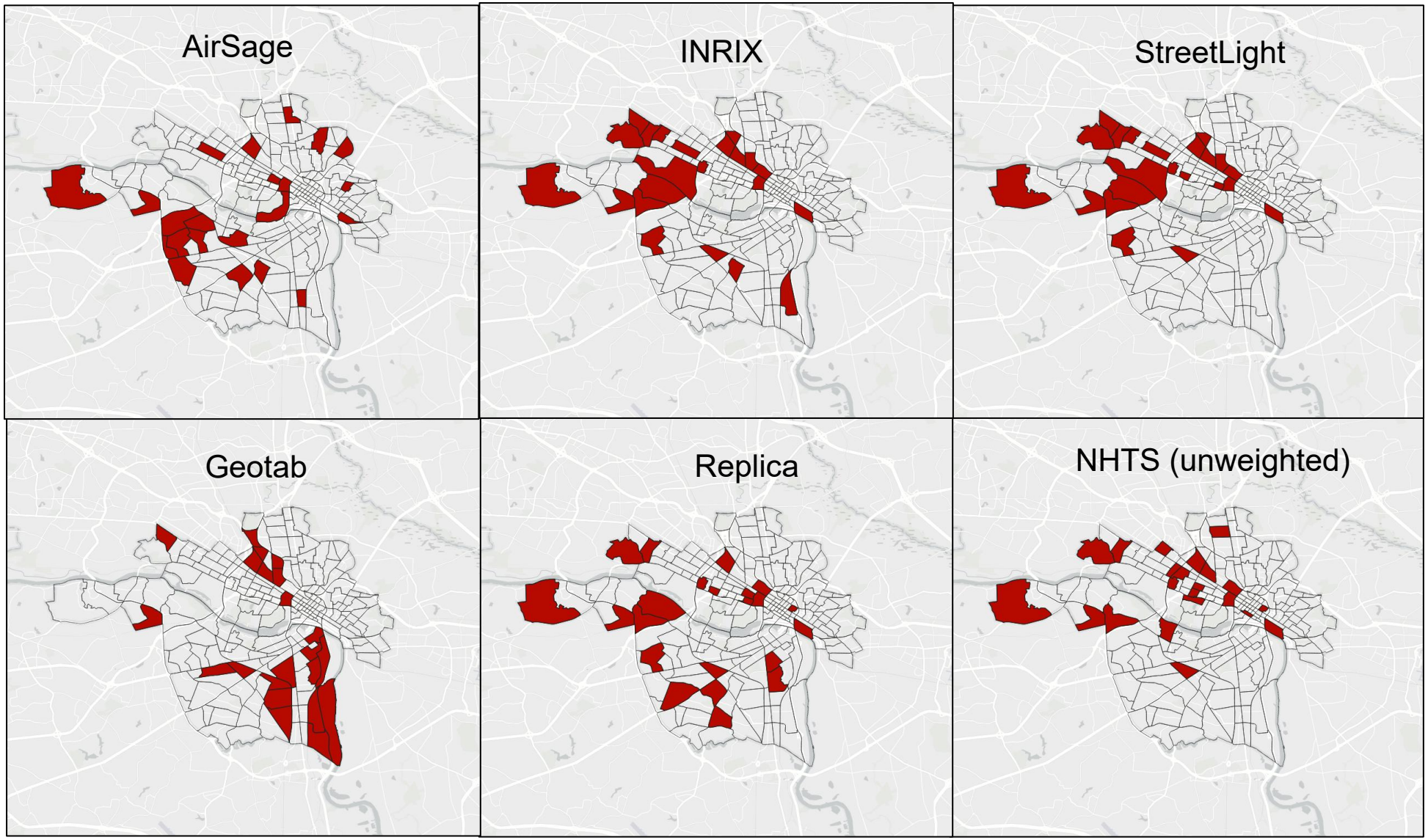

Figure 18 – Map view of top 10% of origins for trips in Richmond

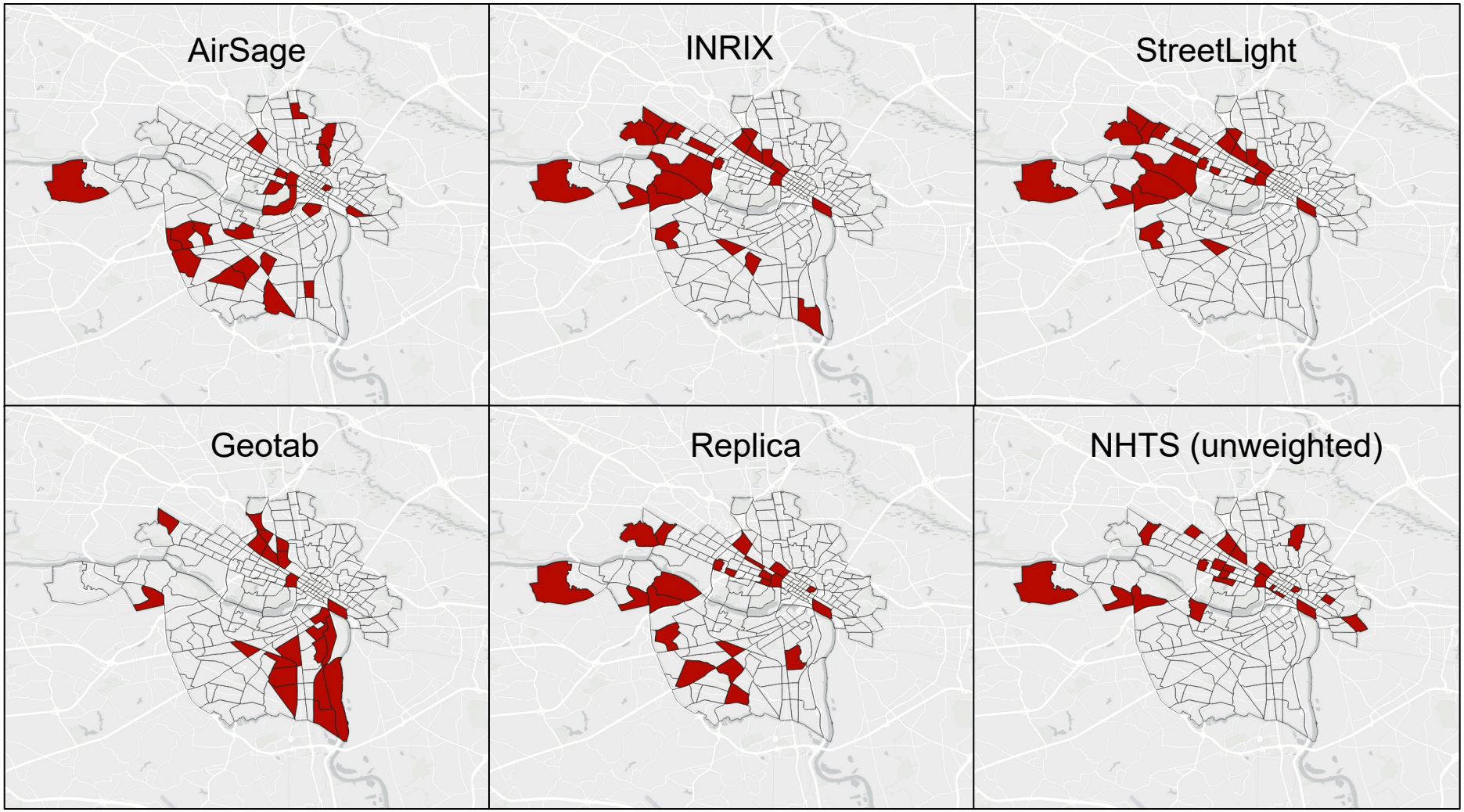

Figure 19 – Map view of top 10% of destinations for trips in Richmond

Figures 20 and 21 illustrate the alternative approach for quantifying how well the top origins and destinations match (similar to Figures 14-15 for Chesterfield County), with Figure 20 focusing on origins and Figure 21 on destinations. Each figure contains a heatmap whose cell values represent the percentage of common origins (or destinations) ranked in the top 10% by trip count – enabling comparisons that are less sensitive to slight differences in ranking. In the case of Richmond County, this involves finding the top 22 origin or destination TAZs (i.e., 10% of the 219 TAZs in Richmond County) for each data source and calculating percentage overlap between each pair. For example, if a pair of data sources share 10 of the same top 22 origins, their overlap percentage would be 10/22*100 = 45.5%.

The highest overlaps are observed between INRIX and StreetLight in both origin and destination heatmaps (91% for both). The NHTS data (unweighted) has the highest percentage overlap with StreetLight, INRIX, and Replica datasets for top Origins (55% for all) and top Destinations (45% for all). However, it is worth emphasizing that (1) we are comparing un-weighted and weighted products together (i.e., ones that represent observed samples and others that have been scaled to estimate population trends) and (2) we are including a freight vendor that represents different trips.

