The Big and Small of Big Location Data for Transportation Analysis

Vijayaraghavan (Vijay) Sivaraman, PhD, PE
Texas A&M Transportation Institute

JUNE 7th, 2024, COMPASS WORKSHOP SERIES

Overview

01

Data Sources:
Telecom, Smartphone Applications and Vehicles

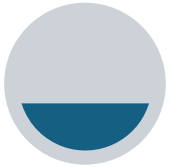
02

Applications:
Visits, Flows, Operational Analysis, Visualization and Dashboards

03

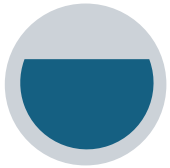
Management:
Data Acquisition, Tools, Skills, Governance
Storage, Security, Access

Workshop Sessions



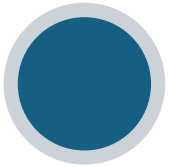
Sources

- Telecom
- Smartphone
- Vehicles



Products

- Data Schema
- Resources
- Management

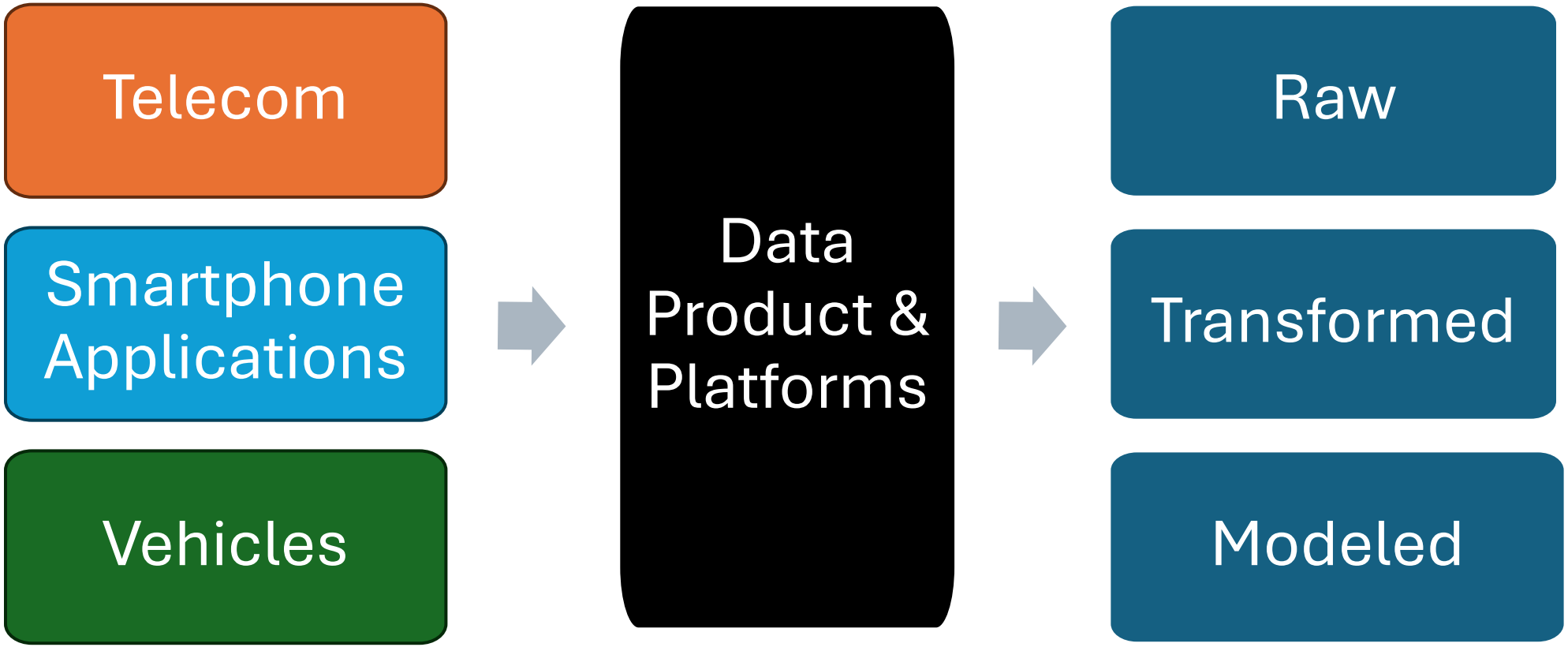


Applications

- Visits
- Trips
- Routes
- Flows

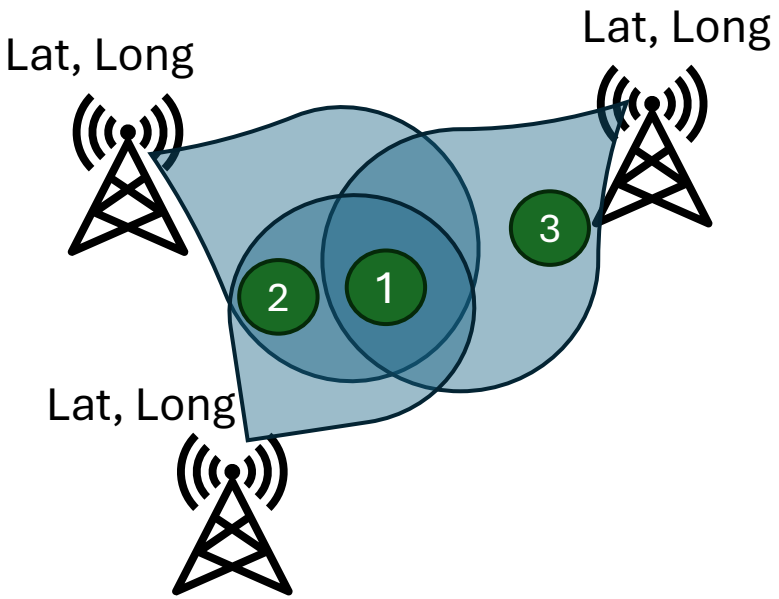
Data Sources

