EVALUATION OF BIG DATA PLATFORMS

for Regional Transportation Applications at COG/TPB

Zhuo Yang
Transportation Data Analyst

Mobility Analytics Subcommittee
March 26, 2026

National Capital Region
Transportation Planning Board

Outline

Introduction

General Comparisons

Programmatic Applications and Findings

Insights and Suggestions

National Capital Region
Transportation Planning Board

Introduction

- Big data has widely benefited industries with its rich data coverage.

- COG hired Kimley-Horn to evaluate big data products available on the market in 2020.

- COG subscribed for StreetLight and Replica data for two years to conduct more in-depth evaluations.

- COG has established an internal user group to evaluate big data across multiple programmatic areas.

- As the initial evaluation period concludes, the results could benefit members of the Mobility Analytics Subcommittee.

National Capital Region
Transportation Planning Board

General Comparisons

Topic	StreetLight	Replica	Teralytics	LOCUS
Data	Vehicle focused	Activity based	Cell-tower-based	LBS based
Applications	Broad	Broad	A few	A few
Accessibility	Summaries	Summaries, raw data	Summaries	Summaries
API & Automation	API for analysis cases	database	NA	NA
Tech Requirements	Easy for standard analyses	Friendly for both non-programmers and advanced analysts	Easy for predefined use cases	Easy for predefined use cases
Docs & Support	Online courses	Online documentations	Limited	Limited

Evaluation Flow

① Research Analysis Selection

② Ground-Truth Data Selection

③ Comparisons, Spatial and Temporal

Analyses Topics

- Origin-Destination (OD) Trips Analysis

  - County and Tract Level Trip Comparisons
  - Interactive Research Dashboard

- Multimodal Analysis

  - Mode Share Analysis
  - Transit Access and Egress Analysis

- AADT and VMT Comparisons

  - by Functional Class,
  - for External/through Counts
  - and by Jurisdiction

- Special Markets Research

  - Active Transportation
  - Survey Assistance
  - Intercity Transit Study

National Capital Region
Transportation Planning Board

Ground-Truth OD Trip Dataset Selection

- Candidates:

  - Public Use Microdata Sample (PUMS)
  - Census Transportation Planning Products (CTPP)
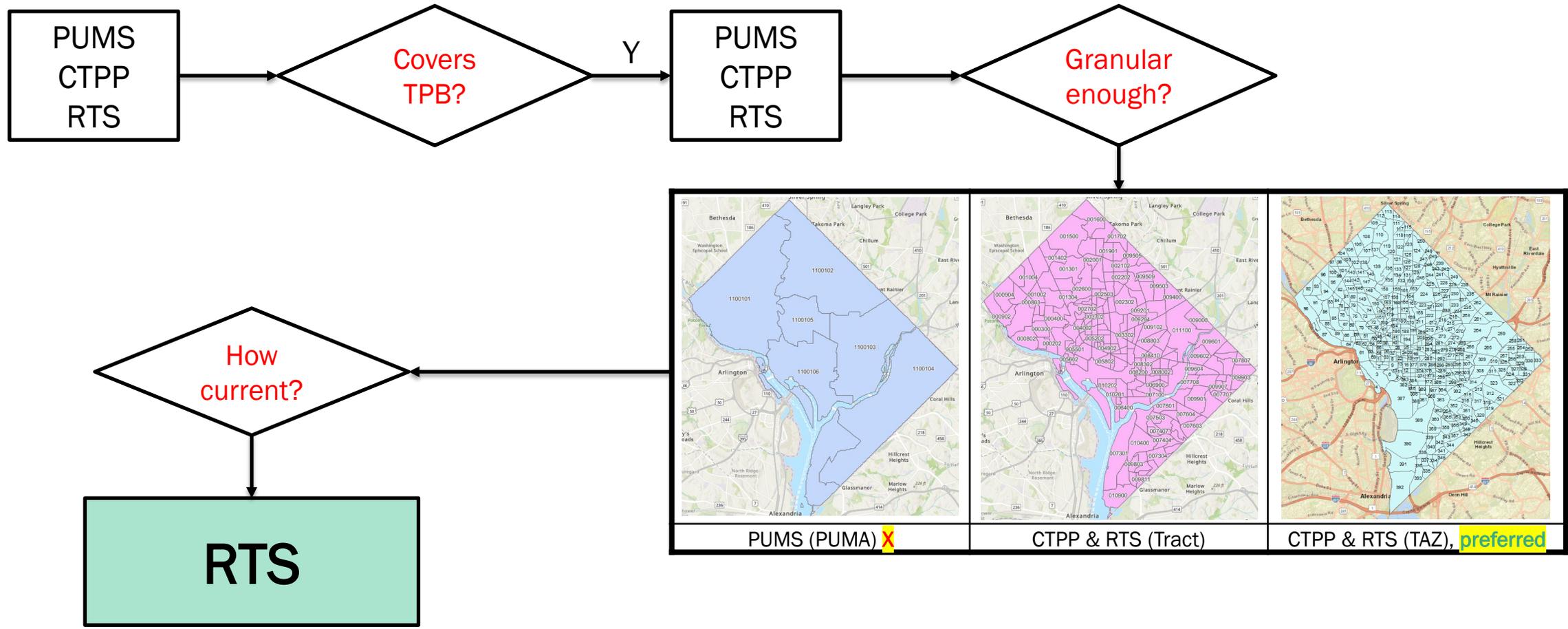 - Regional Travel Survey (RTS)

- Key Criteria:

  - Granularity
  - Coverage
  - Recency
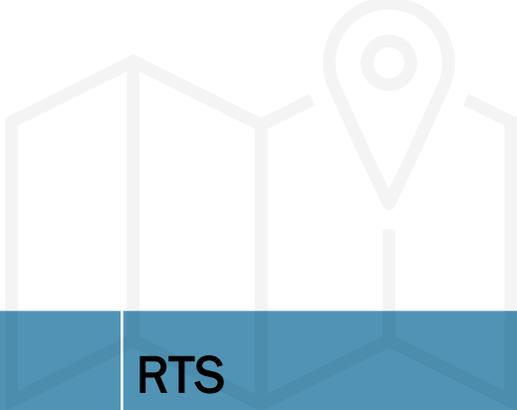
	PUMS	CTPP	RTS
Granularity	PUMA	TAZ	Address
Coverage	USA	USA	TPB modeled region
Recency	2022	2012-2016	2017-2018

National Capital Region
Transportation Planning Board

Selection Flow

Analysis 1. OD Comparisons

Settings:

- 2019 StreetLight and Replica were included in the comparisons

    - Teralytics: Lack of comparable data availability and granularity

    - LOCUS: Trial started after the analysis

- Vehicle Trips or Person Trips

    - STL is only available for Vehicle Trips

    - Both RTS and Replica were adapted to be comparable.

- Difference Ratio: $ratio_{g,b} = \dfrac{trips_{g,b} - trips_{g,r}}{trips_{g,r}}$

    g represents analysis geographic unit;
    b represents big data products;
    r represents RTS data.

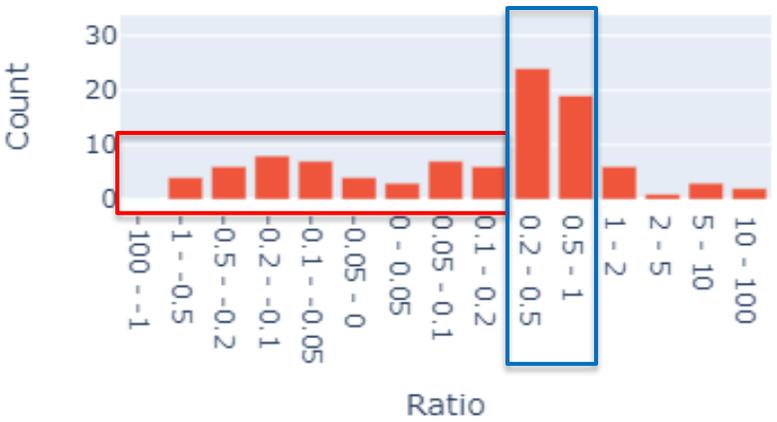- Both county-level and tract-level comparisons were examined.

