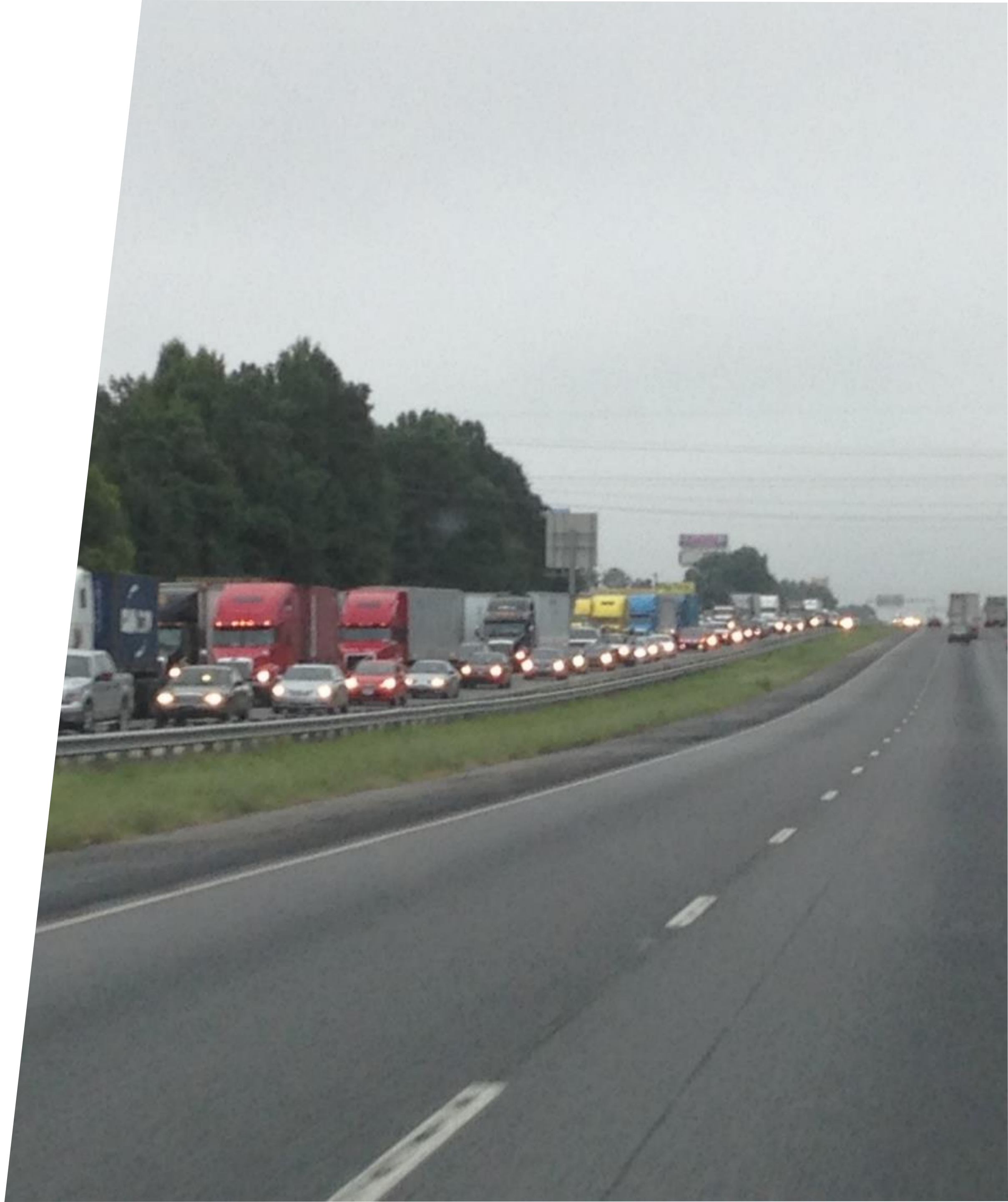
The Promises and Perils of Data for Travel-Activity Behavior Analysis

Presentation to:
WTS FAU Student Chapter

November 19, 2021

RS&H



Presenter

Zahra Pourabdollahi, PhD, PE

- Travel Demand Modeling
- 10+Years Modeling Experience
  - Truck Touring & Behavior-Based Freight Models
  - Florida Behavior-based Tourism Model
  - Statewide Passenger & Freight Models
- Serves on TRB Committees Standing Committees
  - *Freight Planning and Logistics (AT015)*
  - *Traveler Behavior and Values (AEP 30)*

RS&H

Agenda

Travel-Activity Data

Analyzing E-shopping Trends –NHTS Data

Revealing Commercial Truck Trip Patterns – GPS Data

Summary & Discussion

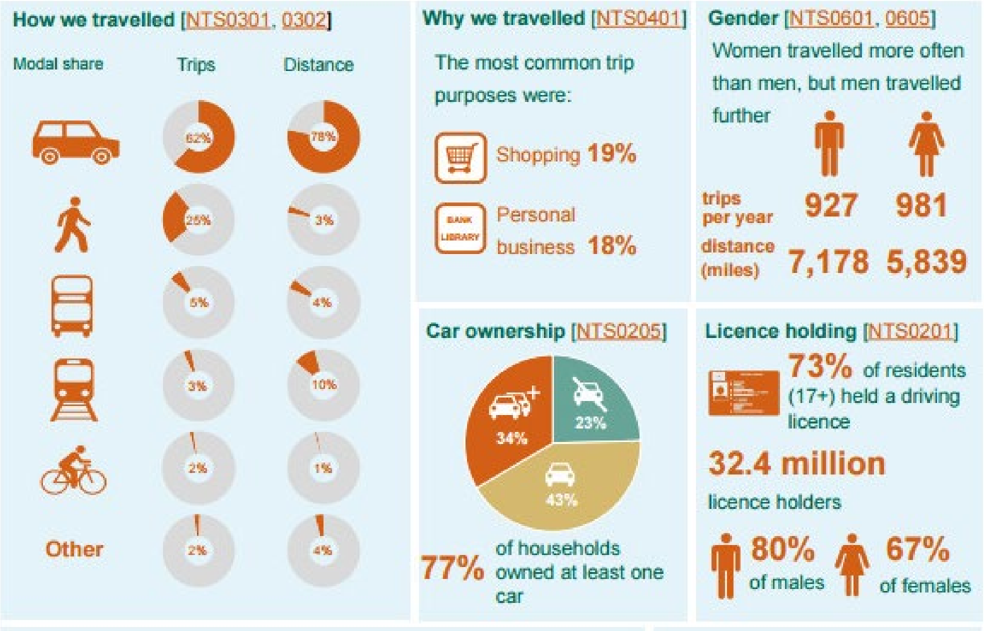
Mobility Analysis

Travel Activity Data

Travel Activity Data

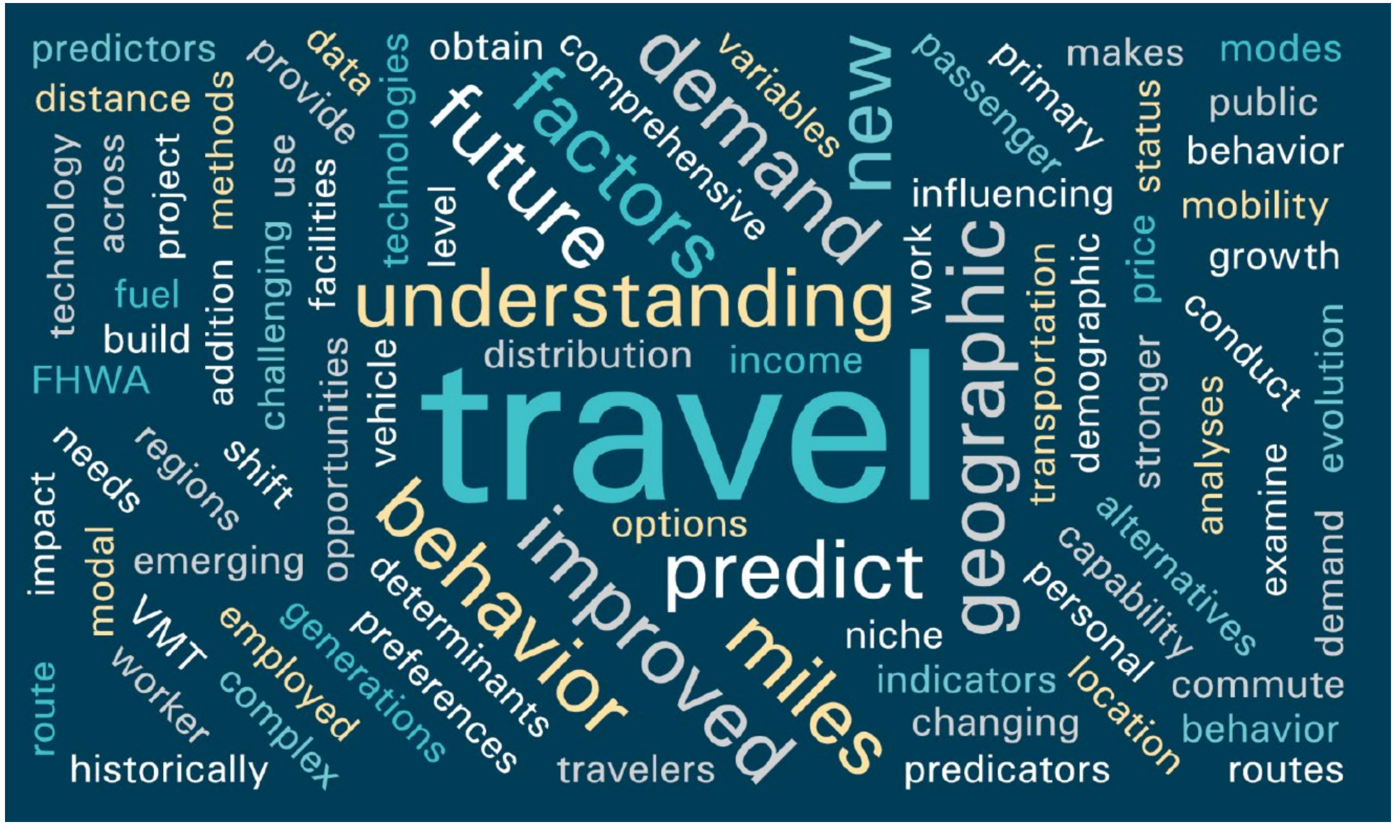
» ***Data Types***
- Travel Survey Data
- Traffic Data (e.g., traffic count data)
- GPS Data
- Bluetooth and Mobile Device Data
- Social Media Data

- Public domain data
- Proprietary data



Travel Activity Data

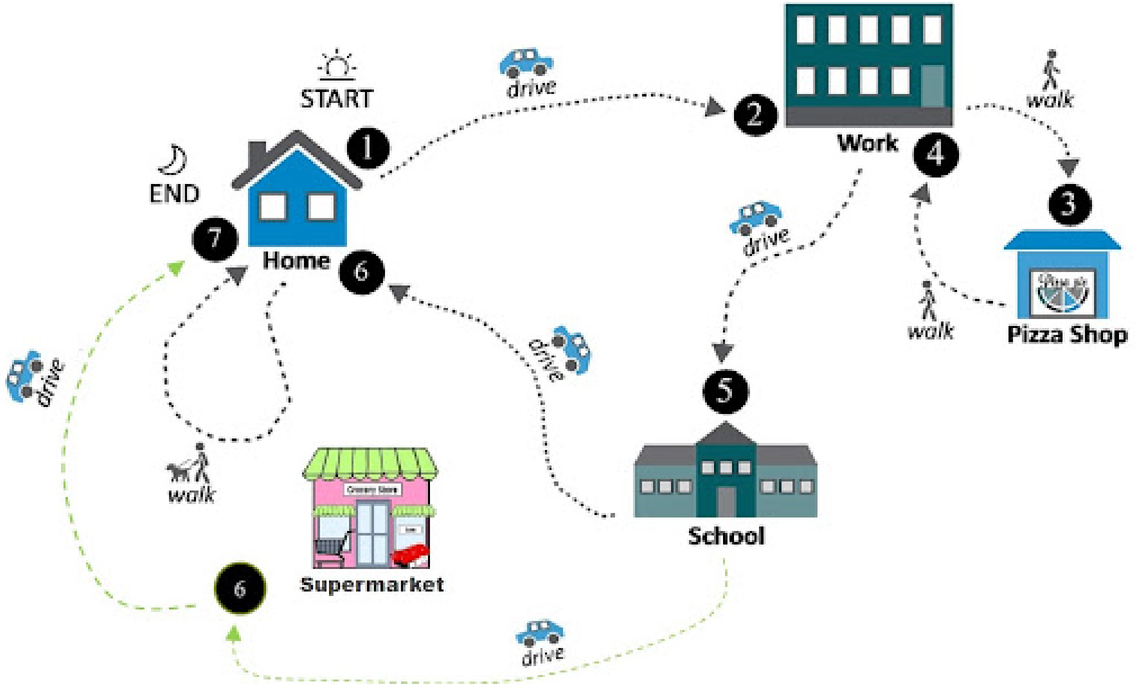
» ***Purpose***
- Identifying Patterns
- Mobility Analysis
- Developing Travel Demand Models
- Understanding The Impacts of Travel Patterns on Geographical Distribution and Land Use
- Inform Decision Makers for Evaluating Plans/Polices