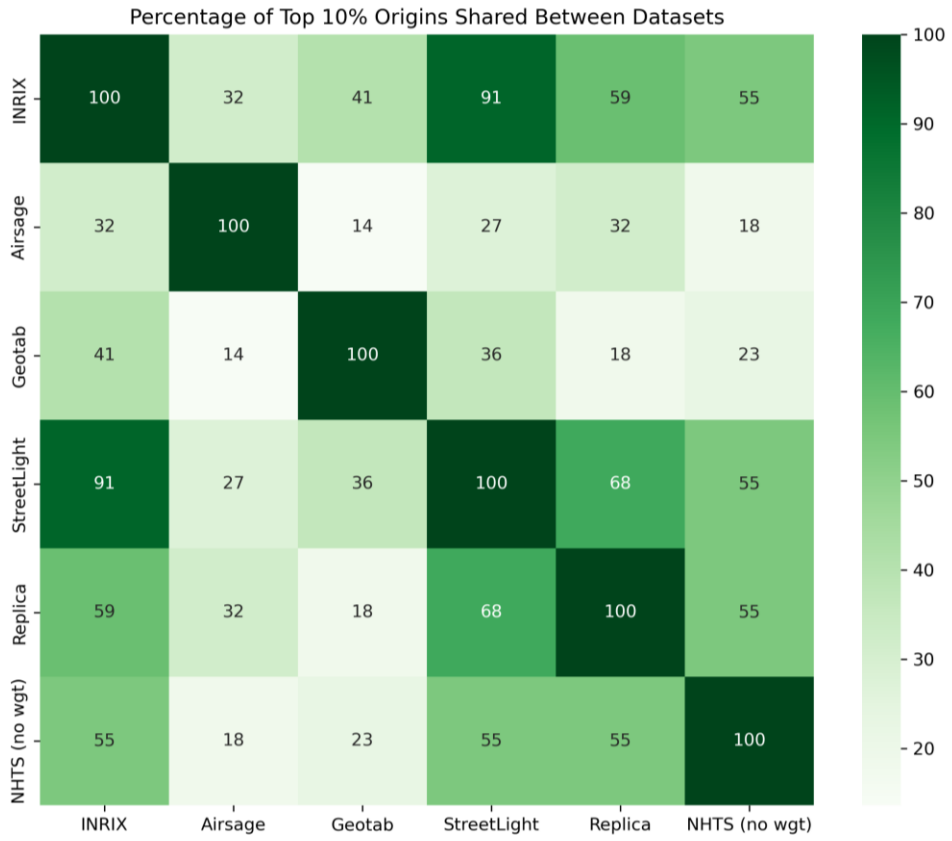

Figure 20 – Percentage of top 10% origin zones shared between sources in Richmond

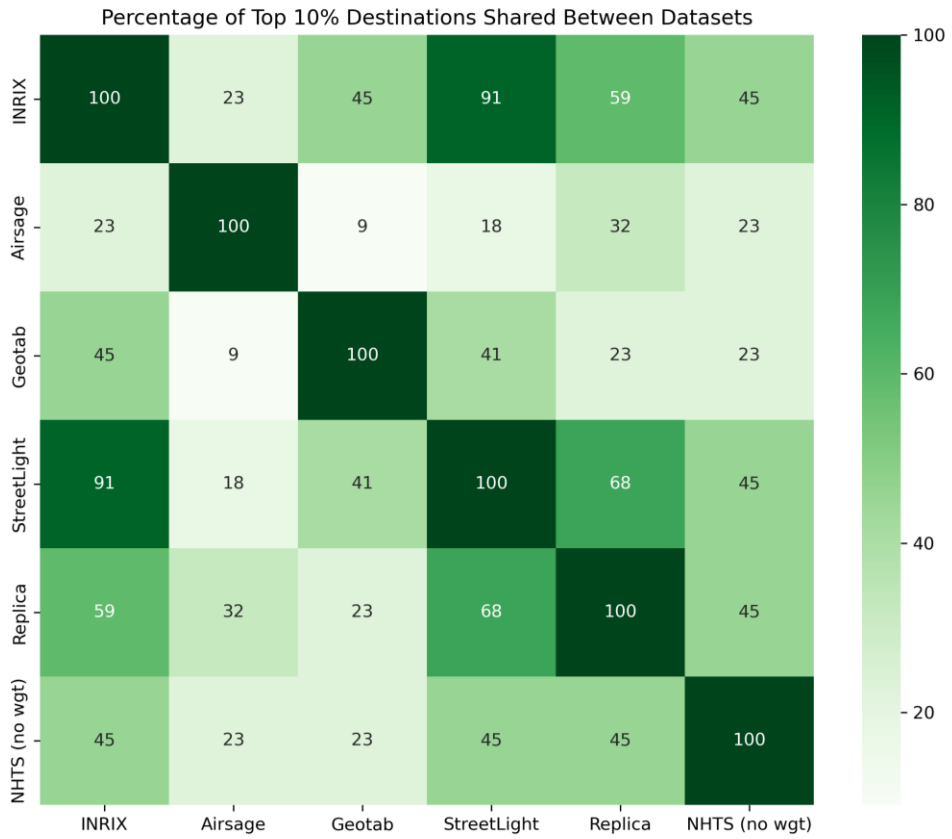

Figure 21 – Percentage of top 10% destination zones shared between sources in Richmond

Goochland County (Rural)