Findings at the County Level

Difference in Ratio Compared to RTS2018
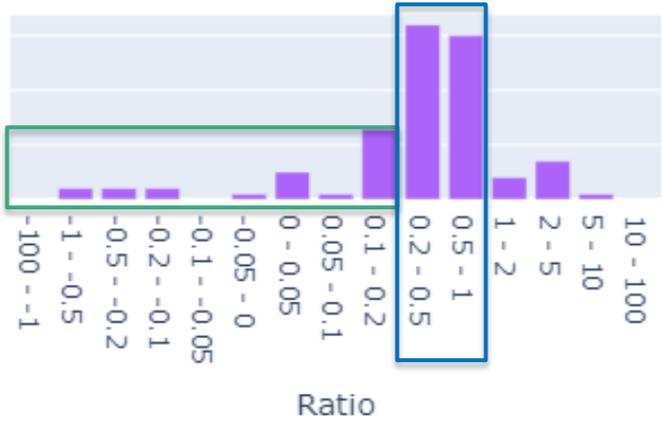
- **Ratio Concentration:** from 0.2 to 1
- **Variability:** Replica > STL
- **Max Deviation:** Replica > STL

Findings at the Tract Level

	Replica	StreetLight
Count	32,003	30,829
Median	-0.77	-0.73
Maximum	93	157
S.D	1.0	1.5

Statistics of Trip Difference Ratios between Big Data Products and RTS

Ratio Range	Replica	StreetLight
[-1, -0.5)	71%	66%
[-0.5, -0.2)	11%	11%
[-0.2, -0.1)	2%	3%
...		
[0.2, 0.5)	3%	4%
[0.5, 1)	3%	4%
[1, 2)	2%	3%
...		

Distributions of Difference Ratios

- **Matched Tract OD Pairs:** Replica > STL

- Deviations:
  - Both generally **underestimated** trips
  - Both products **diverged more** substantially
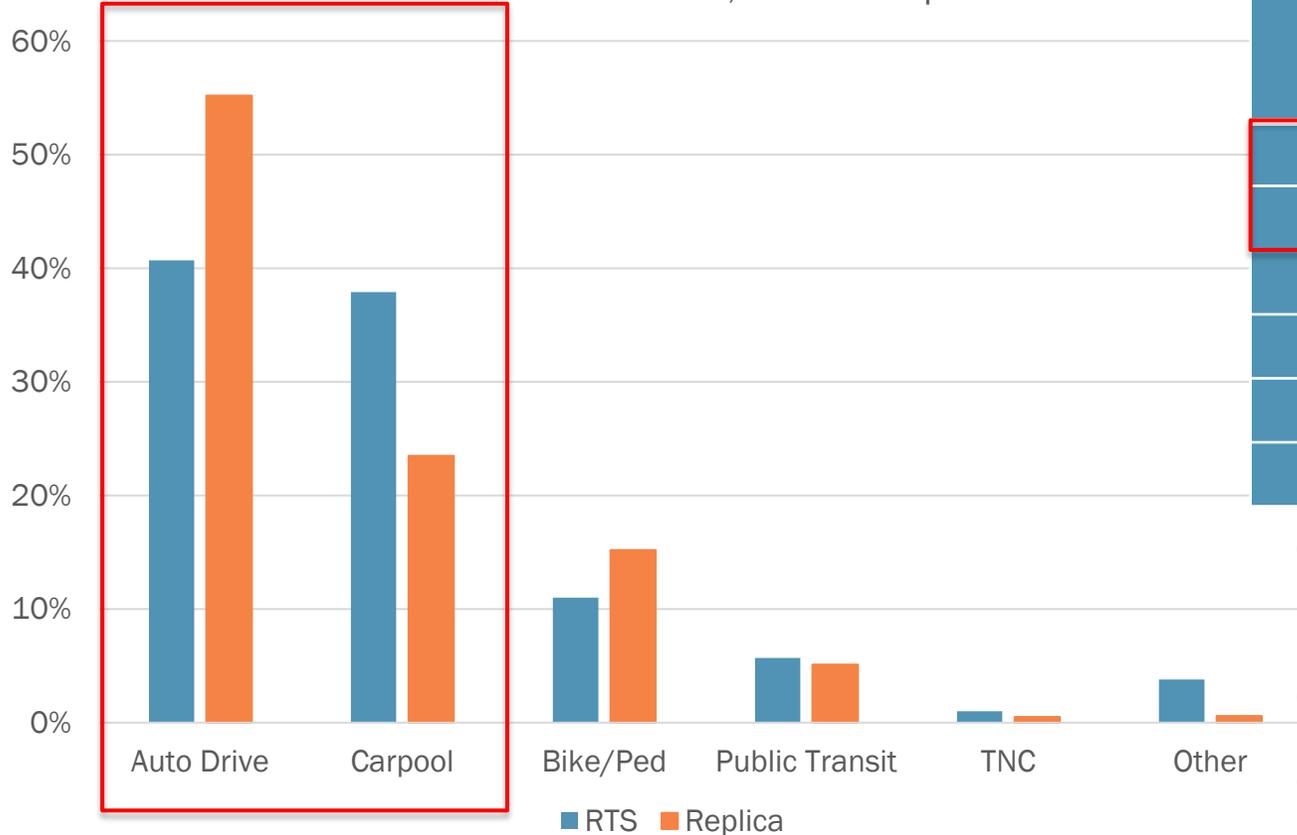
- **Variability:** STL > Replica

OD Explorer Dashboard
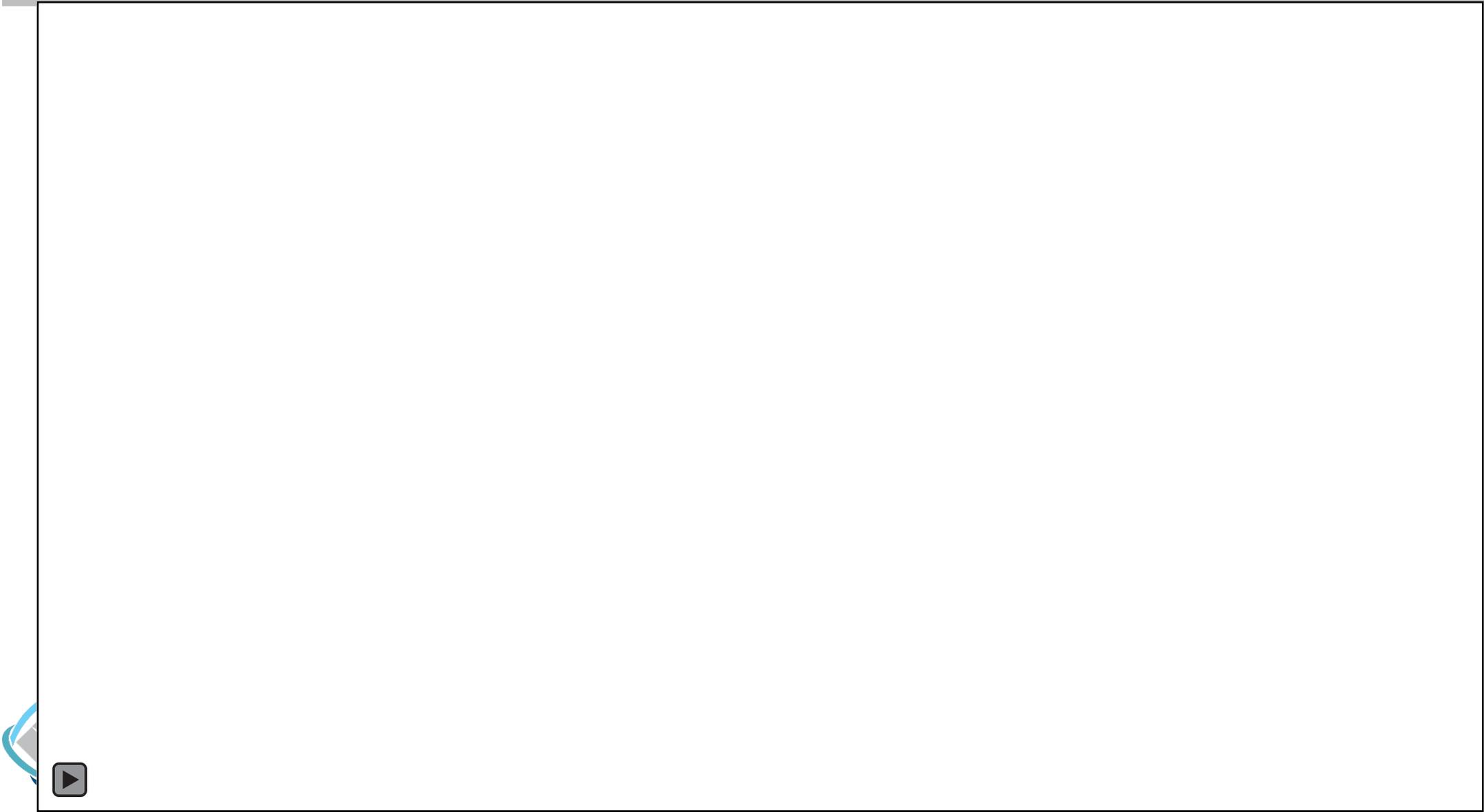
Analysis 2: Mode Share Findings

Mode Share Distribution, RTS v.s. Replica

Mode	RTS	Replica	Difference in Share
Auto Drive	40.7%	55.2%	14.5%
Carpool	37.9%	23.5%	-14.4%
Bike/Ped	11.0%	15.2%	4.2%
Public Transit	5.7%	5.1%	-0.6%
TNC	1.0%	0.5%	-0.5%
Other	3.8%	0.6%	-3.1%

- Only Replica was evaluated for mode share analysis.
- General auto mode is dominant.
- Auto Drive overestimate appears to come mainly from Carpool.