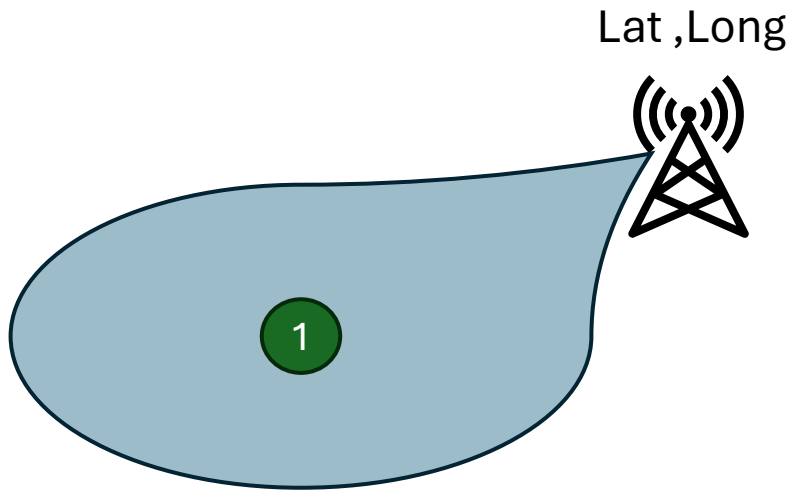
Medium of Mobility Data Sources



Telecom

- Location Estimation (coarse)
- Call/Text Signal Transmission
- Carrier sample (wide coverage)
- Diverse and Persistent sample





- Value :
  - Regional planning
  - External studies
  - Long distance travel analysis
- Trade Offs:
  - Rural coverage (function of towers)
  - Geographic bias (carrier cost/choice)
  - Could see change with 5G

Telecom Data : Typical Schema

- Typical Schema
  - Location
  - Timestamp
  - Device ID
  - Message Type : Call, Text/SMS etc.
- Characteristics
  - Persistence in pings and devices
  - Wider coverage
  - Coarser location resolution
- Use Type
  - Better represents large zone travel
  - Regional and Long-Distance Flows
  - External Analysis
  - Internal Regional Flows (Resident | Visitor Segments)