Travel Activity Data

- Activity-Based / Agent-Based Travel Demand Modeling
- Human/Freight Mobility Analysis

understanding and forecasting how individuals/goods perform activities and move in time and space

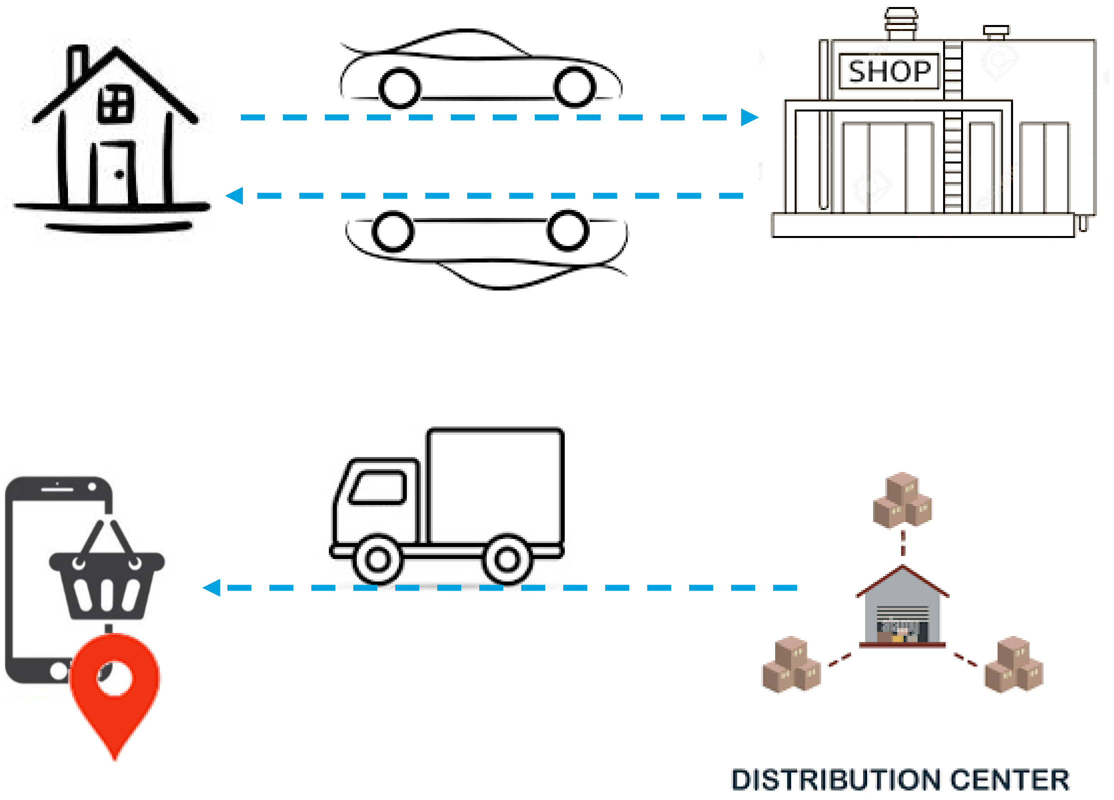


RS&H

Analyzing E-shopping Trends
NHTS Data

Analyzing E-shopping Trends - NHTS Data

» **E-shopping**



Comparing growth: US **ecommerce** vs. **total retail*** sales

Year-over-year growth, 2010-2020

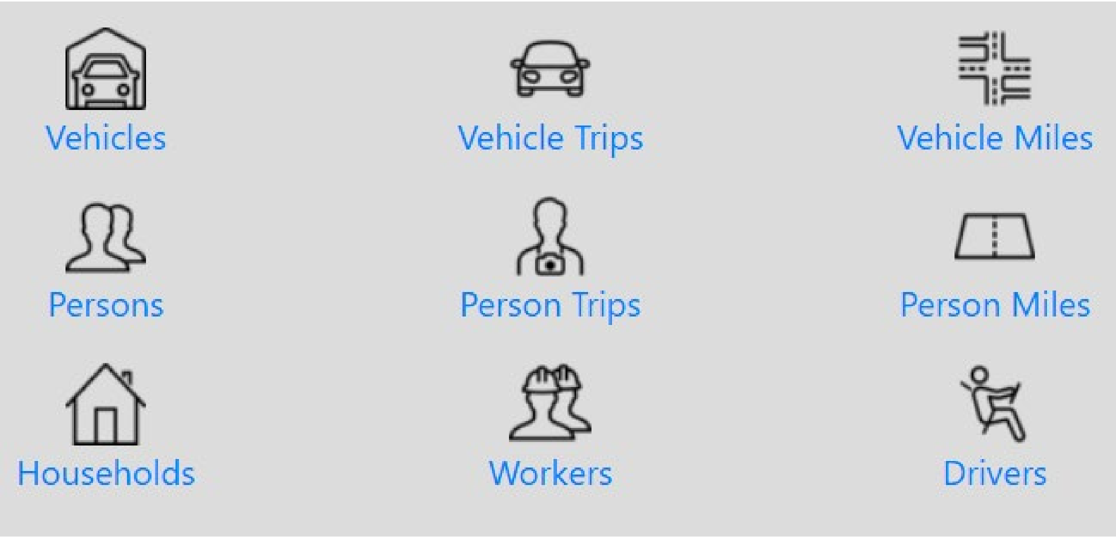
What is the net effect on regional travel trips?

Analyzing E-shopping Trends - NHTS Data

- **Examine Frequency of E-shopping in The U.S. and Its Influential Factors**

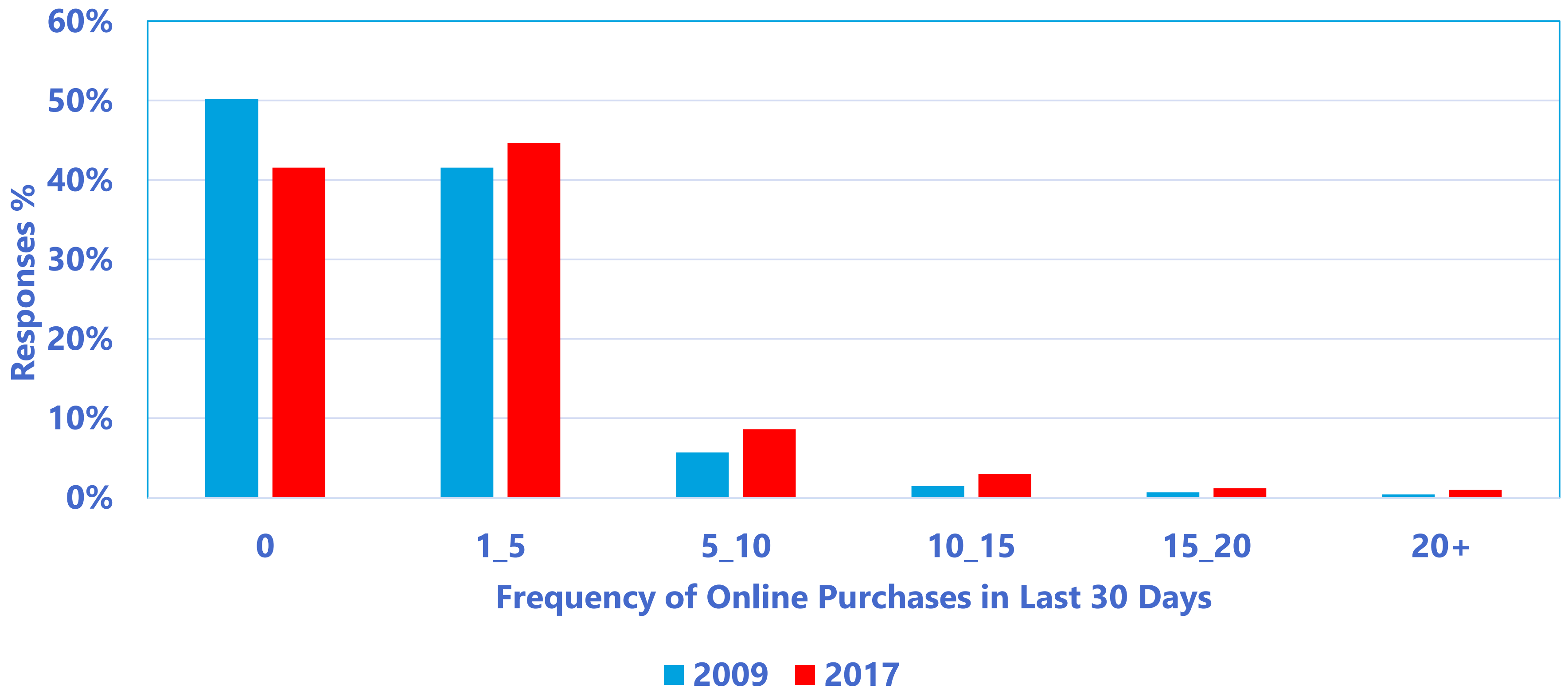
  (Who are the e-shopper and how frequently they shop online)

- 2017 National Household Travel Survey
  - Conducted by the Federal Highway Administration (FHWA),
  - Is the authoritative source on the travel behavior of the American public
  - The only source of national data to analyze trends in personal and household travel
  - It includes daily non-commercial travel by all modes, including characteristics of the people traveling, their household, and their vehicles.