National Capital Region
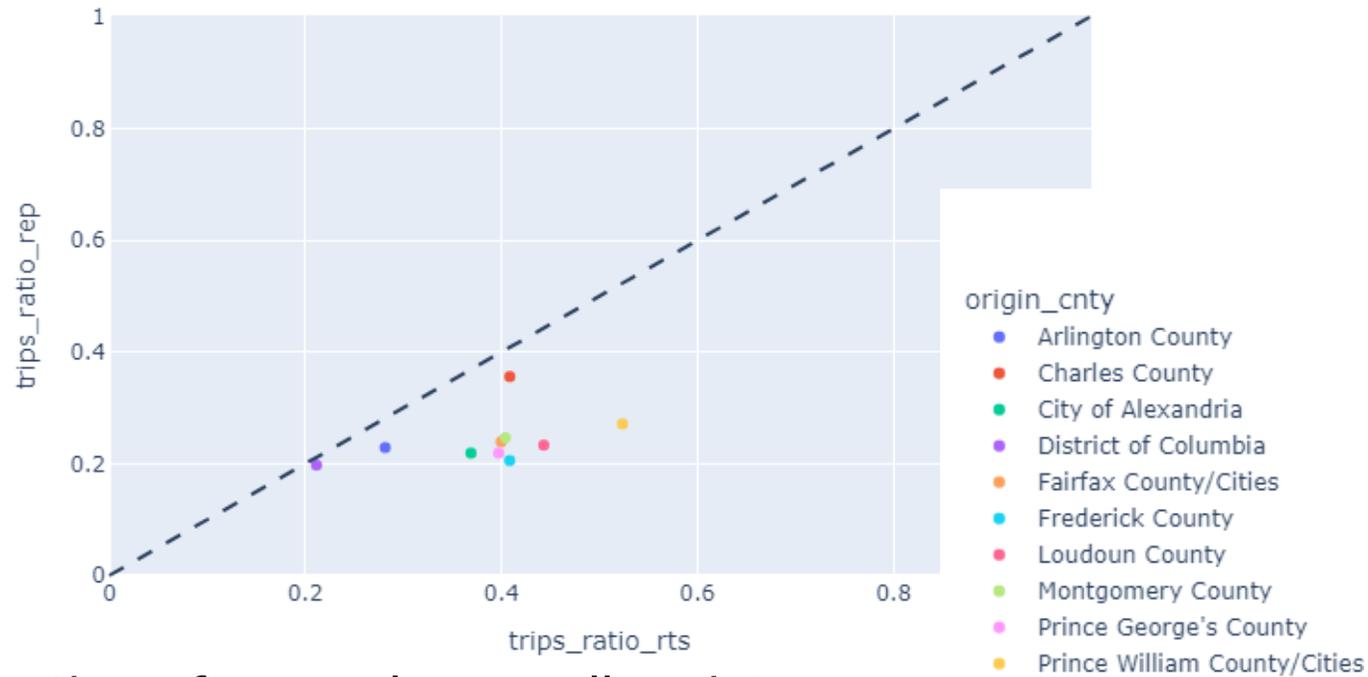**Transportation Planning Board**

Auto-related Mode Share, Jurisdiction Level

Mode Split Comparison by origin_cnty, RTS vs. Replica, Auto_Drive
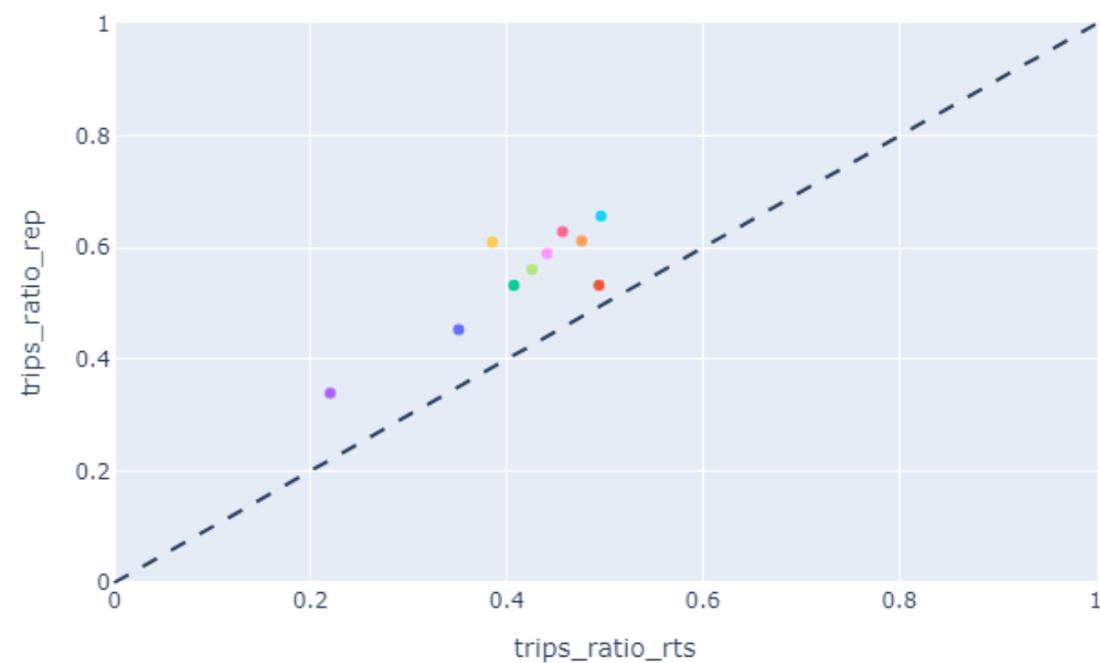
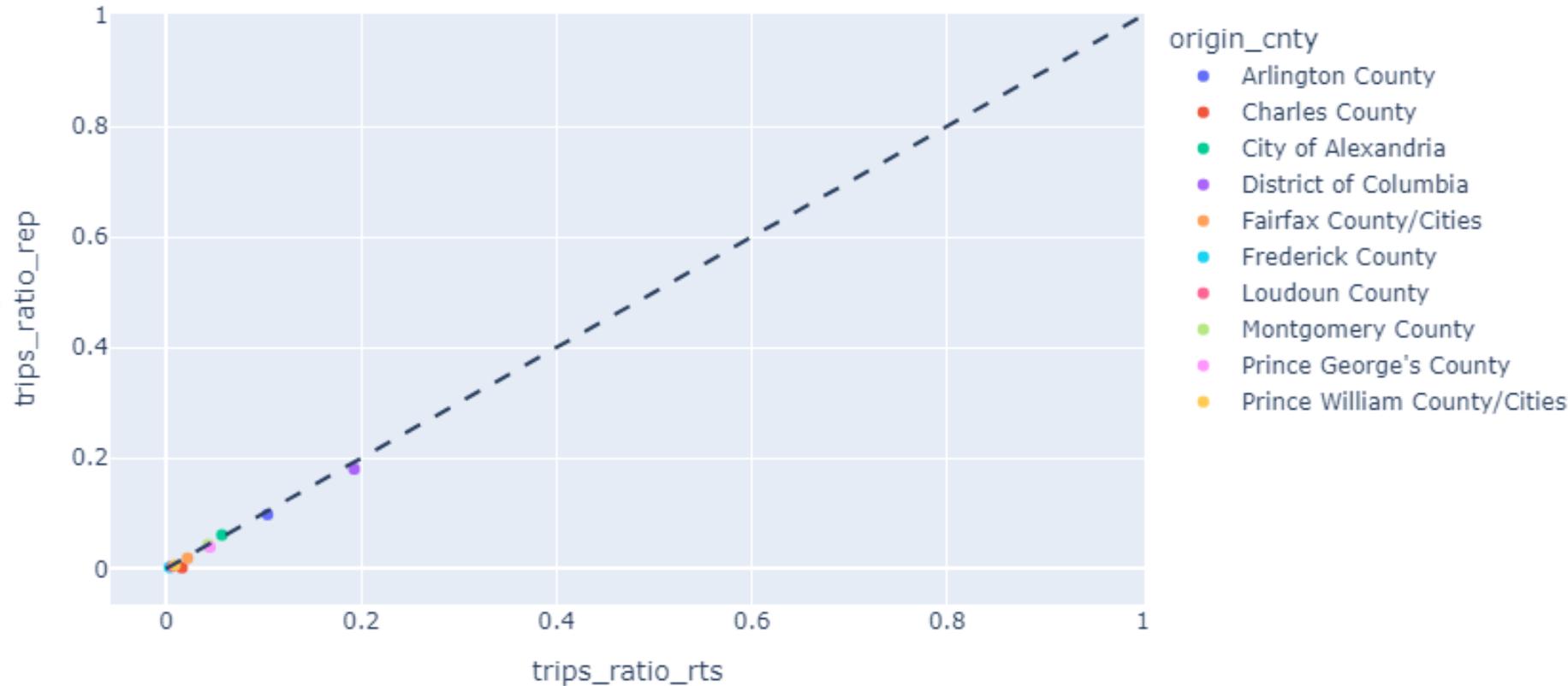Mode Split Comparison by origin_cnty, RTS vs. Replica, Carpool

origin_cnty
- Arlington County
- Charles County
- City of Alexandria
- District of Columbia
- Fairfax County/Cities
- Frederick County
- Loudoun County
- Montgomery County
- Prince George's County
- Prince William County/Cities

- Overestimation of auto drive and underestimation of carpool generally exist across all jurisdictions.