Anonymized/hashed phone and subscriber id

Time Stamp and cell tower location for mobility analysis by subscriber --→ yields activity locations, estimated home/work by night/daytime locations


```
[
  {
    "equip_hash": "146dabee-0d43-4fdf-8a04-89b02fd704a3",
    "sub_hash": "d601f36e-353b-4266-87e4-6f4c2d7312e8",
    "call_start_timestamp": "2023-03-24 13:08:23",
    "call_end_timestamp": "2023-11-29 08:30:24",
    "base_station_calling_id": "BSID4434",
    "base_station_calling_latitude": -39.293004,
    "base_station_calling_longitude": 37.261319,
    "activity_type": "SMS",
    "cell_service": "3G"
  },
  {
    "equip_hash": "3b6e7564-c35a-41ce-9150-cface05f8bf3",
    "sub_hash": "c2bd612c-5142-4b6f-981e-952c671d63c7",
    "call_start_timestamp": "2023-10-13 03:03:34",
    "call_end_timestamp": "2023-11-23 15:16:29",
    "base_station_calling_id": "BSID6210",
    "base_station_calling_latitude": -63.742168,
    "base_station_calling_longitude": -146.070934,
    "activity_type": "Voice",
    "cell_service": "3G"
  },
  {
    "equip_hash": "3420cc8c-0a25-4458-947f-32ee8ef27fdd",
    "sub_hash": "a1b6a826-4dcb-47b1-80cd-8bd17744c644",
    "call_start_timestamp": "2023-04-17 11:16:18",
    "call_end_timestamp": "2023-12-14 10:06:24",
    "base_station_calling_id": "BSID7478",
    "base_station_calling_latitude": 33.916246,
    "base_station_calling_longitude": -62.730307,
    "activity_type": "SMS",
    "cell_service": "2G"
  }
]
```

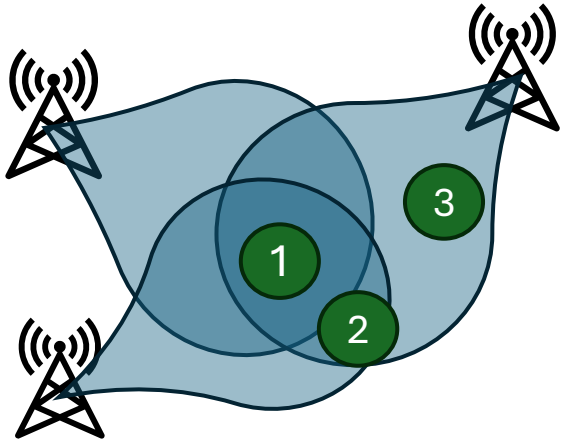
Smartphone Applications

Application dependent (Weather, Games)

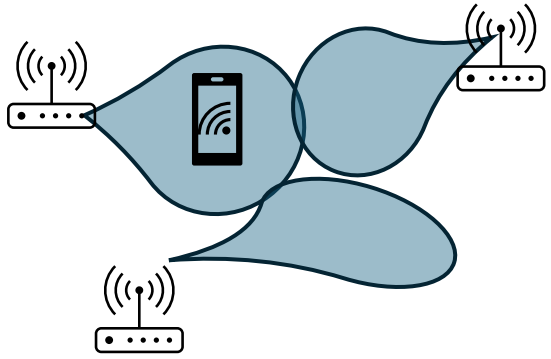
OS dependency (iOS & Android)

Location:

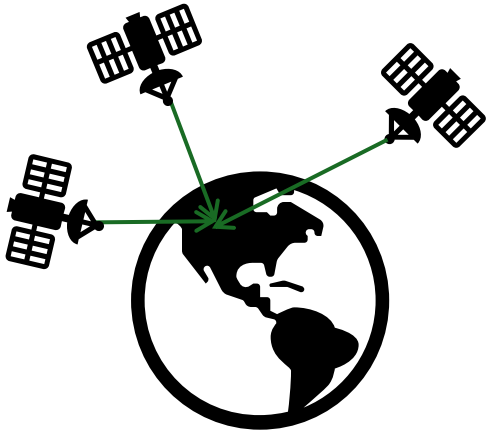
- Telecom, GPS or Wi-Fi
- App configuration dependent



Telecom


Wi-Fi

Data Collected For : Advertising


GPS

- Value :
  - Device persistence
  - POI Visits / Trip Ends/ Generation
  - Select regional travel patterns
- Trade Offs:
  - App Dependent Sample
  - Lack of persistent pings
  - Variable Location accuracy

Smartphone Applications: Typical Schema



- Typical Schema
  - Location
  - Timestamp
  - Advertising ID
- Characteristics
  - Device Persistence
  - App activation dependent pings
  - Specific coverage – use areas such as malls etc.
  - Location Resolution depends on : Telecom, Wi-Fi, GPS
- Use Type
  - Better represents visits to points of interests, stores
  - Trip Generation, Visitor Analytics, Regional Travel
  - External Analysis (Needs routing – no breadcrumbs)
  - Internal Regional Flows (Resident | Visitor Segments)