Analyzing E-shopping Trends - NHTS Data

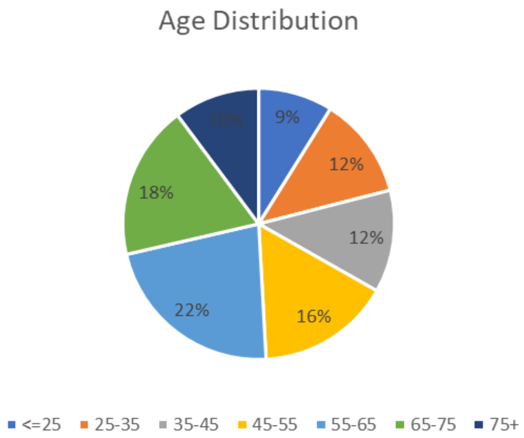
» Examine Frequency of E-shopping in The U.S. and Its Influential Factors
  - 2009 NHTS: “number of times respondent has purchased online for delivery in last 30 days”
  - 2017 NHTS: frequency of internet use and smart devices use to access internet

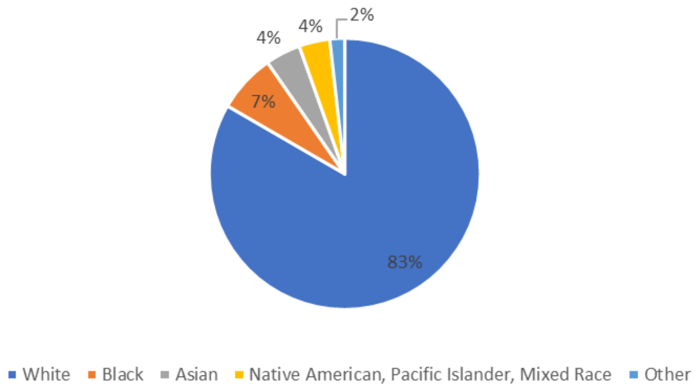


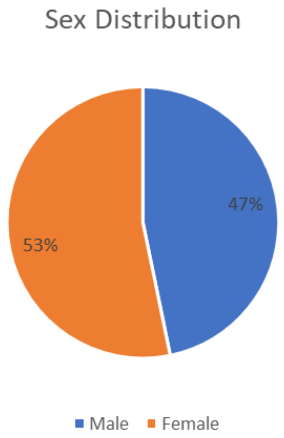
RS&H

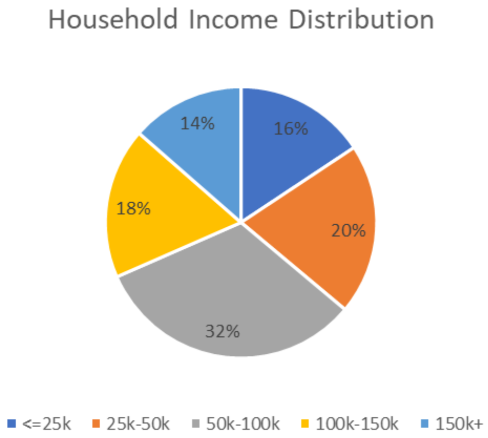
Analyzing E-shopping Trends - NHTS Data

» 2017 NHTS Data
  – a sample of 235,805 useable records



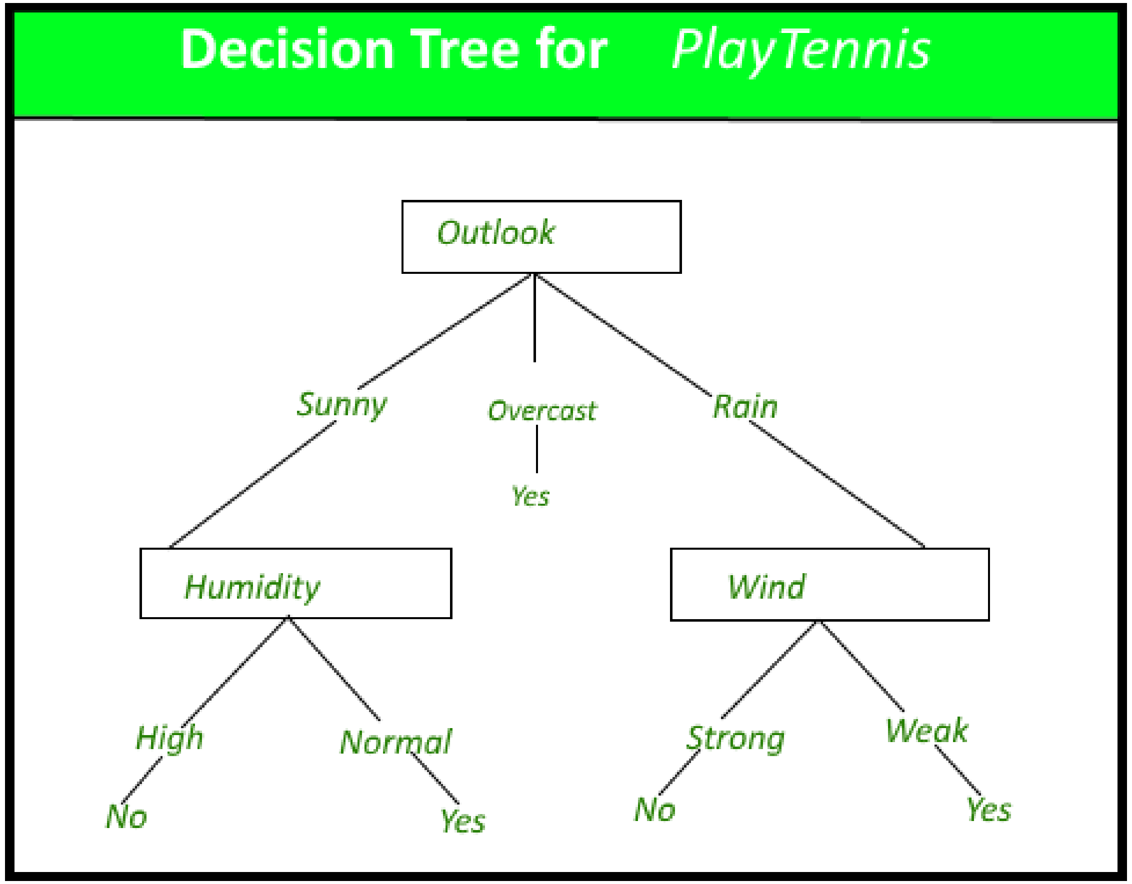






Analyzing E-shopping Trends - NHTS Data

» Analysis Method
- Classification Decision Tree Approach
- Train data (70% of records) and test data (30% of records)
- R
- Classification and Regression Tree (CART) algorithm
- Splitting method: Entropy (Information Gain)
- Growth control: Complexity Parameter (CP)

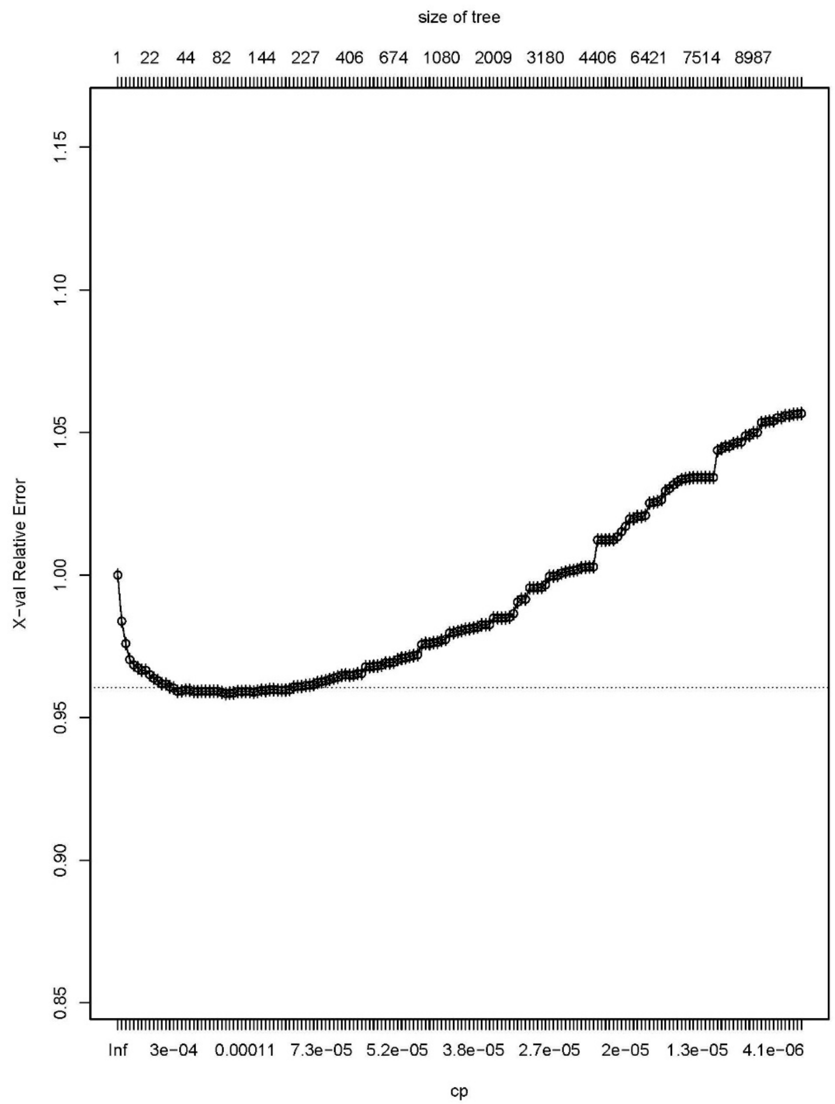


RS&H

» The Process

1. Classify dataset into train and test subsets
2. Develop unrestricted tree to fit the train data
3. Observe CP value changes across tree size and find optimal CP value
4. Measure predictive accuracy of full tree on test dataset (calculate cross-validation error)
5. Prune the tree using optimal CP value and cross-validation error
6. Measure predictive accuracy of pruned tree on test dataset
7. Compare accuracies and finalize the model if criteria are met

Analyzing E-shopping Trends - NHTS Data



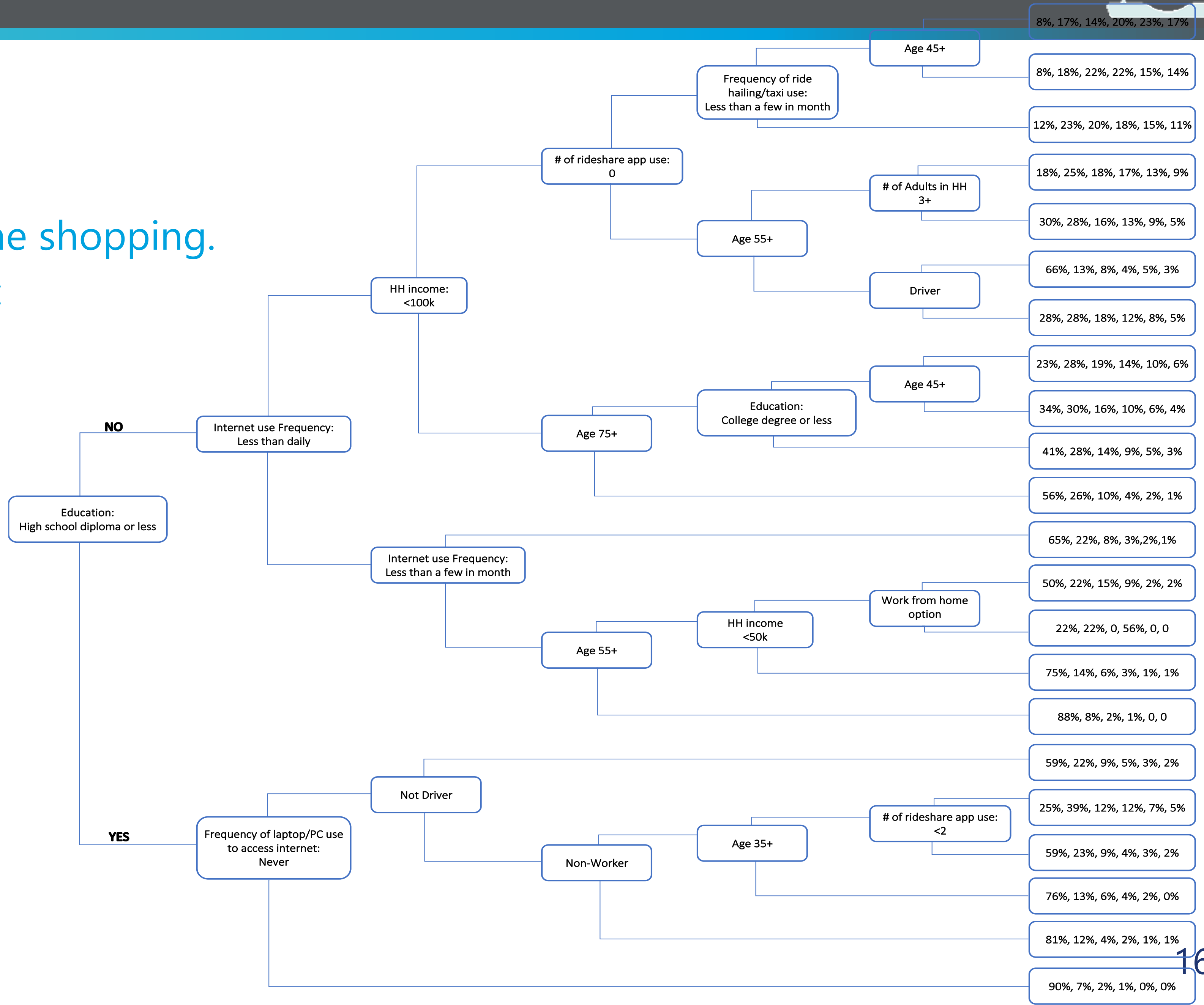
	Nodes	Split	Depth	CP	X validation error	Accuracy
Base Tree	19151	9575	30	0	1.0566	0.5984
Pruned Tree	63	31	13	0.000376	0.9619	0.6493
Final Tree	43	21	6	0.000424	0.9651	0.6421

RS&H

Analyzing E-shopping Trends - NHTS Data

» Results

- the dependent variable: frequency of monthly online shopping.
- classified into 6 categories: 0, 1-2, 3-4, 5-6, 7-10, 10+



Analyzing E-shopping Trends - NHTS Data

- Results
  - The significant factors in determining frequency of monthly online shopping include:
    - *socio-demographic characteristics*
    - *socio-economic characteristics*
    - *technology adoption/internet usage*
    - *Travel related attribute (ridesharing/driving)*

Analyzing E-shopping Trends - NHTS Data

- Results
  - the least likely e-shoppers:
  - With high school diploma or less, never use laptop or PC for accessing internet.
  - 55 years and older, college/university/higher education, low internet usage.
  - With high school diploma or less, have access to internet, non-drive, non-worker.
  - 35 years and older, workers, with high school diploma or less, with access to internet, non-driver.