- Lack of carpool data causes the underestimation of carpool mode[1].

Source: Replica. *How does Replica calculate auto passenger (carpool) trips*? Replica Help Center.

National Capital Region
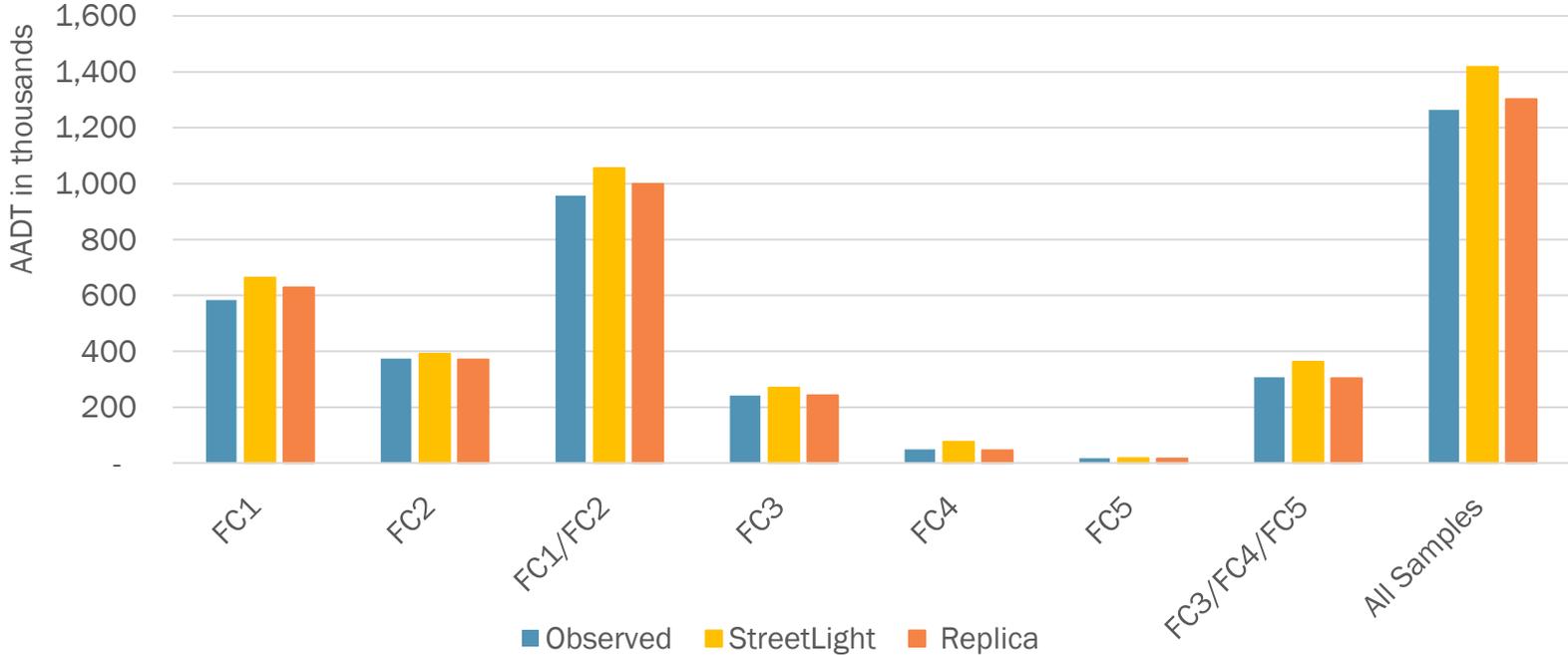**Transportation Planning Board**

Transit Mode Share, Jurisdiction Level

- Transit share is well estimated by Replica.

- District of Columbia has the highest share of transit trips.

Mode Split Comparison by origin_cnty, RTS vs. Replica, Public_Transit

origin_cnty
- Arlington County
- Charles County
- City of Alexandria
- District of Columbia
- Fairfax County/Cities
- Frederick County
- Loudoun County
- Montgomery County
- Prince George's County
- Prince William County/Cities

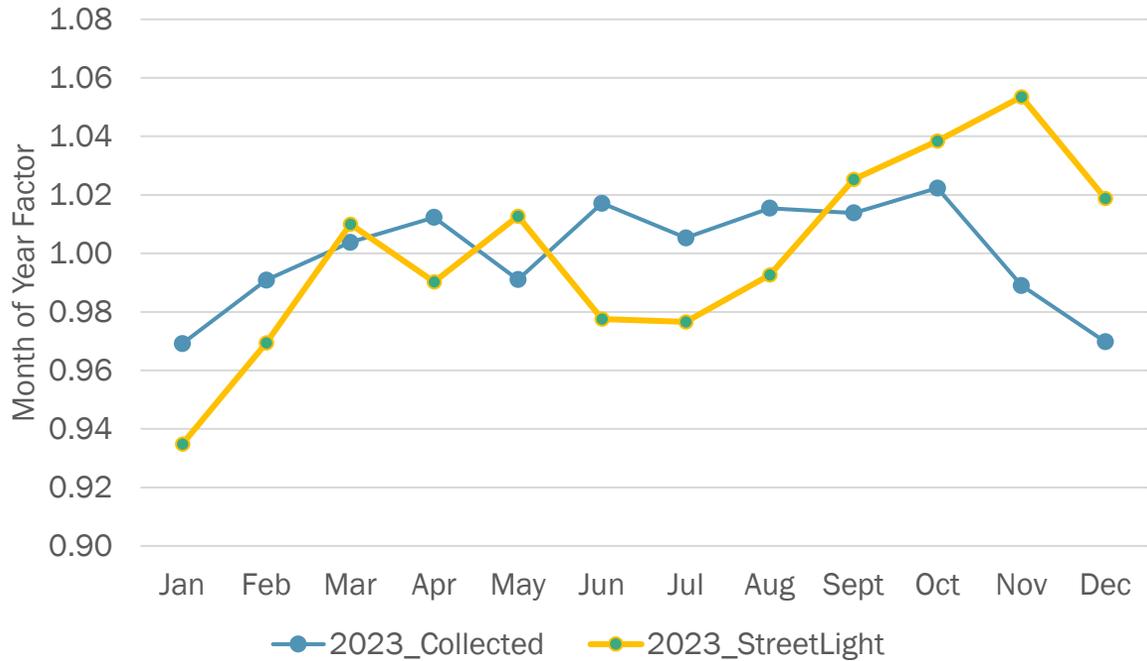# Analysis 3. AADT for Selected Roadways
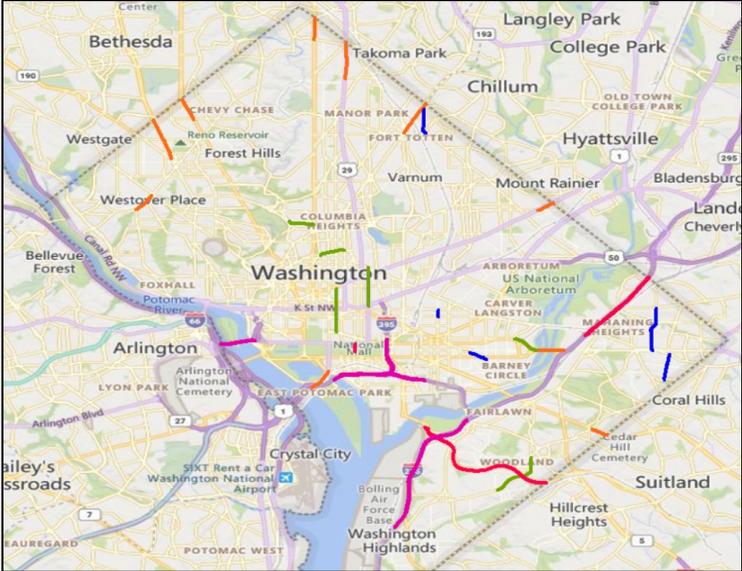
2023 AADT Comparison

- 31 road links were selected for comparisons with traffic counter data.

- Both products produced reasonable AADT estimates

- Replica was generally closer to observed counts.