Goochland County was selected as a county that reflects rural travel behavior. While this is an important perspective to capture, it should be noted that in contrast to Chesterfield and Richmond, *the subset of NHTS trips starting and ending in Goochland County is extremely small*, with just 77 unweighted trips in 2022. The results are presented below, *but this small sample size should be considered when comparing vendor results to NHTS.*

Figures 22-23 illustrate the distribution of travel time and distance for trips in Goochland County, respectively. Figure 22 shows that at least 90% of NHTS travel time observations are less than or equal to 25 minutes, with the peak in the 10–15-minute range. The shapes of vendor distributions are all slightly different, with one peaking at 0-5 minutes, another at 5-10 minutes, and the third at 10-15 minutes. Figure 23 shows that compared to more urban areas such as Chesterfield County and Richmond, Goochland County exhibits a greater number of long-distance trips (while also showing a high number of short trips), indicating that residents may need to travel to more distant locations due to the rural nature of the area. Although the trip distance distribution is different for each vendor, the presence of longer trip distances was observed across datasets.

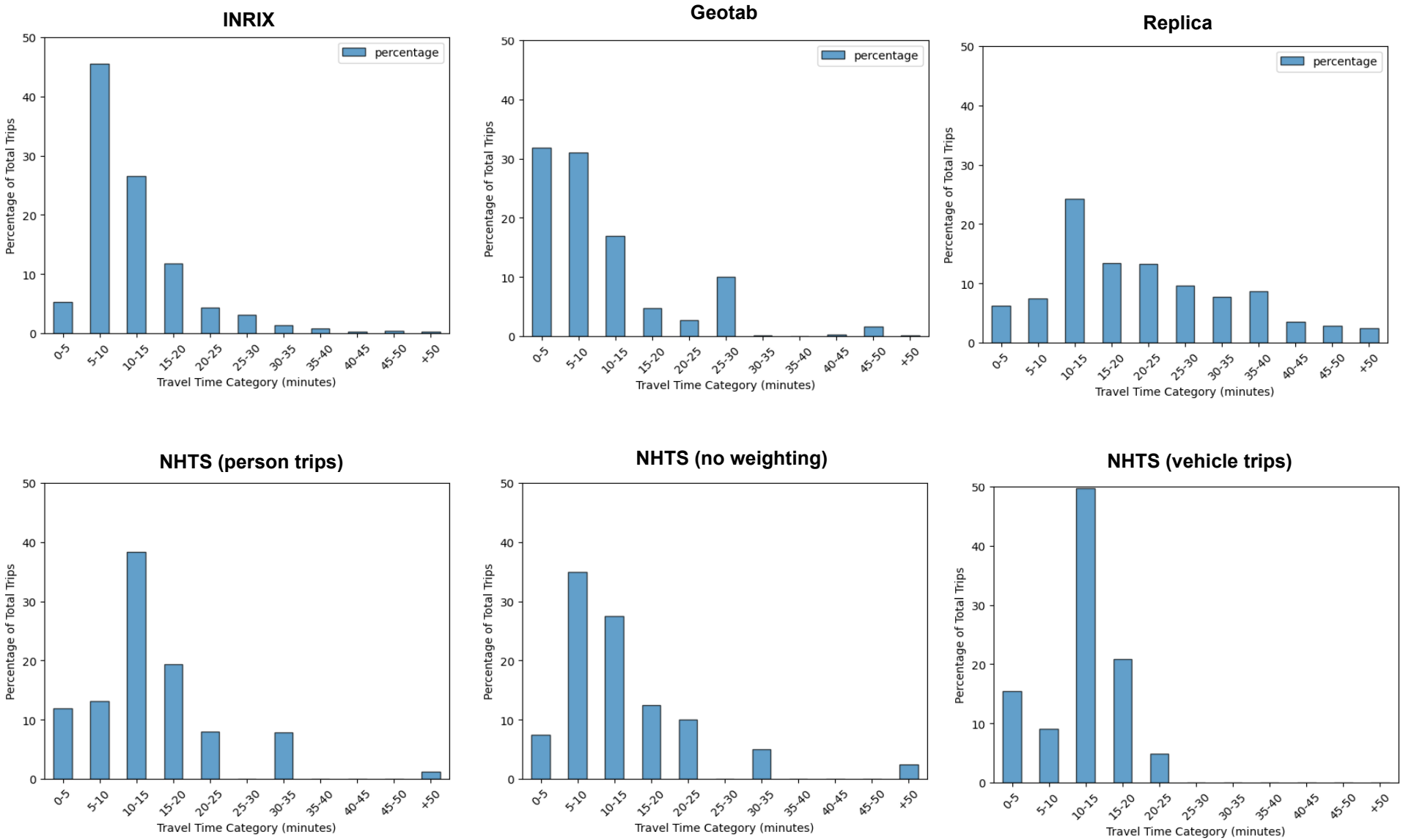

Figure 22 – Travel time distributions for trips in Goochland County.