Analyzing E-shopping Trends - NHTS Data

- Results
  - the most frequent online shoppers:
    - *Higher education (college/university or higher).*
    - *Daily access to internet*
    - *High income households (100K+)*
  - The propensity of online shopping in frequent e-shoppers increases for individuals:
    - *with ride hailing applications on their device and use ridesharing services more than a few times per month.*
    - *with flexible work schedule and WFH option*

Analyzing E-shopping Trends - NHTS Data

» Results

- The influential factors in determining e-shopping frequency:
  - *age, household size,*
  - *employment, education, income, and work from home option,*
  - *technology adoption, frequency of internet usage,*
  - *Travel related attributes: ridesharing application usage, and frequency of using ride hailing or taxi services for travel, and driving status*
- insignificant covariates:
  - *Sex,*
  - *Build environment attributes,*
  - *Health related attribute,*
  - *Other transportation related characteristics*

Activity-based modeling (in-home & shopping activity)
Last mile delivery / urban freight planning

RS&H

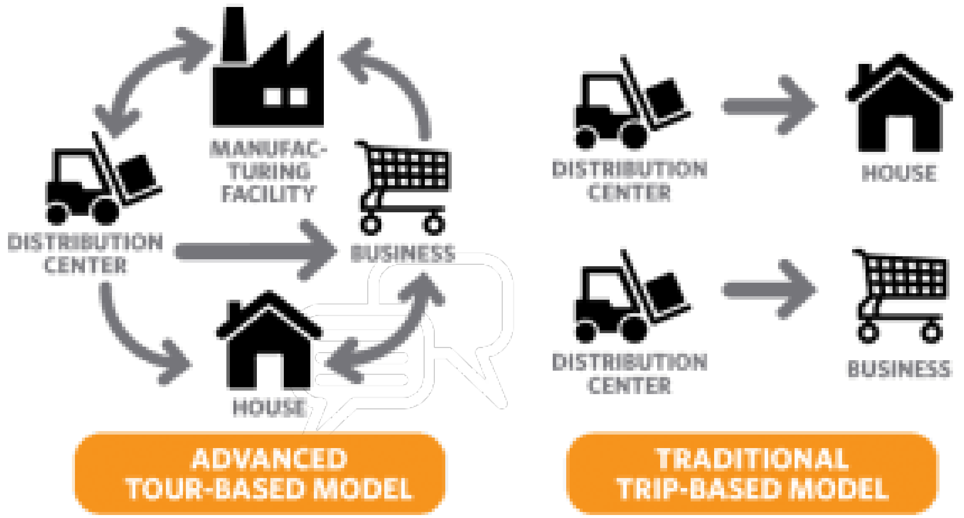
Revealing Commercial Truck Trip Pattern GPS Data

Revealing Commercial Truck Trip Patterns - GPS Data

Traditional Trip-based Truck Model

Truck Trip Generation → Trip Distribution → Trip Assignment

- ❑ Disregard temporal and spatial interrelations between truck trips



Tour-based Model

Tour Models →
- Generation model
- Tour Purpose
- Start Time of Day Choice
- Vehicle Type Model
- Commodity Type

Stop Models →
- Stop Frequency
- Stop Activity
- Stop Location
- Stop Duration

Trip Accumulation →
- End of tour model
- Trip accumulation

Trip Assignment

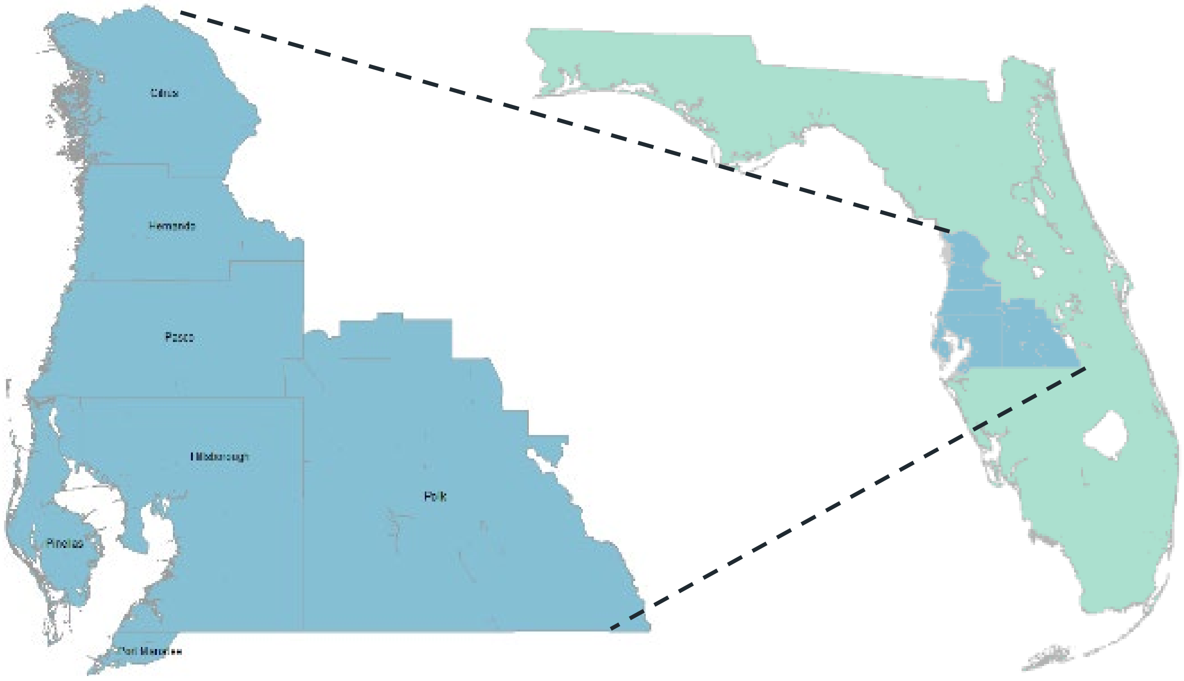
Revealing Commercial Truck Trip Patterns - GPS Data

» GPS Data
- Provide spatial and temporal information on truck movements
- Require extensive pre/post processing
- Proprietary
- Limited info on cargo and other travel information

» Today's Discussion
- **Utilizing GPS data to reveal regional truck trip patterns for modeling purpose**

Revealing Commercial Truck Trip Patterns - GPS Data

FDOT – District Seven (Tampa Bay Region)

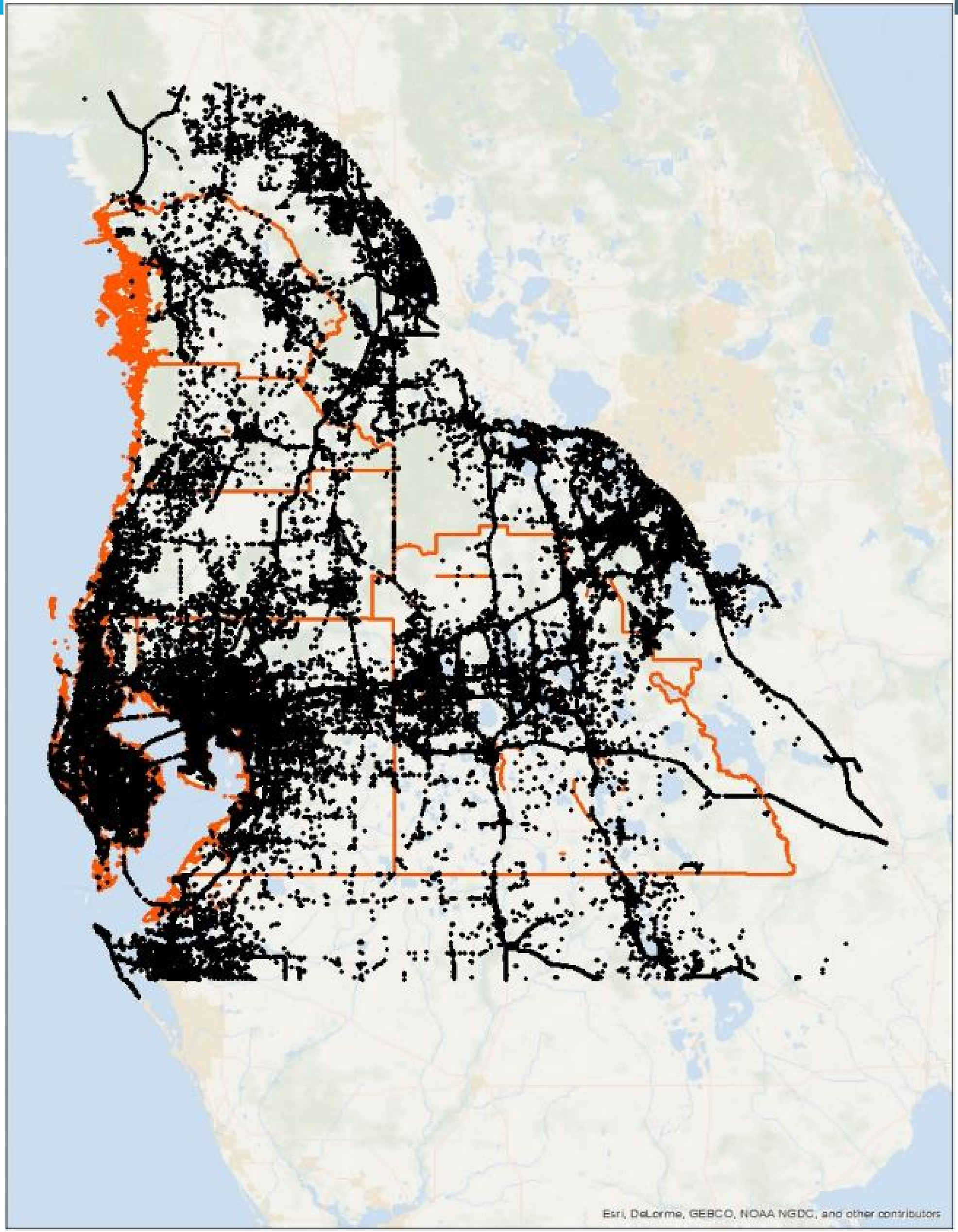


ATRI GPS Data For Trucks

Revealing Commercial Truck Trip Patterns - GPS Data

ATRI Data	Description
Spatial Coverage	Seven Counties + 10 mile external buffer
Temporal Coverage	8 Weeks (between October and July)
# of Records	96.4 Million
# of Unique Truck IDs	110K