MOY and DOW Factors by StreetLight
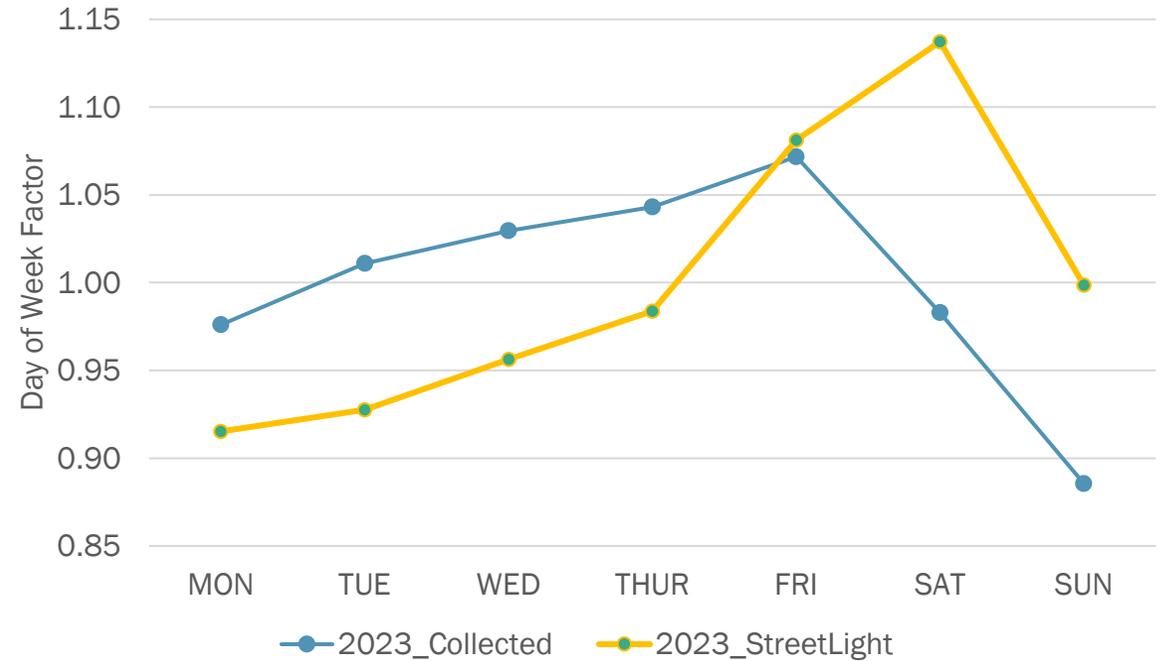
- Day-of-Week and Month-of-Year traffic patterns are not very reasonable.

- MOY Factors underestimated for summer while overestimated for winter.

- DOW Factors are overestimated for weekends.

Annual Growth Factor by Replica

- Replica is reasonable for lower FC but a little off for higher FC.

- Negative growth factors are questionable compared to observed data.

Functional Class	Replica	Observed
FC1/FC2	-2.14%	2.94%
FC3/FC4/FC5	0.93%	2.86%
All Samples	-1.44%	2.92%

National Capital Region
Transportation Planning Board

Analysis 4. AAWDT by External Counters
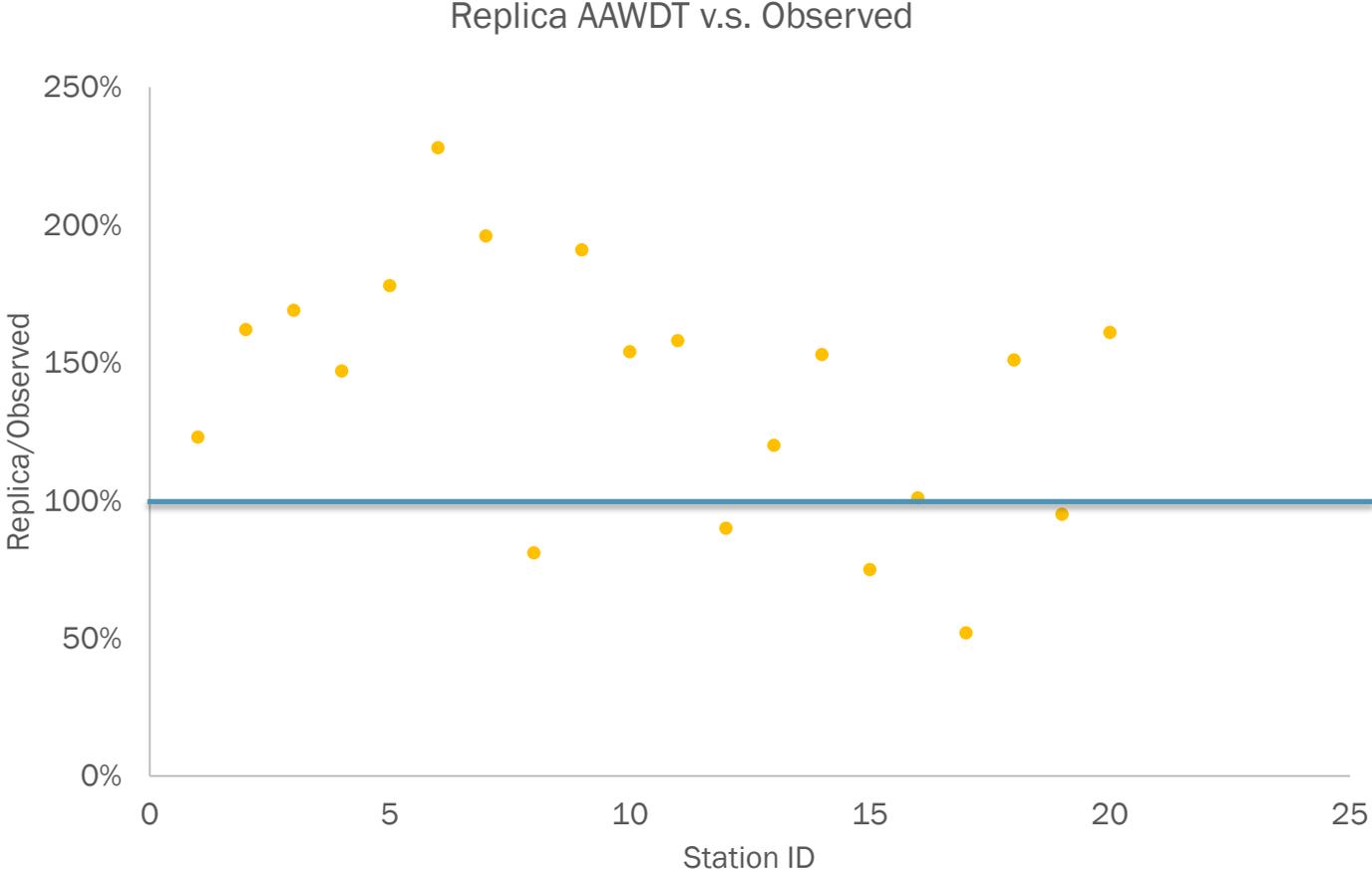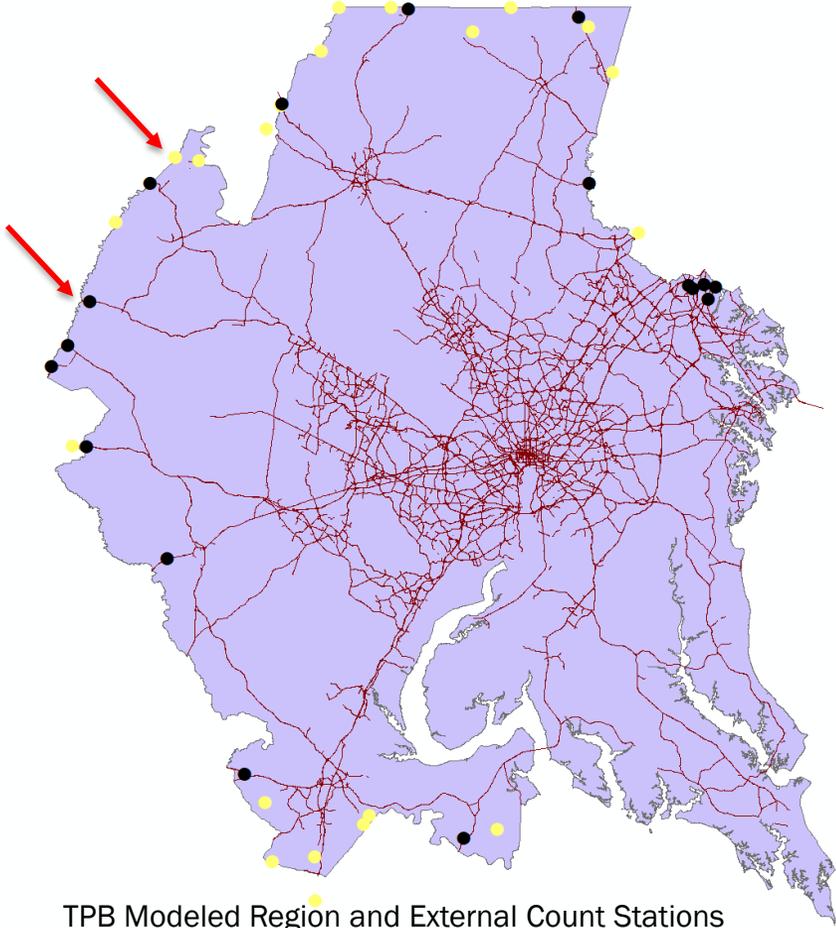
TPB Modeled Region and External Count Stations