[3] AirSage and StreetLight travel times are not included due to how their data was queried at the time of data collection. However, trave time information is available for both vendors' OD products.

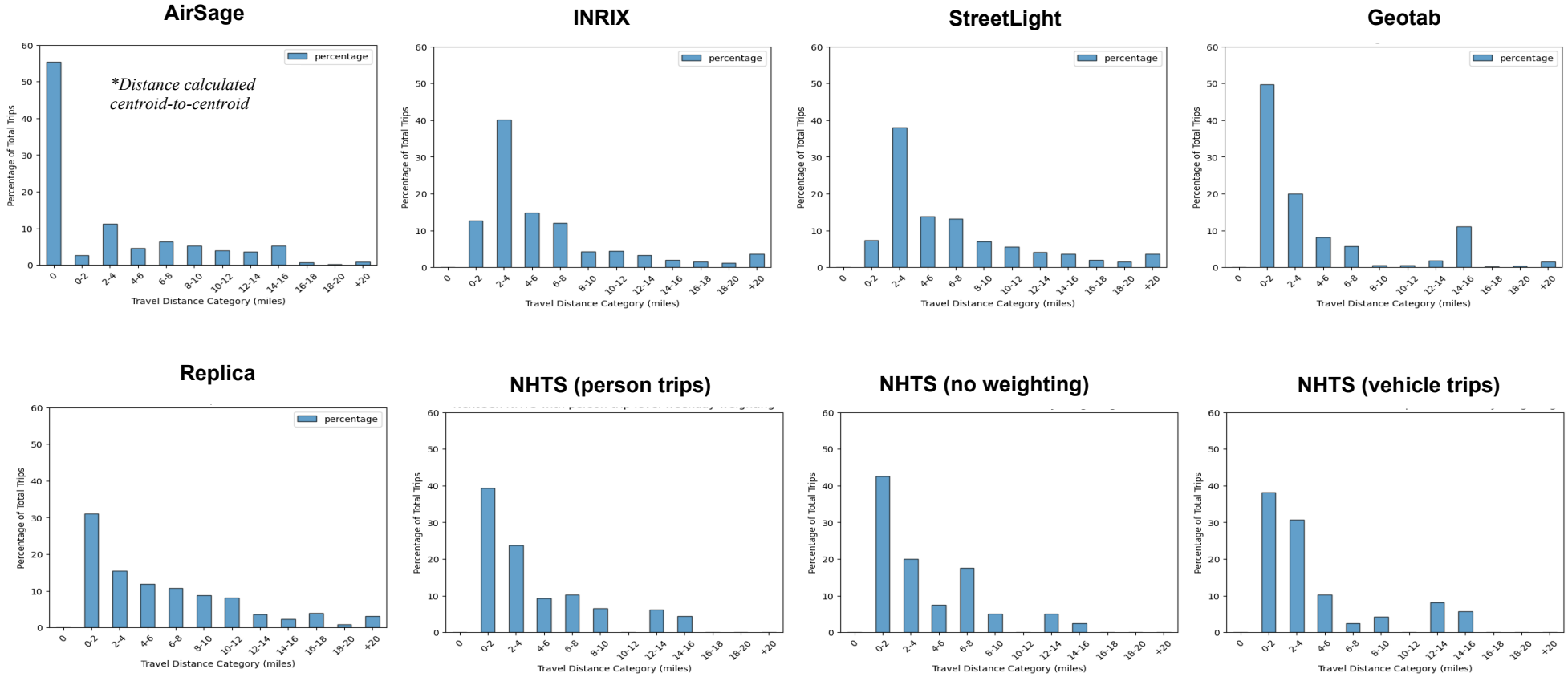

Figure 23 – Travel distance distributions for trips in Goochland County.

Top Origins and Destinations

Table 6 summarizes each dataset's coverage in terms of number (or percentage) of unique origin zones, destination zones, and OD zone pairs in Goochland County. Unlike the previous two examples (Chesterfield County and Richmond), which had 242 and 219 TAZs, Goochland only has 30 TAZs. Of the total 30 TAZs in Goochland County, most vendors' datasets contain close to the maximum possible unique origin and destination TAZs – both during typical weekday and weekend periods -- with the exception being Geotab, which has trips starting and endings in about 83% of TAZs on typical weekdays. The NHTS is lower during typical weekdays (about 47% of unique origin and destination TAZs, respectively), and drops further to about 27% for O's and D's during weekend periods.

Additionally, there are 900 possible OD pairs (30 * 30 TAZs) in Goochland County, and vendor datasets show a wide range of observed pairs. Replica's dataset showed the most unique OD pairs (88% and 83% for typical weekdays and weekends, respectively), followed by INRIX (79% and 73%), StreetLight (67% and 66%), AirSage (28% and 23%), and Geotab (6% for typical weekdays). For the NHTS dataset, 2-4% of OD pairs were captured. *Again, this is meant to be descriptive, rather than a measure of accuracy*. There are not necessarily trips between every possible – or even the majority of -- OD pairs.