Anonymized Advertising Id and Phone Operating System

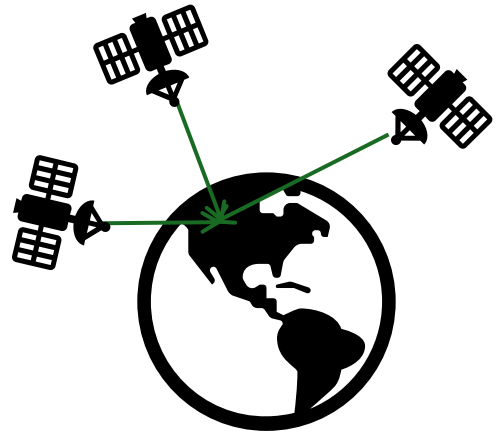
If location estimated using Wi-Fi (there could be an IP)


```
[
  {
    "adv_hash_did": "6b2e3e18-3c51-4e13-8a91-c3f9e8f78e3f",
    "os_type": "iOS",
    "device_ping_timestamp": "2023-03-14 12:34:56",
    "device_ping_latitude": 37.7749,
    "device_ping_longitude": -122.4194,
    "device_ip_address": "192.168.1.1",
    "location_accuracy": 10.5
  },
  {
    "adv_hash_did": "8f42e2b7-5a54-4b3d-85d9-0b7e5f4b63b8",
    "os_type": "Android",
    "device_ping_timestamp": "2023-07-22 15:45:30",
    "device_ping_latitude": 40.7128,
    "device_ping_longitude": -74.0060,
    "device_ip_address": "10.0.0.1",
    "location_accuracy": 20.0
  },
  {
    "adv_hash_did": "3c1a4c69-5f72-4a2e-b8f1-7a9e4b5a2d0e",
    "os_type": "iOS",
    "device_ping_timestamp": "2023-11-01 08:12:45",
    "device_ping_latitude": 34.0522,
    "device_ping_longitude": -118.2437,
    "device_ip_address": null,
    "location_accuracy": 15.3
  }
]
```


Location accuracy could vary based on estimation technology (Tower, Wi-Fi, GPS and its availability) and is also dependent on App design

Vehicles

- Mobility Data (GPS Enabled)
  - Waypoints
  - Trips
  - Trajectories
- Vehicle Behavior (Not all is available)
  - Hard braking
  - Acceleration
- Data Collected For: Telemetry



GPS

Value :

- Relatively High Location Accuracy
- Traffic Operations Analysis
- Trip Persistence

Trade Offs:

- Lack of Device persistence (planning)
- Trip End Definition (On/Off) – Linking
- Vehicle Type Bias (OEM)

Fleet Vehicles: Typical Data Schema

- Typical Schema
  - Trip ID, Provider ID, Device ID (may or may not be persistent)
  - Timestamp
  - Location Latitude, Longitude
  - Vehicle Weight Class (1: < 14K lbs., 2 : 14k-26K lbs., 3: > 26k lbs.)
  - Provider Driving Profile (Consumer, Taxi, Delivery/Service, Fleet)
  - Probe Source Type (1- Embedded GPS or 2-Mobile Device)
- Characteristics
  - Location is obfuscated (still better than cell phone location)
  - Additional attributes, predominately represents commercial travel
  - Not all devices have persistent ID across trips or over days
  - Includes trajectories (route data)
- Use Type
  - External travel analysis
  - Upcoming applications to understand internal commercial travel
  - Truck parking and operational analysis (congestion)


```
[
  {
    "trip_id": "f68274c0-f94e-46ea-97b3-c5e45a4b4ae5",
    "device_id": "DEV295",
    "provider_id": "PROV007",
    "mode": 2,
    "start_date": "2023-09-18T14:23:09",
    "end_date": "2023-12-02T02:17:15",
    "start_latitude": -11.044691,
    "start_longitude": -101.034402,
    "end_latitude": 32.755458,
    "end_longitude": -39.197758,
    "provider_type": 2,
    "driving_profile": 1,
    "probe_source_type": 2,
    "vehicle_weight_class": 1
  },
  {
    "trip_id": "080702dc-daec-4df4-bf9a-140cc27d584f",
    "device_id": "DEV466",
    "provider_id": "PROV010",
    "mode": 5,
    "start_date": "2023-07-12T13:39:46",
    "end_date": "2023-08-15T06:06:23",
    "start_latitude": -85.143501,
    "start_longitude": -113.895258,
    "end_latitude": 10.674406,
    "end_longitude": 122.747675,
    "provider_type": 1,
    "driving_profile": 2,
    "probe_source_type": 1,
    "vehicle_weight_class": 2
  },
  {
    "trip_id": "ddbce643-4b0b-497f-ac5f-4b27ae258389",
    "device_id": "DEV376",
    "provider_id": "PROV010",
    "mode": 5,
    "start_date": "2023-03-02T14:44:36",
    "end_date": "2023-09-06T09:03:56",
    "start_latitude": -82.088141,
    "start_longitude": 52.335376,
    "end_latitude": 50.565306,
    "end_longitude": 175.105253,
    "provider_type": 1,
    "driving_profile": 3,
    "probe_source_type": 2,
    "vehicle_weight_class": 3
  },
```