Replica AAWDT v.s. Observed

National Capital Region
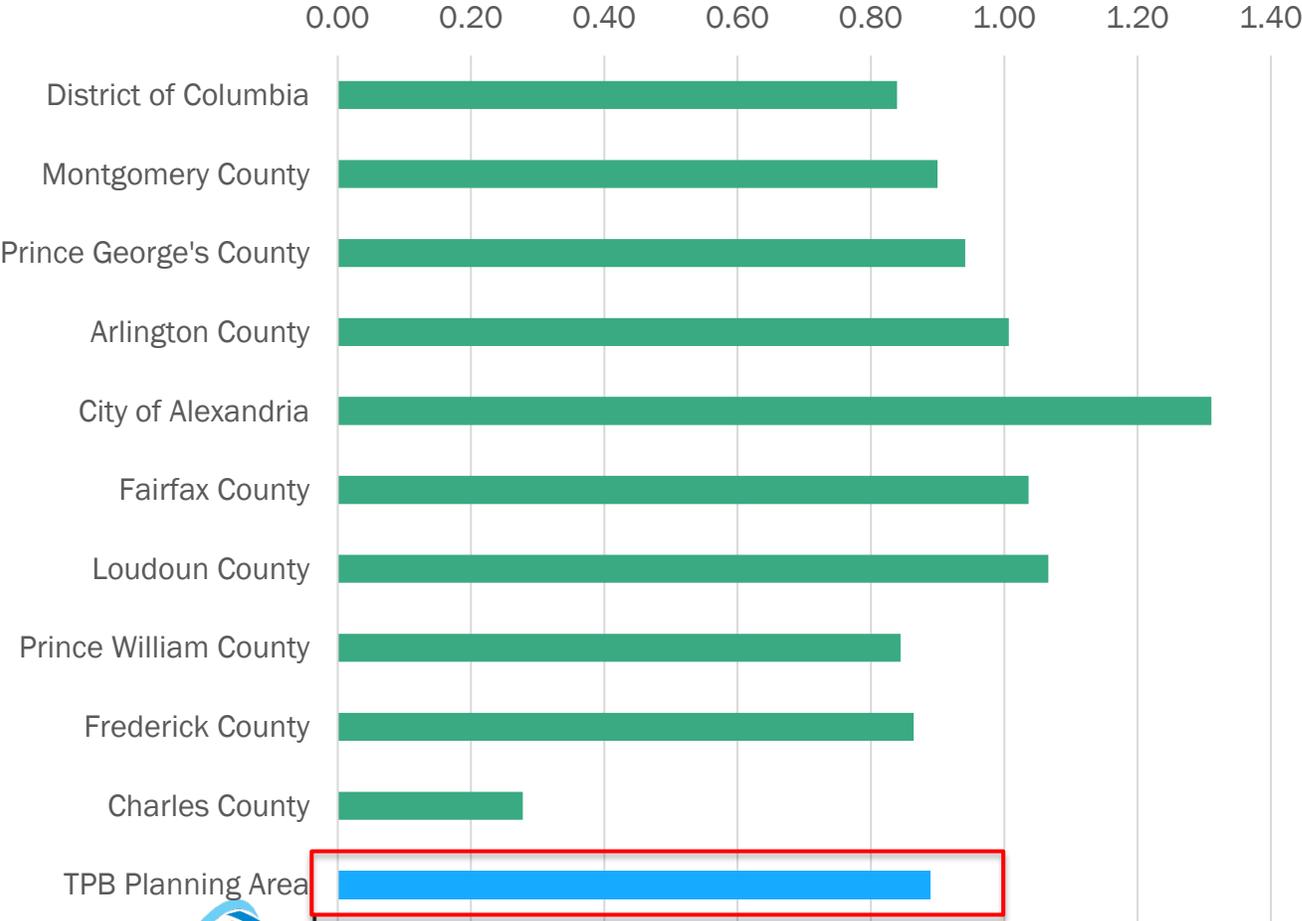**Transportation Planning Board**

Analysis Findings

- Only Replica was analyzed, as it uniquely covers the entire TPB modeled region.

- Stations may not be able to be matched due to the inconsistencies in the road network.

- Volumes at matched locations show a wide range of discrepancies, spanning from <u>52%</u> to <u>228%</u>.

- Replica's external and through-trip data should not be considered "ground-truth" data.
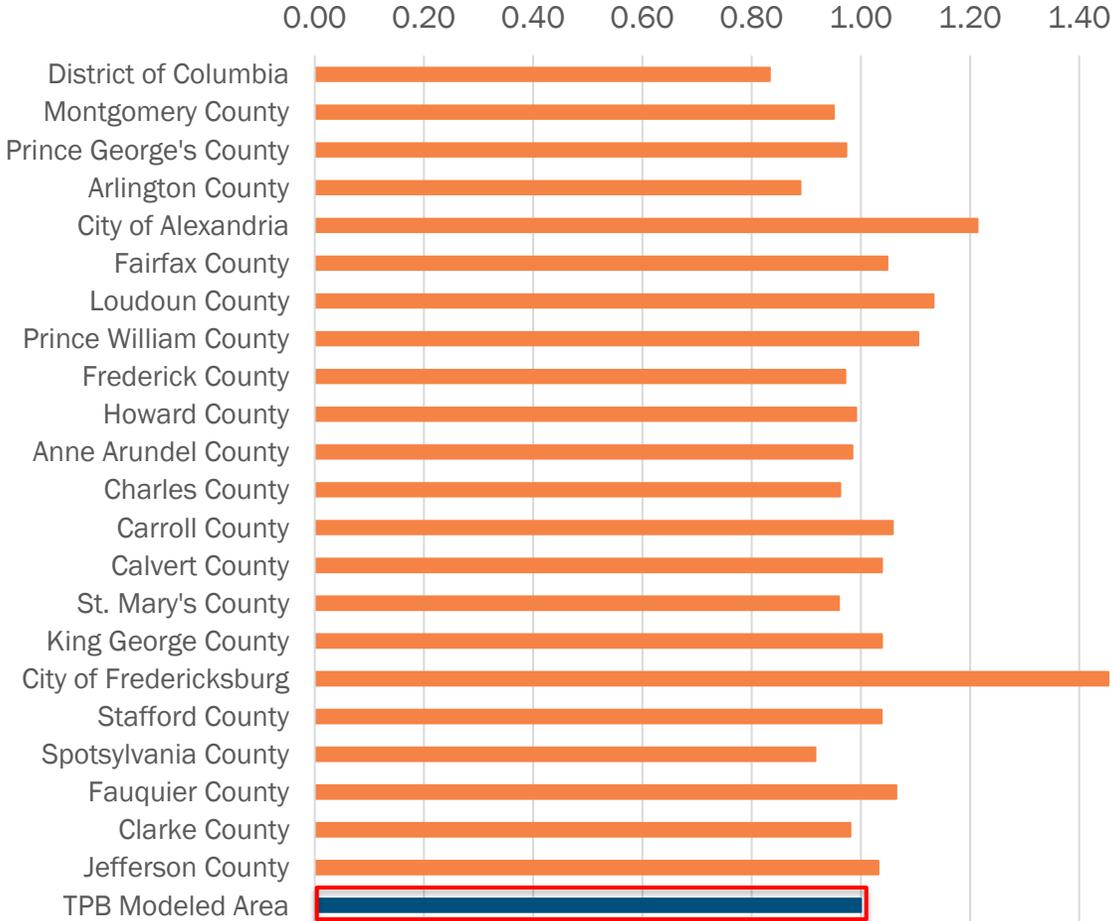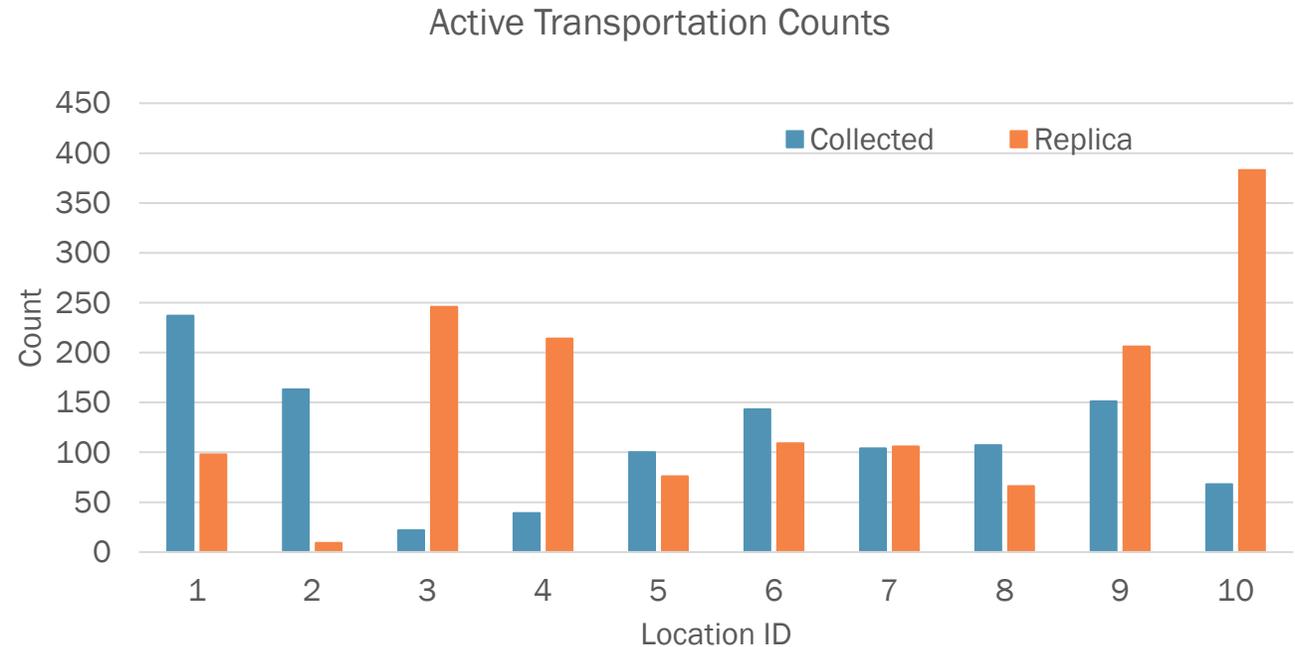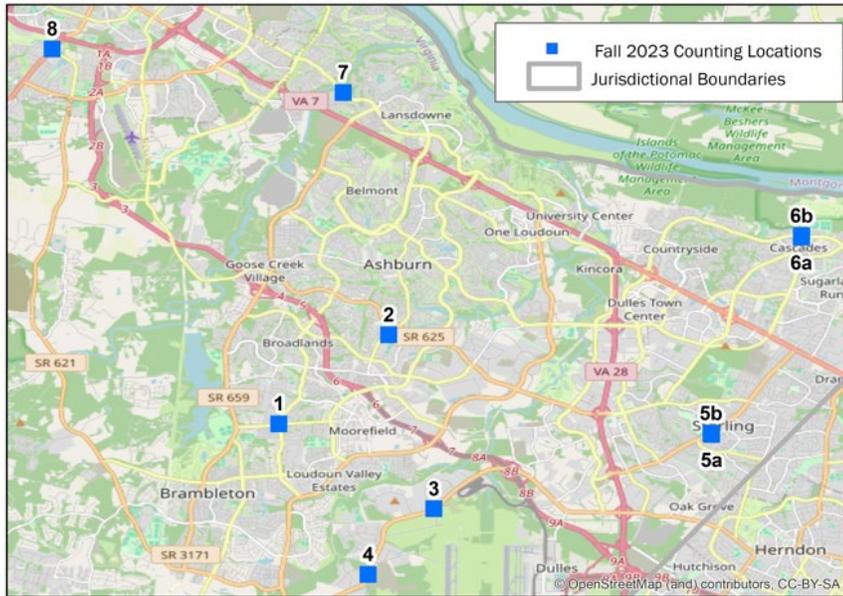
Analysis 5. VMT by Jurisdiction