Table 6 – Summary of unique O's, D's, and O-D pairs for each data source in Goochland County

Data Source	Day Type	Unique Origin TAZs*	Unique Destination TAZs*	Unique OD Pair TAZs**
AirSage	Typical Weekday	30 (100%)	30 (100%)	249 (28%)
	Weekend	29 (97%)	29 (97%)	209 (23%)
INRIX	Typical Weekday	30 (100%)	30 (100%)	710 (79%)
	Weekend	30 (100%)	30 (100%)	658 (73%)
StreetLight	Typical Weekday	30 (100%)	30 (100%)	603 (67%)
	Weekend	30 (100%)	30 (100%)	593 (66%)
Geotab	Typical Weekday	25 (83%)	25 (83%)	58 (6%)
Replica	Typical Weekday	30 (100%)	30 (100%)	791 (88%)
	Weekend	30 (100%)	30 (100%)	750 (83%)
NHTS (Unweighted)	Typical Weekday	12 (40%)	12 (40%)	18 (2%)
	Weekend	8 (27%)	8 (27%)	8 (1%)

* 30 total TAZs in Goochland County
** 900 total OD Pairs (=30*30)

Figures 24 and 25 illustrate the top 5 origin and destination TAZs in terms of the number of trips reported by vendors and the NHTS dataset (unweighted) in Goochland County. Note that the top 5 zones were selected because the top 10% -- as was used for the other counties – only represents 3 zones in Goochland. These maps show that while there is some slight overlap between top O's or D's between sources, clear visual patterns are not detected. However, as noted previously for Chesterfield and Richmond, trip counts are bunched closely together, and minor differences can alter the ranking of top origins or destinations.