Connected Car Vehicles: Typical Schema

- Typical Schema
  - Trip | Journey (depends on ignition status)
  - No persistence (cannot consolidate trips by vehicle)
  - Vehicle type, fuel, events (operational applications)
- Characteristics
  - Trips and Trajectories
  - OEM dependent (biased towards a brand/newer cars)
  - Persistent location and best resolution
- Use Type
  - Better represents trajectories
  - Operation analysis (Safety, Queuing, Intersection Use)
  - Planning (explored for non-commercial trip table)


```
{
  "data_record_id": "1fbb6e3b-441e-4b89-82c1-0e8b9a1e345c",
  "trip_journey_id": "a4d1e7f4-5a3d-4d8b-91a1-3e9c6e4c8d2b",
  "vehicle_type": "sedan",
  "vehicle_fuel": "gas",
  "vehicle_year": "2021",
  "record_timestamp": "2023-06-15 12:45:30",
  "record_latitude": 34.052235,
  "record_longitude": -118.243683,
  "record_speed": 65.5,
  "record_heading": "N",
  "record_ignition_status": "on",
  "event_type_id": "event123",
  "event_type": "seat belt"
},
{
  "data_record_id": "2e5b9c76-5a4e-49a1-8212-2c9e7b5d8e2a",
  "trip_journey_id": "b2c4d6e3-6f4b-4e7d-92b1-4f8d7e5c8d3c",
  "vehicle_type": "suv",
  "vehicle_fuel": "diesel",
  "vehicle_year": "2018",
  "record_timestamp": "2023-07-22 15:50:45",
  "record_latitude": 40.712776,
  "record_longitude": -74.005974,
  "record_speed": 55.0,
  "record_heading": "E",
  "record_ignition_status": "off",
  "event_type_id": "event456",
  "event_type": "door status"
},
{
  "data_record_id": "3f9b8e12-7c4d-48b1-9312-3e7f9e6b8d2c",
  "trip_journey_id": "c3d5e8f7-7f5b-4d9c-93b2-5f9d8e6c8d4d",
  "vehicle_type": "pick-up",
  "vehicle_fuel": "hybrid",
  "vehicle_year": "2021",
  "record_timestamp": "2023-08-30 10:15:20",
  "record_latitude": 51.507351,
  "record_longitude": -0.127758,
  "record_speed": 45.2,
  "record_heading": "S",
  "record_ignition_status": "on",
  "event_type_id": "event789",
  "event_type": "braking"
}
]
```

Location Estimation Technologies

- Phone Based: Legacy Source, Larger Sample and Wider Coverage, Low Resolution Location Estimate, Regional Applications – 4G/LTE…..5G (finer resolution in the future ?)

- Smartphone Application: Current Source, Variable Sample – highly dependent on app, High Location Resolution, Lacks persistence to consistently identify trips, Local Site Visits/Trip Ends

- Vehicle Based: Persistent and High-Resolution Sample, Link Level/Operational Applications, Specific to Vehicles. Roadway Traffic Operational Analysis

Medium	Cellular Phones	Smartphone Applications	Vehicles
Telecom	X	X	
Wi-Fi		X	
GPS		X	X

Trivia 1

- Which source has the largest sample size in terms of representing passenger travel
  - Telecom - widest subscriber bases (carrier dependent bias – red vs. yellow vs. purple)
- Which source is more conducive to representing trip ends / attractions to point of interest
  - Smartphone Applications – influenced by geofencing – ad targeting function (app dependent)
- Which source is most conducive to representing travel paths/breadcrumbs
  - Vehicles – representing passenger from vehicle (connected car) and/or commercial fleet data (OEM)
- Does any source provide true activity location
  - No - All are estimated based on derived location and stay point estimation procedures & algorithms