VMT Ratios of Teralytics over HPMS by Jurisdiction, 2024

VMT Ratios of Replica over HPMS by Jurisdiction, 2019

Agenda Item 3: Evaluation of Big Data Platforms for Regional Transportation Applications at COG/TPB
March 26, 2026

National Capital Region
Transportation Planning Board

Analysis 6. Active Transportation

Active Transportation Counts

- Replica provides more useful support for active transportation than StreetLight.

- Replica's active transportation volumes were generally more reasonable at an aggregate level than at an individual level.

- Validation against observed counts remains essential for planning.

Analysis 7. Survey Assistance

Trip Counts and Modes Could Inform Sample Plan Development

- It provides a comprehensive picture of where trips are occurring in the region prior to survey data collection, allowing stratification targets to be established.

- Non-auto travel modes such as transit, biking, and walking are often underrepresented in travel surveys.

- Areas with a high concentration of non-auto trips can be oversampled to ensure they are sufficiently captured.
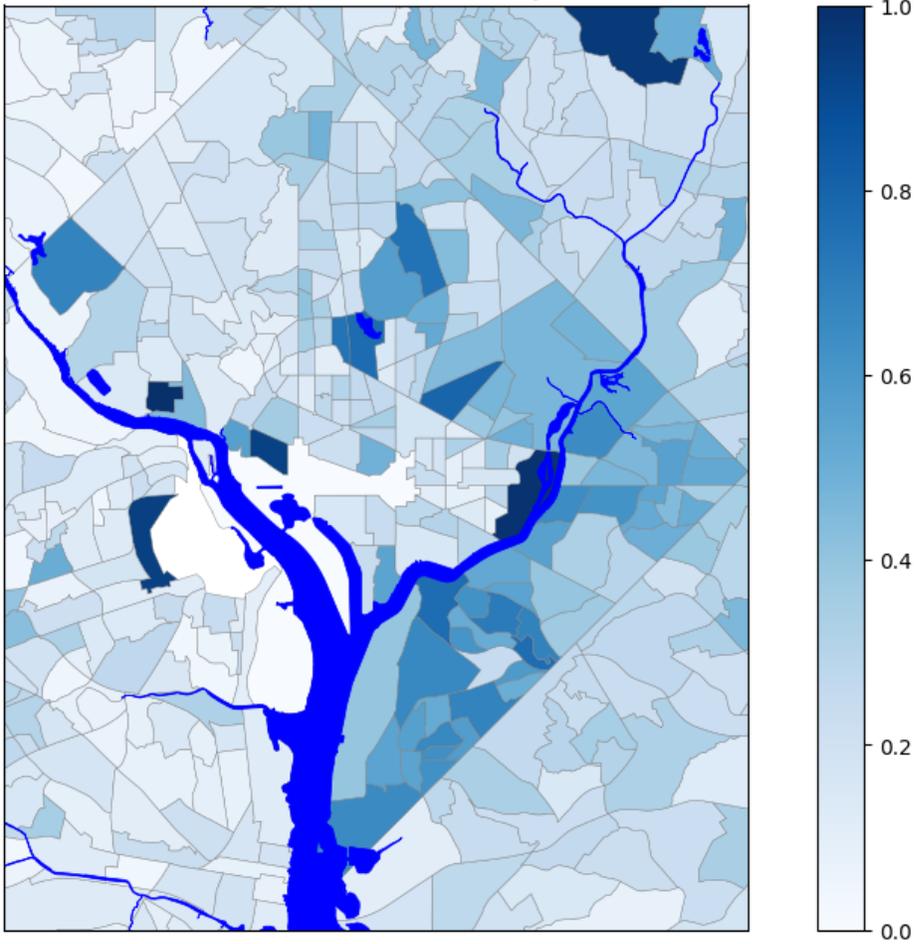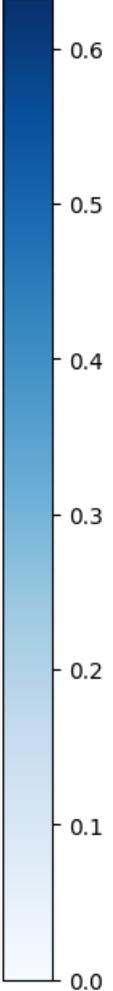
Share of Households with Income < $40,000 by Tract