ATRI Data Attribute	Description
Truck ID	Vehicle Identifier (Dynamic IDs)
X	Degrees longitude
Y	Degrees latitude
Time/Date Stamp	Time and date
Spot Speed	Travel speed (mph)
Heading	Travel direction



RS&H

Revealing Commercial Truck Trip Patterns - GPS Data

ATRI Data Processing

Data QC & Cleaning

Movement & Stop Identification

Stop Filtering

Trace Movements

Geo-spatial analysis:

Land-use / Network / Employment info for purpose inference

Tours and Trip Table by Purpose

GPS Pings → GPS Traces → Route, Destinations, Activities

RS&H

Revealing Commercial Truck Trip Patterns – GPS Data

» Data Processing

- Sort pings by Truck ID and Date/Time stamp
- Calculate distance, duration, and speed between consecutive pings
- Identify the status of each record based on the average speed (First, Moving, Stopped, Last)
- Combine Cluster Stops
- Calculate stop duration
- Remove stops less than 5 min
- Remove moving records
- Assign stop and tour number
- Identify TAZ for each stop record using ArcGIS geoprocessing
- Assign land use and employment information to the stops

96.4 Million GPS records

1,710,493 Stops (first-stop-last)

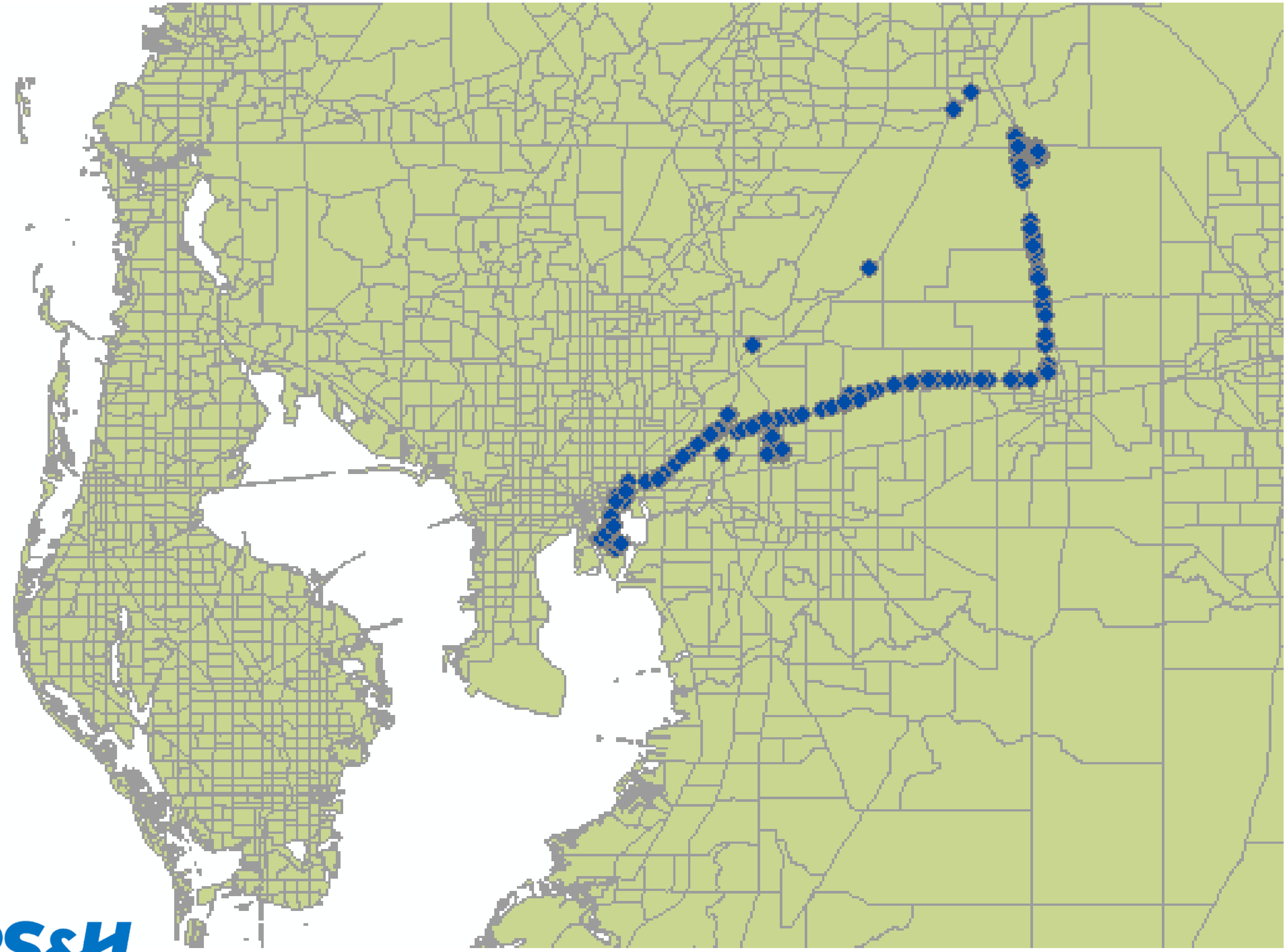
94,928
Unique Truck IDs

325,615
Heavy Truck Tours

1,384,878
Heavy Truck Trips

RS&H

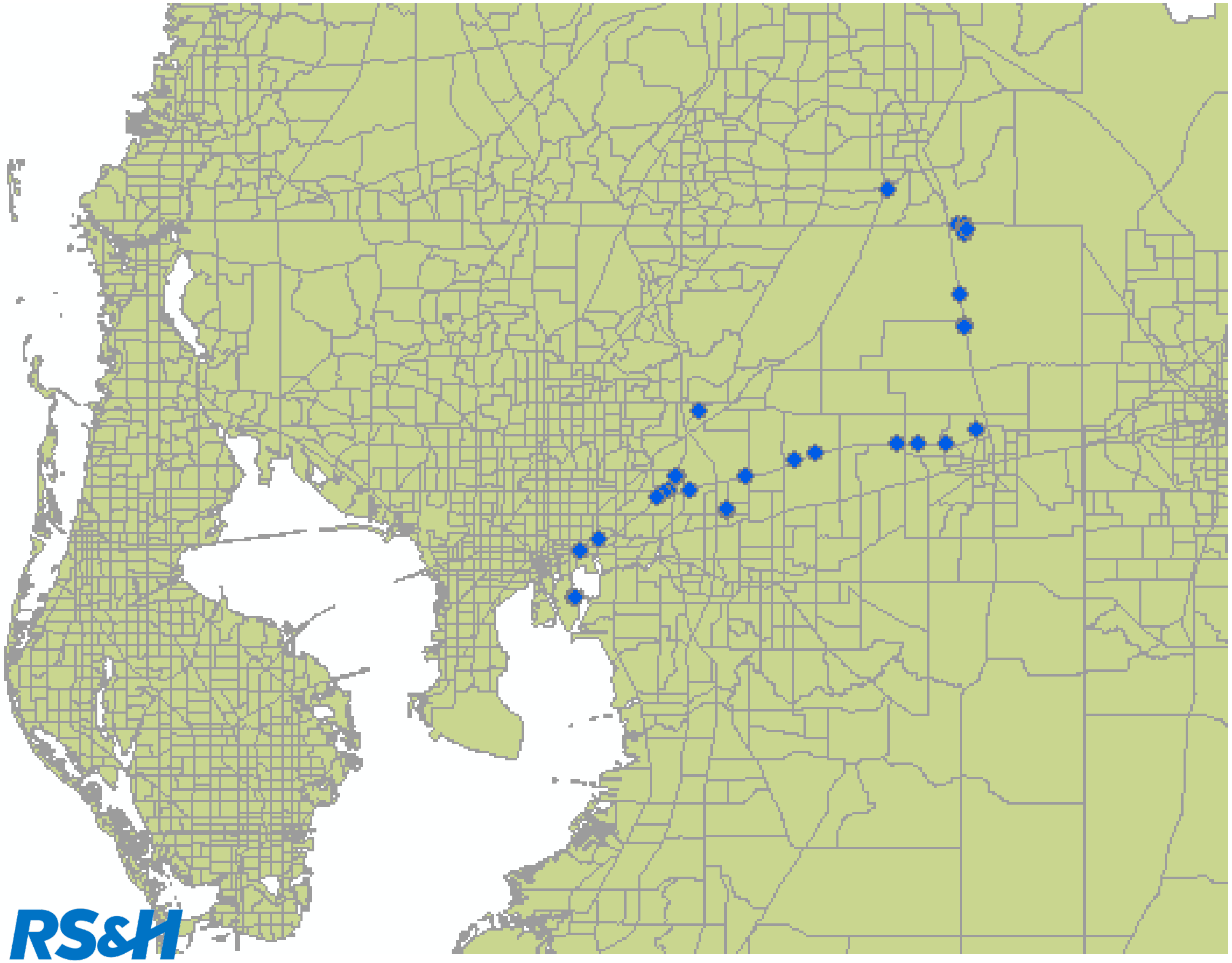
Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 117722	October
GPS events (pings)	596
Identified Tours	6

RS&H

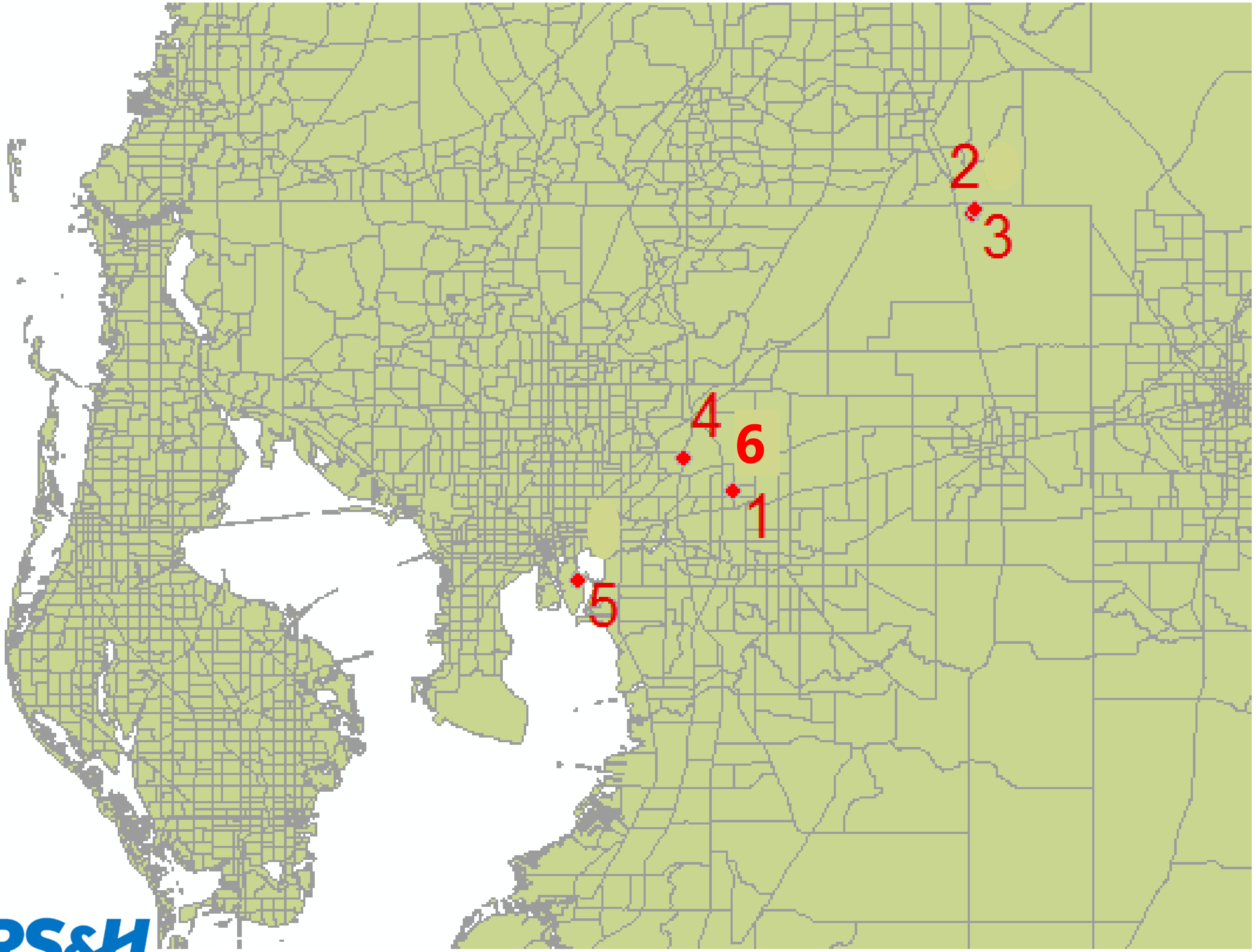
Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 117722 – One example Tour	October
GPS events (pings)	73
Tour #	1
Identified Stops	4