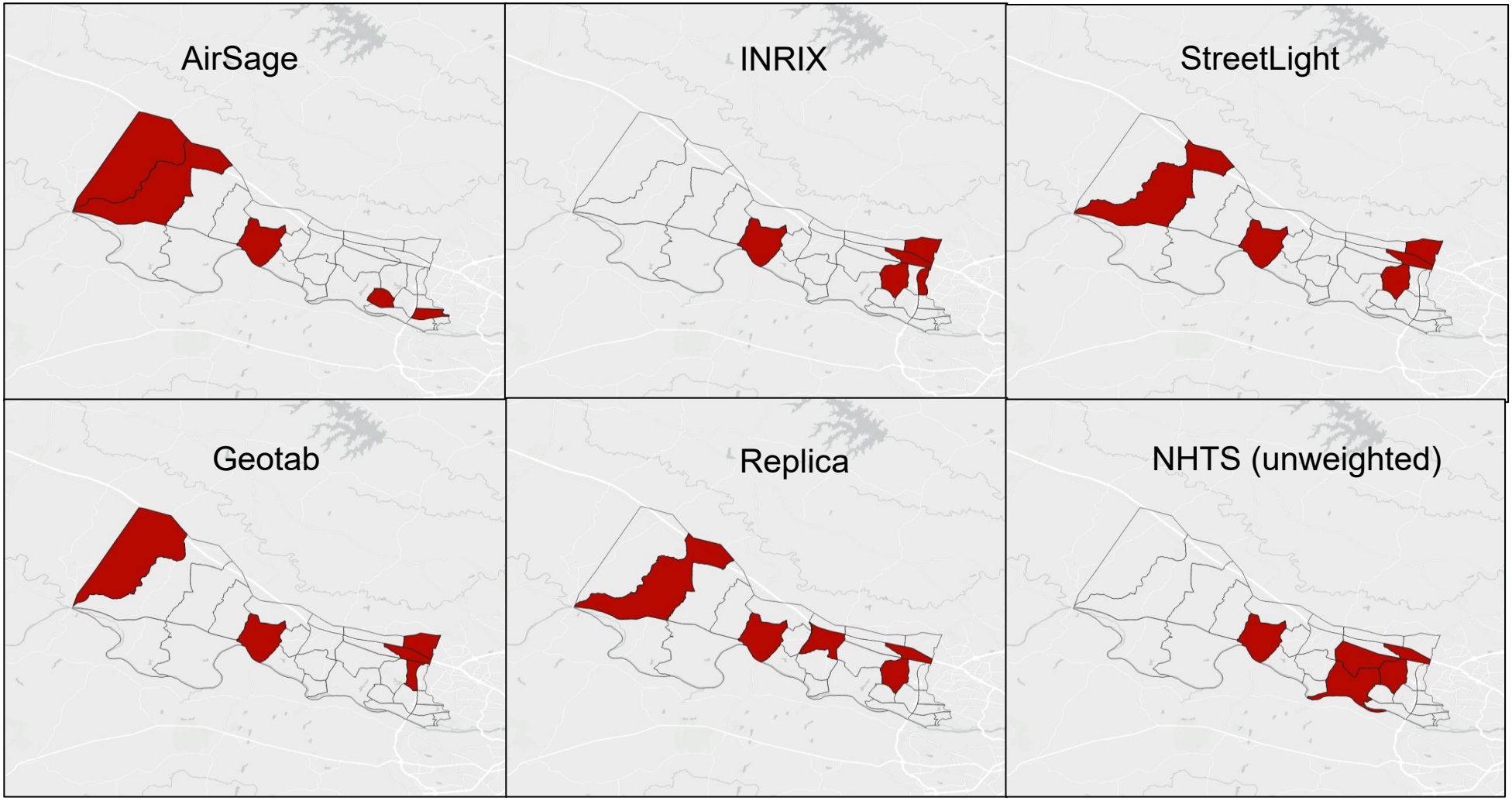

Figure 24 – Map view of top origins for trips in Goochland County

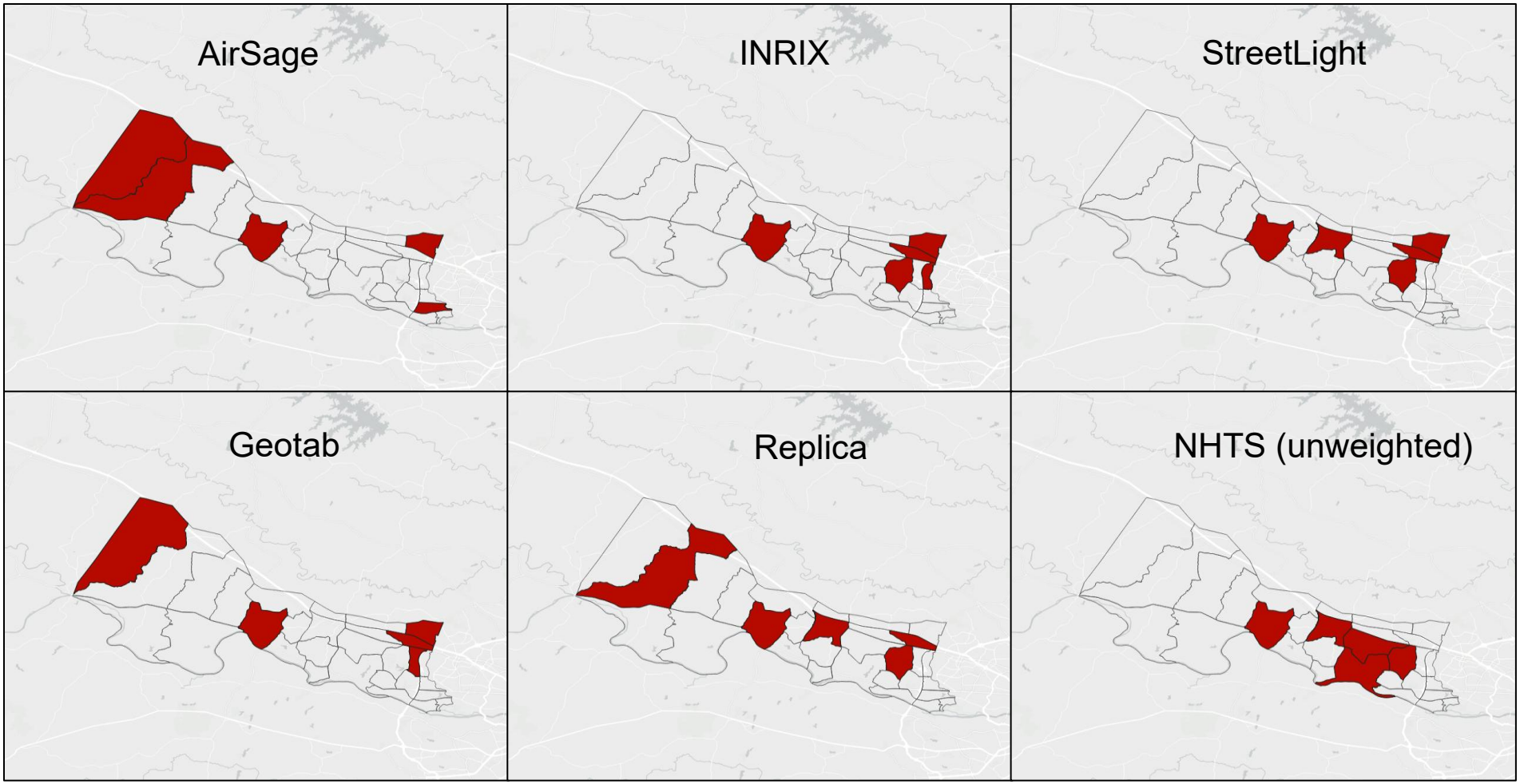

Figure 25 – Map view of top destinations for trips in Goochland County

Conclusions

This validation report represents Phase II of the Eastern Transportation Coalition's first Origin-Destination validation activity, analyzing OD data from 5 TDM vendors (AirSage, Geotab, INRIX, Streetlight, and Replica) and the National Household Travel Survey (NHTS) 'add-on' dataset in Richmond, VA. Whereas the Phase I analysis was focused on the full dataset and county-level flows, the analysis in this report digs into Traffic Analysis Zone (TAZ) level granularity for trips within specific areas. Several key takeaways emerged from the analysis:

- **Major differences between vendor product offerings make direct comparisons challenging.** This finding was highlighted in the Phase I report, but is worth re-emphasizing here. Key differences between vendor products include underlying data sources (connected vehicle data vs smart phone location-based services (LBS) data), the types of vehicles captured (e.g., mixture of all vehicles, freight only), how trips are quantified (e.g., person-trips vs vehicle-trips), trip mode, and whether reported trip counts reflect observed sample probes or population-level estimates. Furthermore, each vendor uses slightly different logic to split GPS waypoints into separate trips, so the same GPS trajectory may result in different trip definitions across vendors. These nuances should be taken into consideration when comparing vendor results to each other or to external data sources. As a result, *even when vendor products disagree with each other – or with an external data source—they may still provide meaningful information.*
    - AirSage reports **person-trips** that reflect population estimates.
    - Geotab reports observed **commercial freight vehicle-trips**
    - INRIX reports observed **vehicle-trips** (mixture of passenger and freight CV data)
    - Streetlight reports **vehicle-trips** that reflect population estimates (CV-data).
    - Replica reports both **vehicle-trips** and **person-trips** that reflect population estimates. For this study, only person-trips are included.

- **The NHTS dataset provides a useful point of comparison, particularly for evaluating travel time and trip length distributions.** Many planners are familiar with the NHTS dataset and its characteristics, so although it does not reflect 'ground truth', it acts as a useful comparative dataset. The ability to consider person-trips and vehicle-trips separately, plus the ability to apply weights to represent population-level behavior, helps facilitate more reasonable comparisons with vendor data.

  However, it appears that the NHTS dataset – particularly when using the weighting factors – is best suited for analyzing travel time and trip distance distributions, rather than identifying OD patterns at the granular TAZ level. Although unweighted NHTS data is used to identify OD patterns in this report, the small number of observations at this detailed spatial scale – particularly in rural areas -- makes it difficult to determine the most important O's and D's.

- **Vendor Travel Time and Trip Length distributions appear reasonable and shifted intuitively across counties**. When viewed individually, vendors' travel time and trip length distributions were generally intuitive, with trip lengths tending to include smaller values in the urban context (city of Richmond) and larger values in rural conditions (Goochland county) – an encouraging result.

Although results all appeared reasonable, the shape of vendor distributions often differed slightly from each other and with respect to the NHTS dataset. These differences also varied across counties. Given that these datasets represent diverse products (as noted in the first point above), such differences are expected.

- **Top OD patterns are difficult to compare at the TAZ level.** Unlike county-level OD flows explored in Phase I (where the same top Origin and Destination counties were common across vendor datasets), top O's and D's were more challenging to distinguish at the more granular TAZ level. Given that there was often little separation in the top trip-producing or attracting zones, the most effective approach for comparison was to check for the percentage agreement in the top 10% origins and destinations for each data source.

  This analysis showed varying levels of overlap in top origins and destinations between different vendors and NHTS data. Across all datasets, StreetLight and INRIX had the highest degree of overlap with each other and to a slightly lesser degree, with Replica. However, differences in top O and D rankings between datasets are expected given the distinct underlying data products and the small size of TAZ geographies.

In summary, although this analysis does not label vendor products as 'sufficiently accurate' or 'inaccurate' – owing to significant differences in underlying data sources and product characteristics, plus the difficulty of establishing 'ground truth' benchmarks -- the differences observed in the TAZ-level OD analysis appear consistent with self-reported data characteristics described in the Phase I report.

Users of the data are encouraged to (1) explore the value and usability of the data as compared to survey data collection mechanisms, and (2) take note of the approach and characteristics of each vendor, including both the source data (person-trips, vehicle-trips, or freight trips), and the various methods of processing (reporting sample trips only, extrapolating to the population, and how trips are defined based on underlying trajectories). These parameters vary by vendor, and in some cases are configurable, so it is important that agencies understand the data characteristics to make sure they are suitable for intended use cases.

Next Steps

Moving forward, the validation team recommends that the next study focus on a smaller study area – perhaps a corridor or other small geography – for which independent data can be collected as a point of comparison. This would be in line with guidance developed in coordination with the Coalition's Technical Advisory Committee (see TDM-VAL-3) which recommended assessing datasets at a variety of spatial scales.