Share of Households with Income <= $40,000 by Tract by Tract
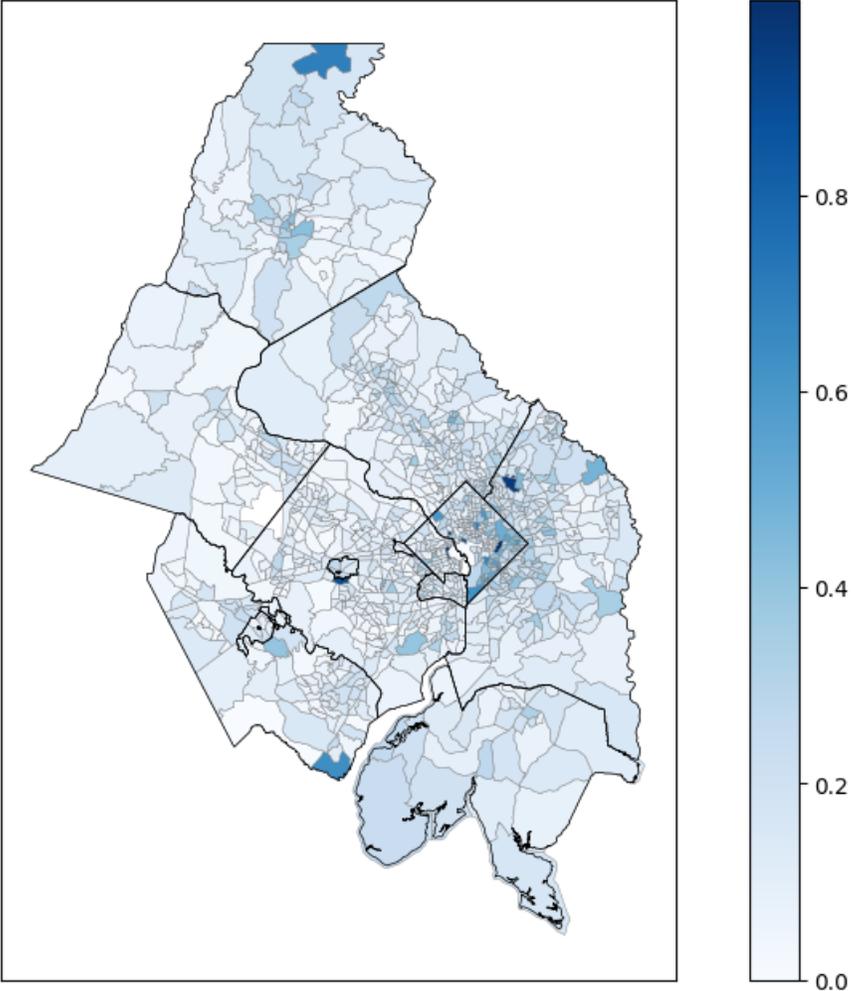
Share of Households with Income <= $40,000 by Tract (Zoomed to DC)

Transportation Planning Board
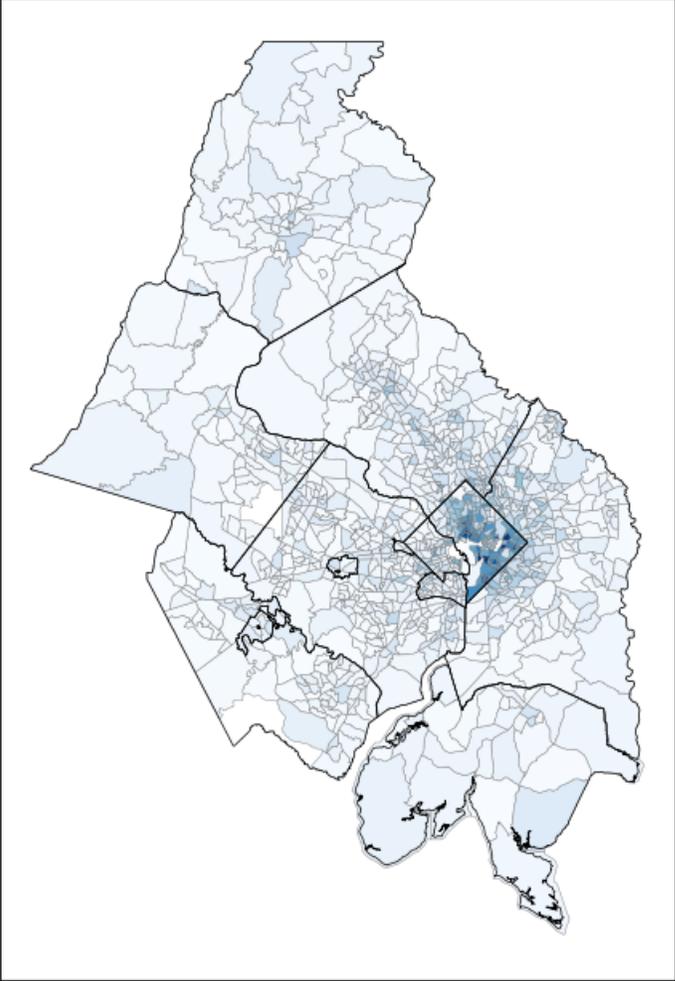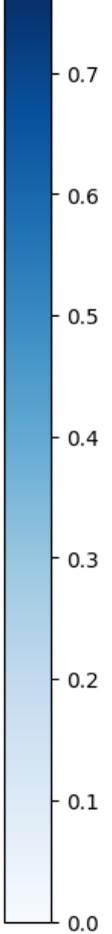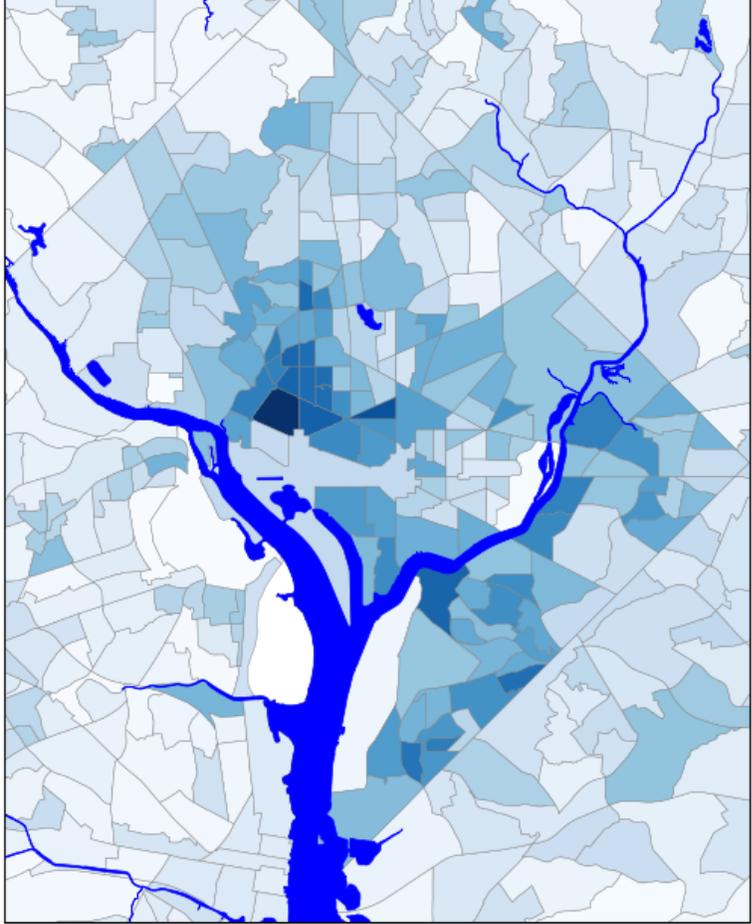
Share of Zero Vehicle Households by Tract

Share of Zero Vehicle Households by Tract

Share of Zero Vehicle Households by Tract (Zoomed to DC)

Programmatic Applications Matrix

Programmatic area	StreetLight	Replica	Teralytics	LOCUS
Traffic volume monitoring	Yes	Yes	Yes	Yes
Mode share	Limited	Yes	No	No
Travel demand modeling support	Limited	Yes	No	No
CMP / congestion monitoring	Limited	Limited	No	No
Transit analysis	No	Yes	No	No
Survey support	No	Yes	No	No
Pedestrian / bike analysis	Yes	Yes	No	No
Data integration flexibility	API	Database access	No	No

Overall Findings and Insights

What the evaluation showed

- Big data performs best at aggregated spatial and temporal levels.

- Reliability weakens at finer geographies, detailed mode categories, and highly specific local use cases.

- Ground-truth validation remains necessary before operational use.

What this means?

- These products can support planning, but not all outputs should be treated equally.

- Some applications are ready for broader use; others remain exploratory or supplemental.

- Product choice should be driven by the specific analytical question.

National Capital Region
Transportation Planning Board

Implications for Decision Makers

- Use product strengths selectively rather than assuming one platform can meet all needs.

- Continue evaluations of big data products from more programmatic areas with observed data sources and survey-based validation.

- Encourage higher usage of big data to explore its potentials.

- Align future procurement decisions with specific use cases such as monitoring, modeling, multimodal analysis, or special studies.

National Capital Region
Transportation Planning Board

Takeaways for MAS Members

- Big data is useful for OD patterns, traffic monitoring, and broad mobility trends analysis.

- Analysis results are more reliable at aggregated levels.

- Caution is required when analyzing fine-scale and mode-specific projects.

- Different products support different use cases.

- Big data works best as a complement to observed data.

National Capital Region
Transportation Planning Board

Acknowledgement

- Dept. of Planning Data and Research

  - Yu Gao, Kenneth Joh, James Li, Ian Newman, and Timothy Canan

- Dept. of Travel Forecasting and Emission Analysis

  - Feng Xie, Ray Ngo, and Anant Choudhary

- Dept. of Multimodal Planning

  - Pierre Gaunaurd and Eric Randall

Zhuo Yang

Transportation Data Analyst

(202) 962-3370

zyang@mwcog.org

Metropolitan Washington Council of Governments

777 North Capitol Street NE, Suite 300

Washington, DC 20002

mwcog.org/tpb

National Capital Region
Transportation Planning Board