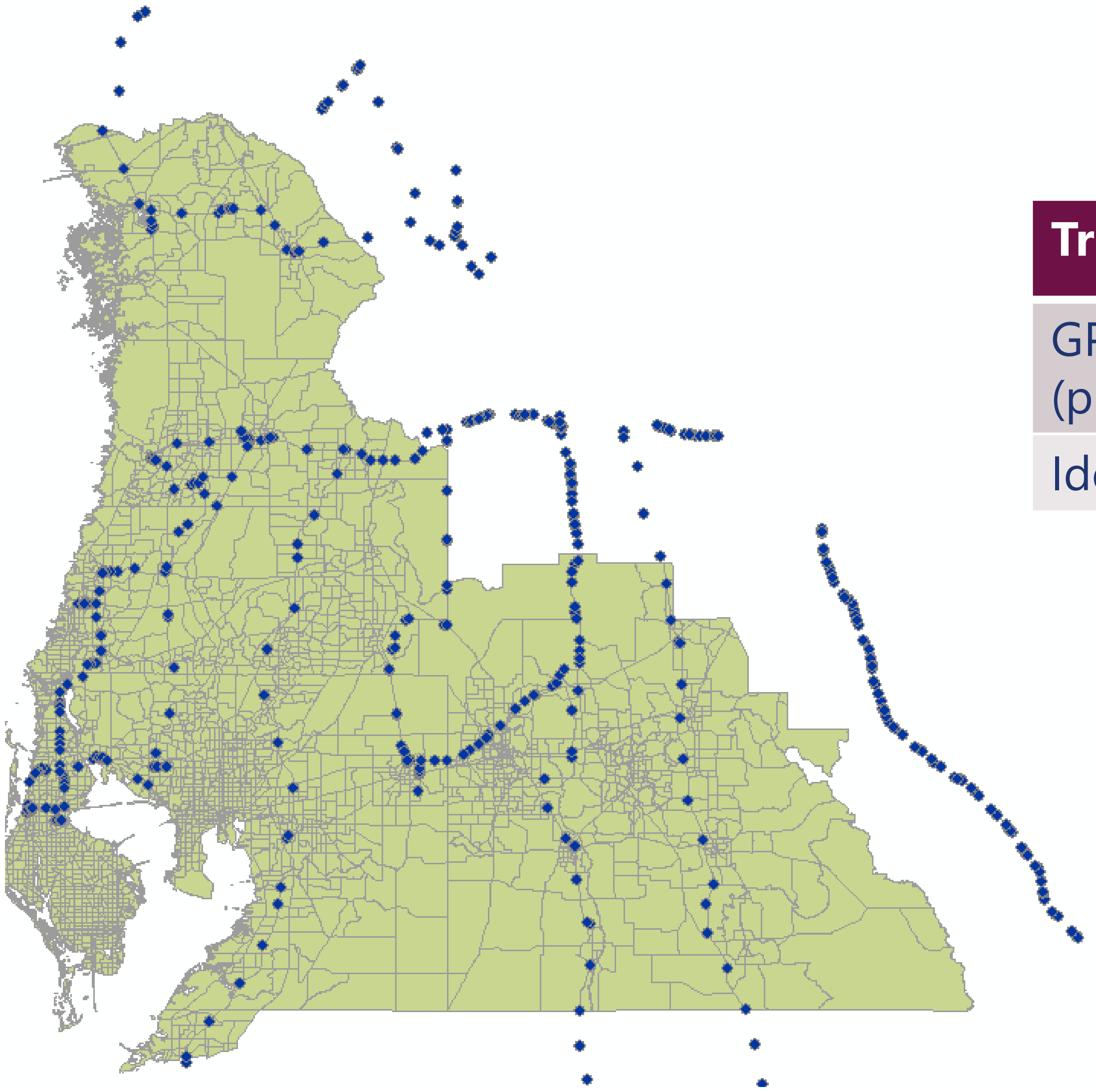
RS&H

Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 117722– One example Tour	**October**
GPS events (pings)	73
Tour #	1
Identified Stops	4
Trips	5

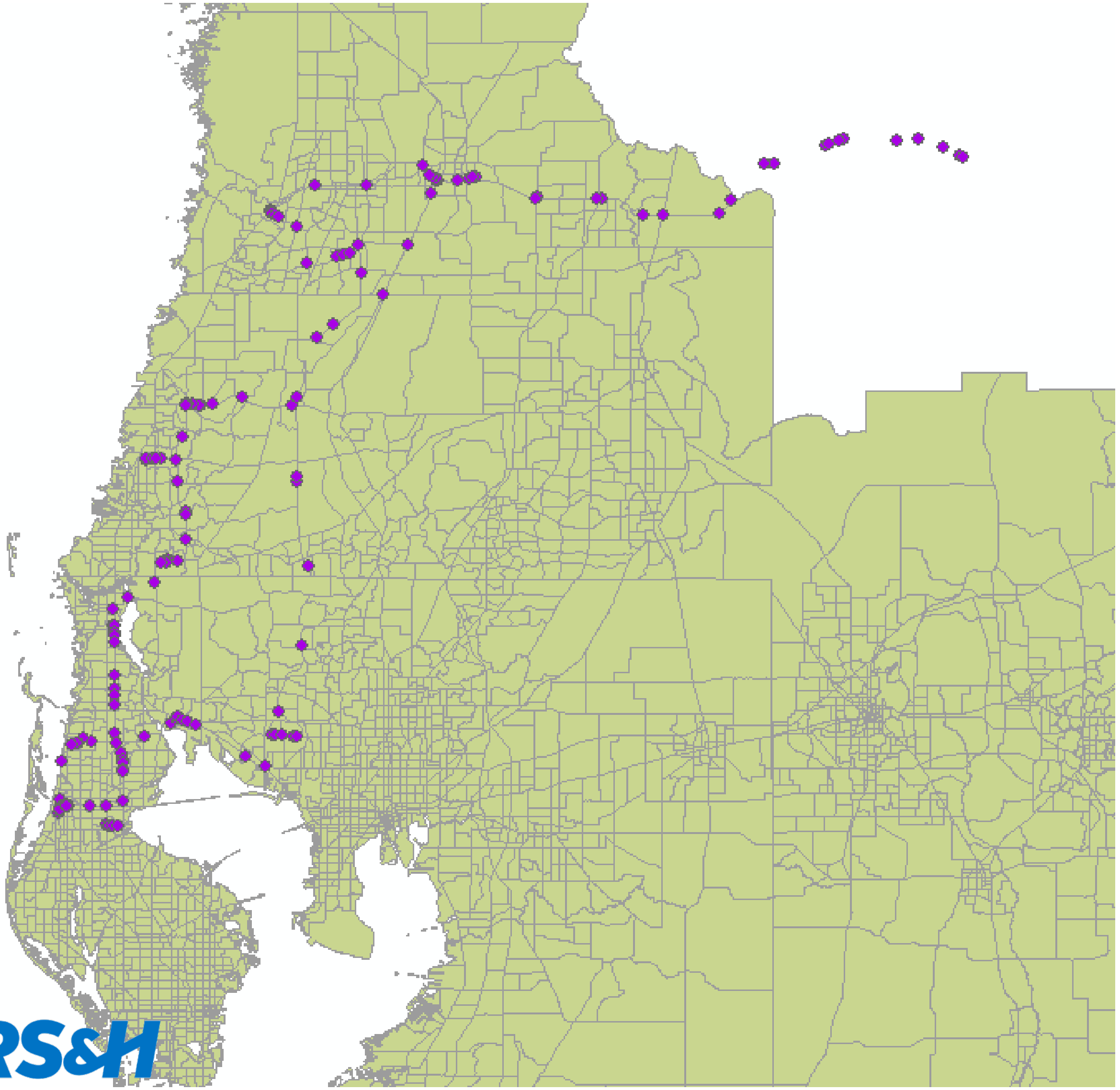
Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 104035	October
GPS events (pings)	553
Identified Tours	5

RS&H

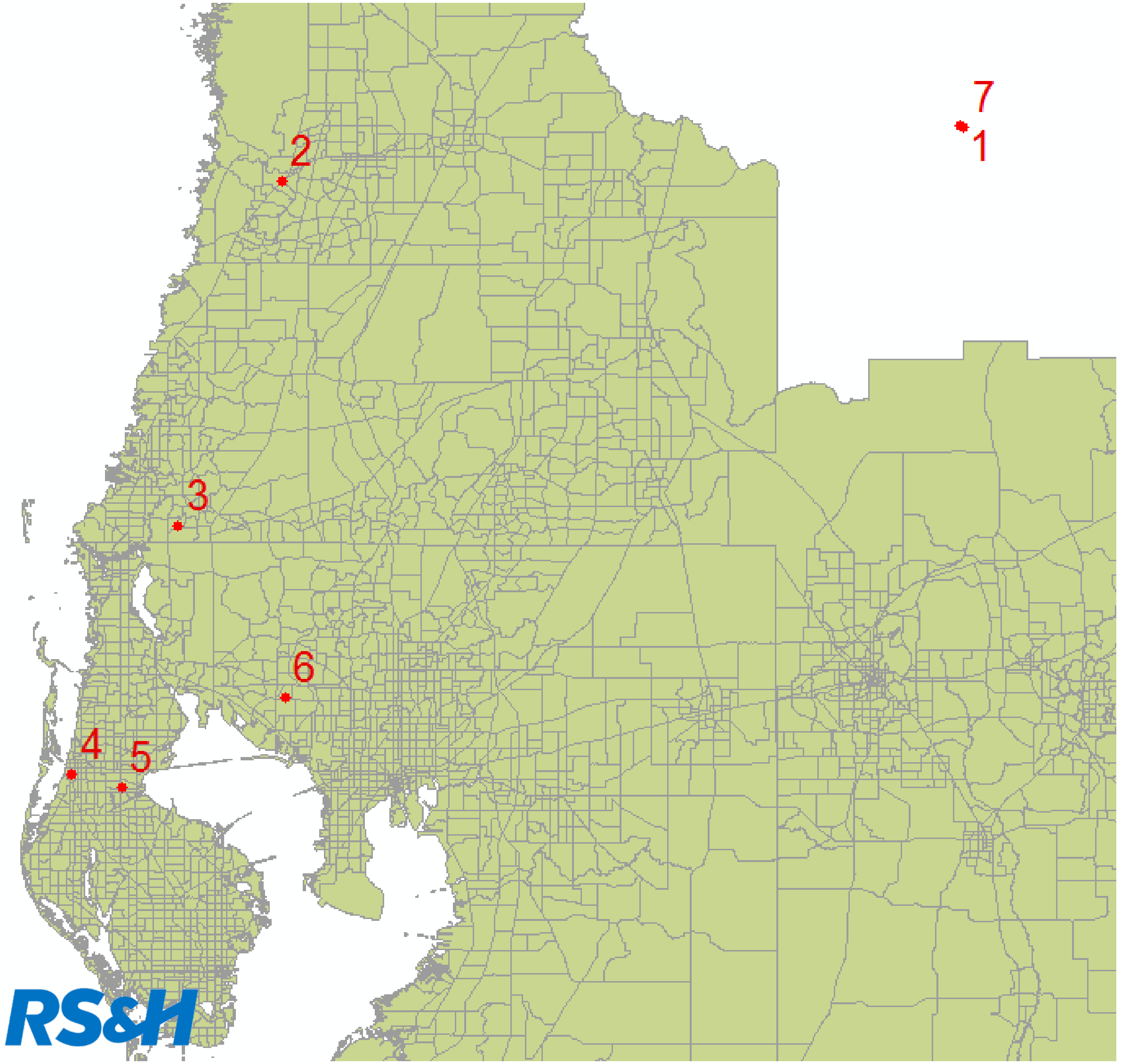
Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 104035 – One example Tour	October 2015
GPS events (pings)	194
Tours #	1
Identified Stops	5

RS&H

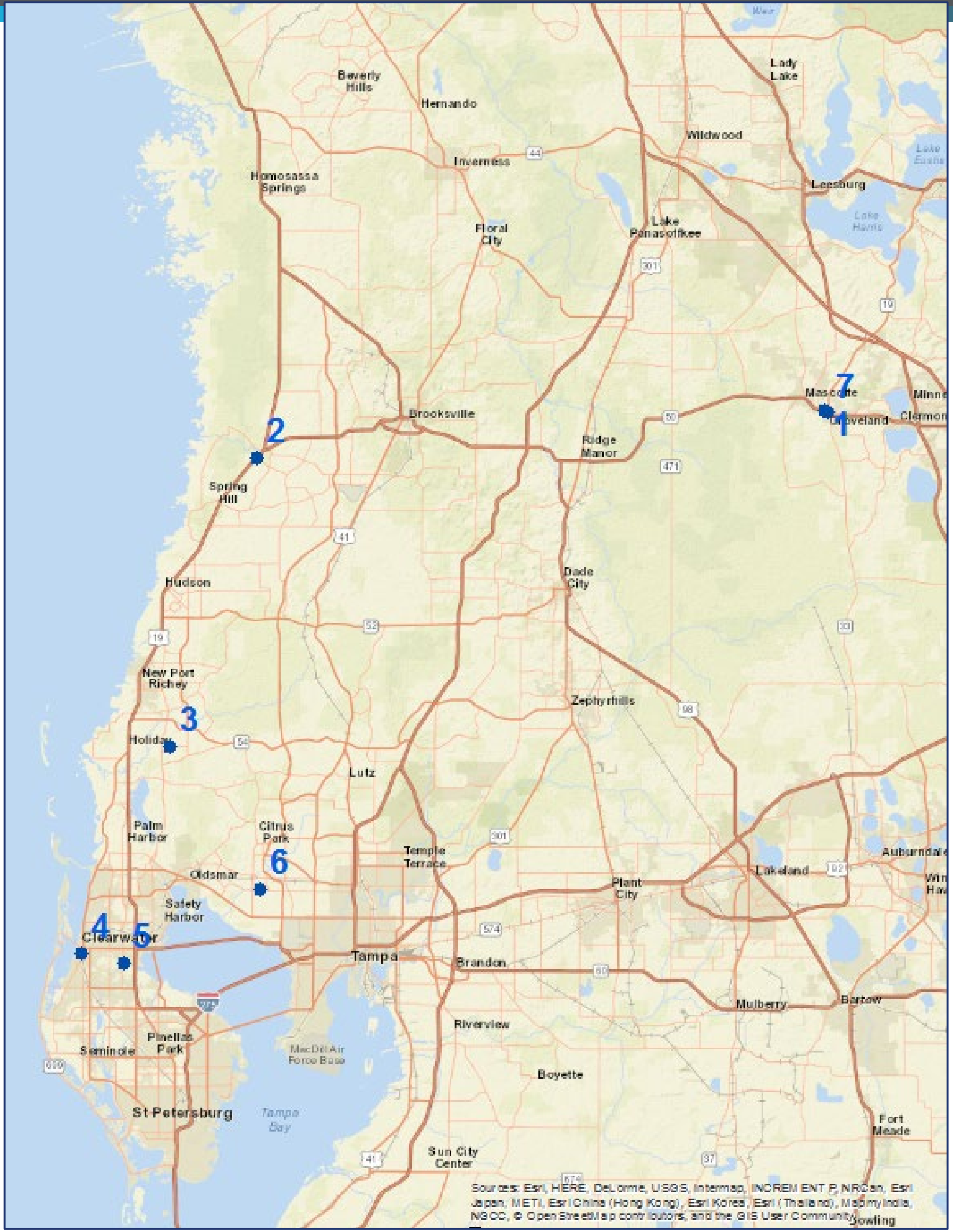
Revealing Commercial Truck Trip Patterns – GPS Data



Truck ID 104035 – One example Tour	October 2015
GPS events (pings)	194
Tours #	1
Identified Stops	5
Trips	6

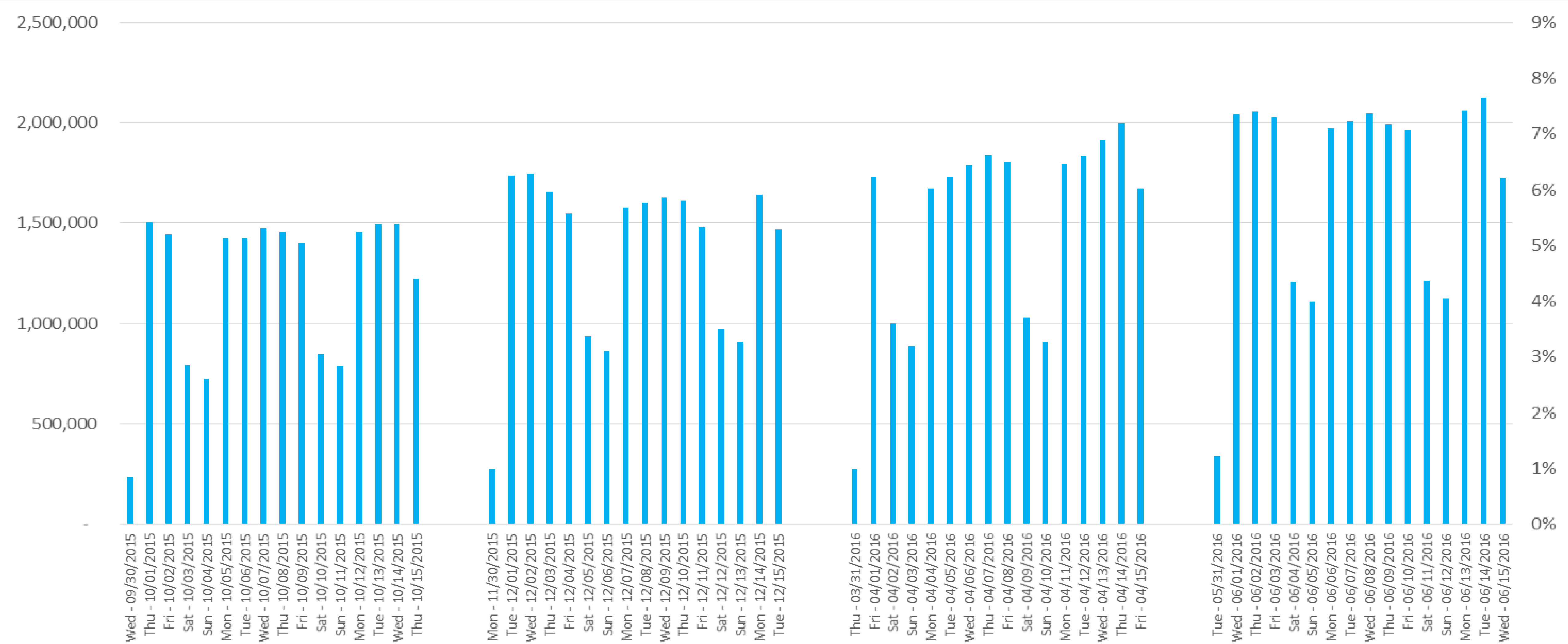
RS&H

Revealing Commercial Truck Trip Patterns - GPS Data



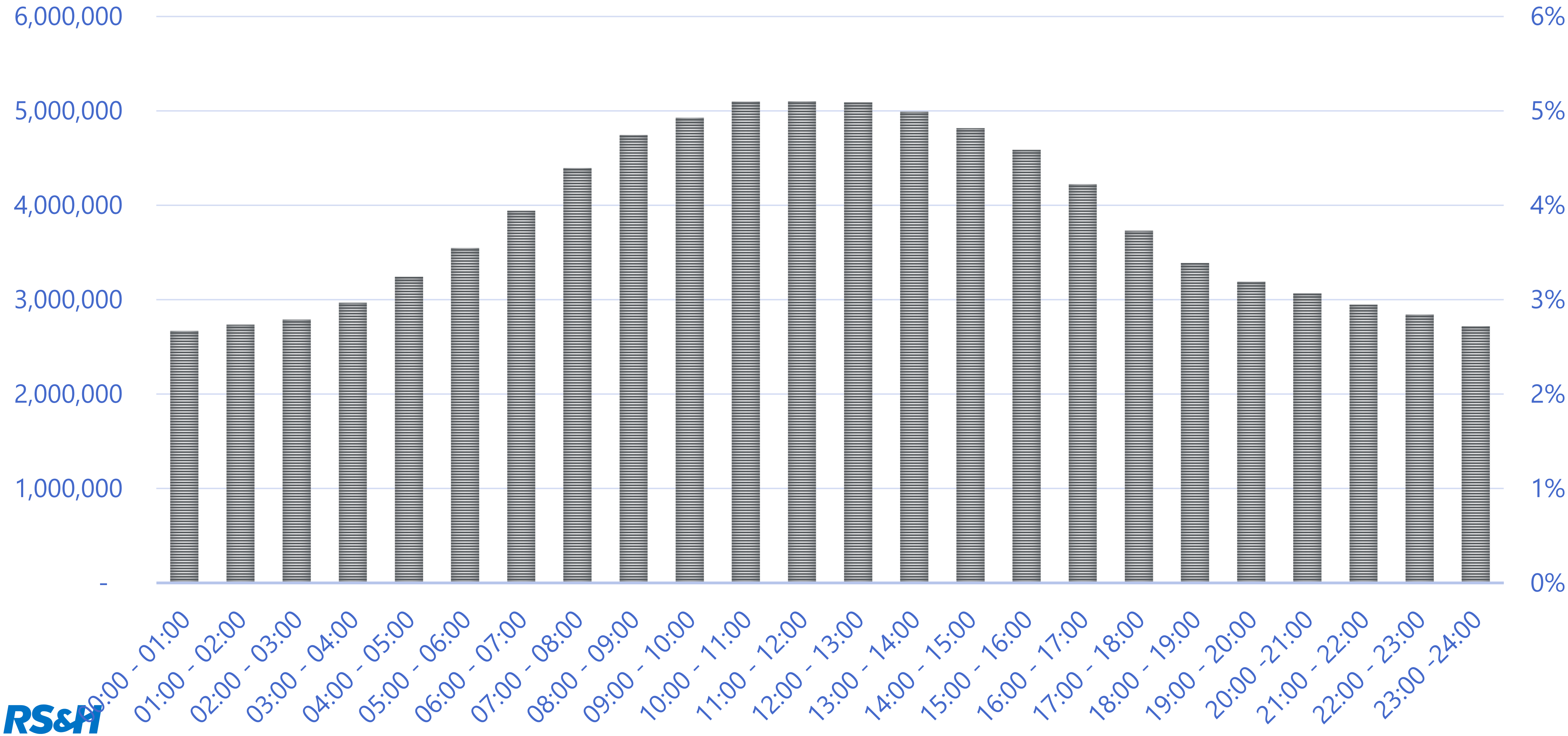
Revealing Commercial Truck Trip Patterns – GPS Data

Temporal Variation of Truck Activities



Revealing Commercial Truck Trip Patterns – GPS Data

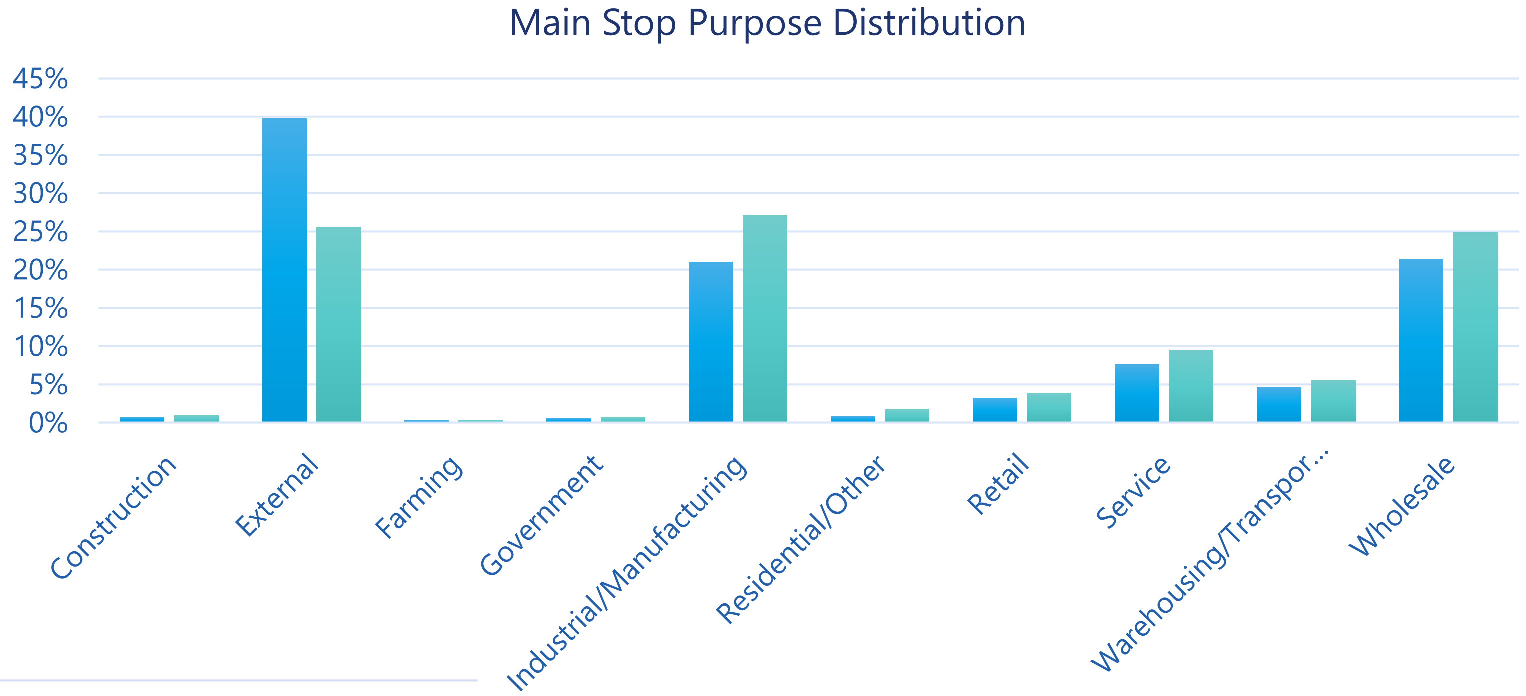
Temporal Variation of Truck Activities

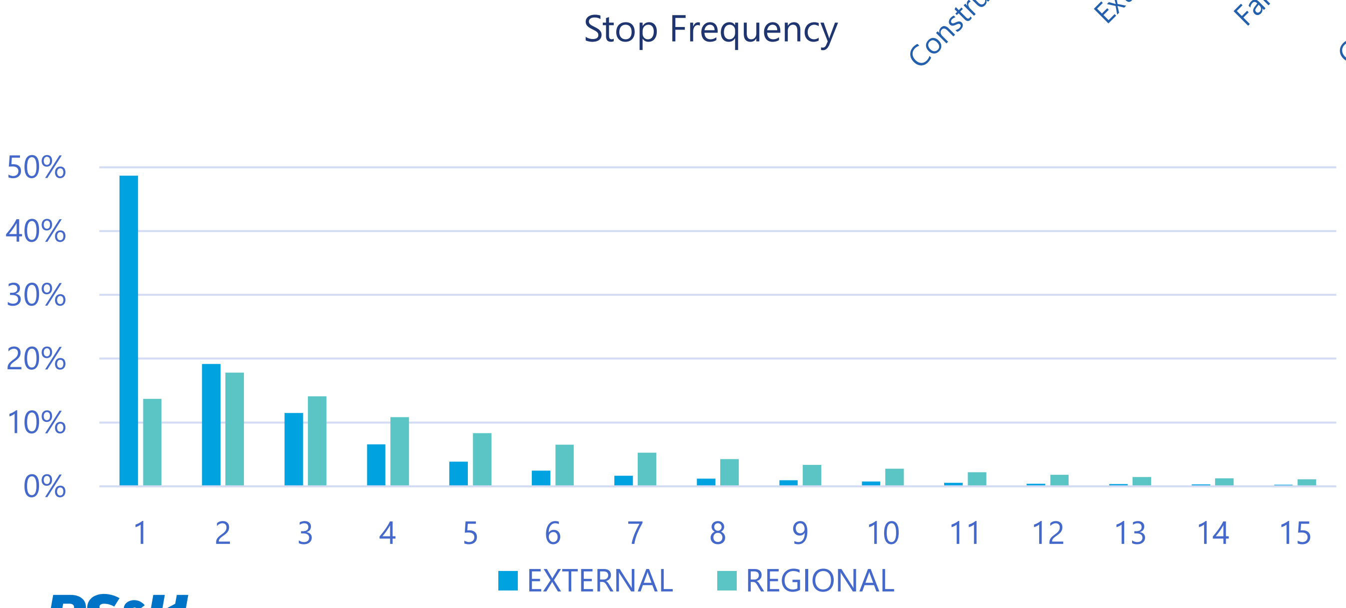


RS&H

Revealing Commercial Truck Trip Patterns - GPS Data

Stop Frequency and Purpose



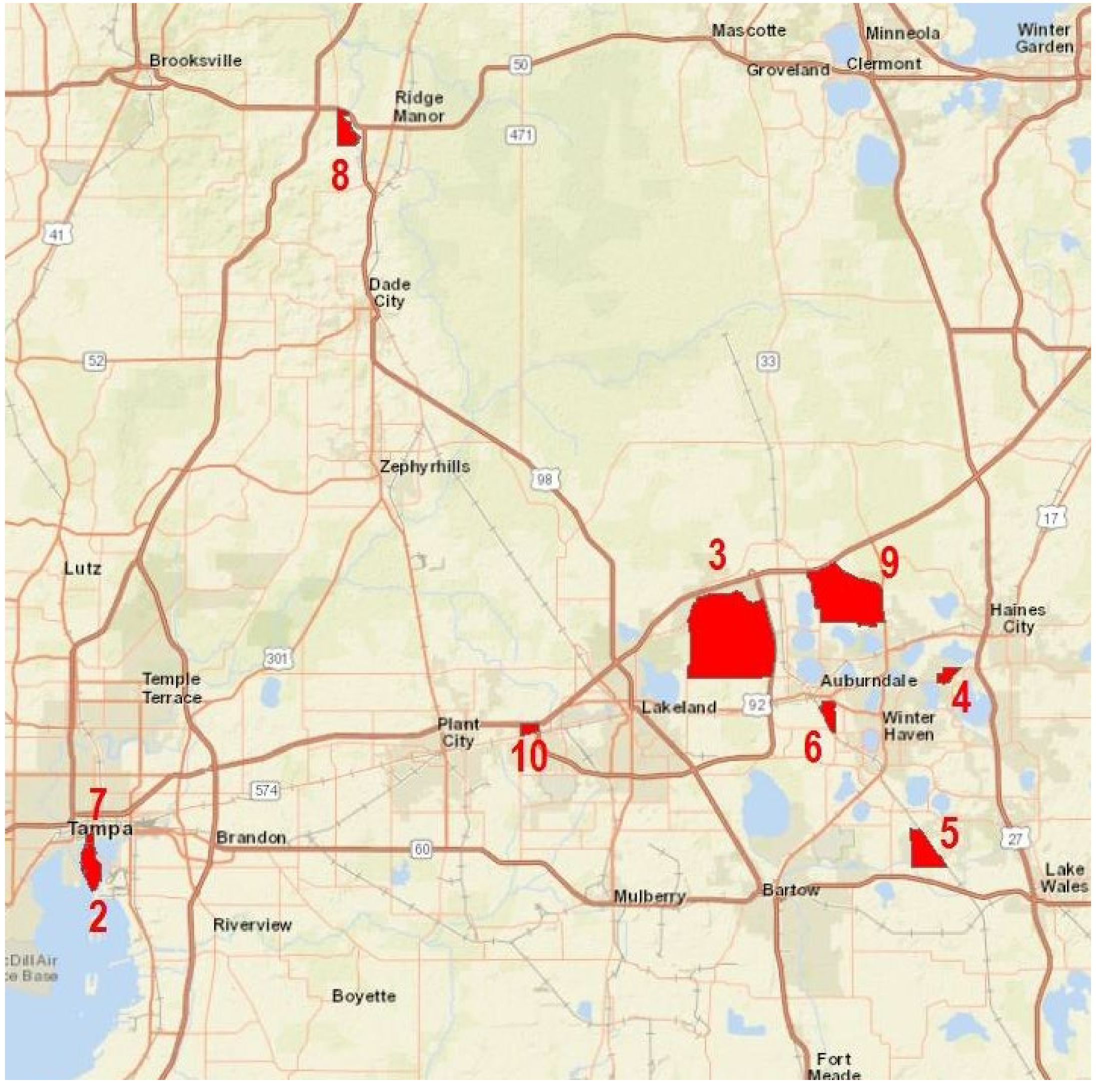


RS&H

Revealing Commercial Truck Trip Patterns - GPS Data

Most Visited Destinations

Rank	TAZ	Description
1, 2	7-367	Port Tampa
3	1-265	Sam's Club Distribution Center and other
4	1-672	Walmart Distribution Center
5	1-467	CSX Winter Haven
6	1-665	Coca Cola
7	7-366	Port Tampa - APWU
8	7-2626	Walmart Distribution Center
9	1-379	Truck Rest Areas; Mobile Modular
10	1-3	Distribution Centers; Warehousing and Transportation Center



RS&H

Summary & Discussion

Summary & Discussion

- Travel activity data enable the understanding of travel patterns
- play a critical role in travel trend monitoring, transportation planning, and policy decision support
- Conventional travel behavior data such as NHTS
  - primary source of travel behavior information
  - high cost,
  - Less frequent,
  - cross sectional data,
  - involve more error,
  - provide detailed info and are self validating

Summary & Discussion

- The passively collected data is one of the most effective data sources that provides invaluable information
  - significant potential as supplementary data input
  - reasonable cost
  - Longitudinal (trend analysis)
  - Extensive processing, expansion, fusion, analysis, imputation and inferring methods and integration with other existing data
  - Validation issues
- Social Media and big data such as Twitter data are among emerging data sources that will enable extracting travel-activity trends

Summary & Discussion

» The fusion and integration of different data types seems inevitable

» Big data

» Advanced data analytical methods are required to overcome significant challenges in developing comprehensive travel-activity data that allows stakeholders to track travel behavior trends

The last decade has witnessed very active development and some overlap in two broad, but separate fields: transportation research and data science/computer science

RS&H

Zahra Pourabdollahi
Zahra.pourabdollahi@rsandh.com
6303 Blue Lagoon Dr, Suite 325
Miami, FL 33126
305-428-3217