Data Products

Data Products

- Raw (Sample) :
  - Delivered as is, with obfuscation : GPS trips, trajectories, waypoints, LBS logs
  - Requires in house storage solutions, staff, and big data skills
  - Custom analysis, integration and visualization
- Transformed :
  - Delivered as aggregated scaled product : trip tables, visits for points of interest
  - Requires specification for study area, period etc.
  - Inputs for in house downstream analysis, resulting visualization etc.
- Modeled :
  - Delivered as aggregated synthetic product : modeled – trips by purpose, mode etc.
  - Accessible via vendor platform as standard product or customized
  - Likely cannot be used for in house analysis, possible to produce corresponding visuals

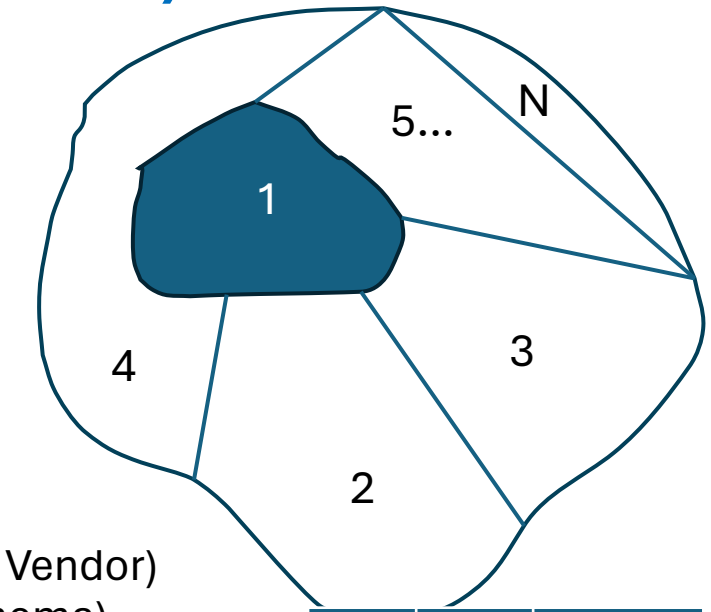
Visits - Points of Interest Products

- Typical Attributes
  - Daily Visits by Polygon (custom ,place key, parcels)
  - Business Name, NAICS
  - TOD, DOW, Home Geographies(Census)
- Characteristics
  - App activation dependent pings
  - Specific coverage – use areas such as malls etc.
  - Resolution dependent on : Telecom, Wi-Fi, GPS
- Applications
  - Visitation Analysis (Real Estate, Parks etc.)
  - Special (Game/Concerts) Events, Weather Events
  - Site Trip Generation (Special Generators)
- Source
  - Smartphone Applications (Primarily LBS data sources)


```
[
  {
    "placekey": "3e3c5f5a-2d5e-4a9e-9b2e-6e3c4f5a6b8e",
    "business_name": "Starbucks",
    "business_lat": 37.7749,
    "business_long": -122.4194,
    "start_date_range": "2023-01-01",
    "end_date_range": "2023-01-07",
    "visits_by_day": 120,
    "visitor_home_blockgroups": ["060750601001", "060750601002", "060750601003"],
    "visits_by_home_blockgroups": [30, 50, 40],
    "median_visit_dwell_time": 15.5,
    "dwell_time_distribution": [10, 20, 30, 40, 20]
  },
  {
    "placekey": "4d6e7f7a-3f7e-4b9e-9d5e-7f6d5f7a8c9e",
    "business_name": "McDonald's",
    "business_lat": 40.7128,
    "business_long": -74.0060,
    "start_date_range": "2023-02-01",
    "end_date_range": "2023-02-07",
    "visits_by_day": 200,
    "visitor_home_blockgroups": ["360610097001", "360610097002", "360610097003"],
    "visits_by_home_blockgroups": [70, 80, 50],
    "median_visit_dwell_time": 20.0,
    "dwell_time_distribution": [15, 25, 35, 45, 30]
  },
  {
    "placekey": "5e7e9f9a-4e9e-4c9e-9e5e-8f7e6f9a8d0e",
    "business_name": "Walmart",
    "business_lat": 51.507351,
    "business_long": -0.127758,
    "start_date_range": "2023-03-01",
    "end_date_range": "2023-03-07",
    "visits_by_day": 300,
    "visitor_home_blockgroups": ["E02004761", "E02004762", "E02004763"],
    "visits_by_home_blockgroups": [100, 120, 80],
    "median_visit_dwell_time": 25.3,
    "dwell_time_distribution": [20, 30, 40, 50, 40]
  }
]
```

Trips (Origin-Destination) Products

- Typical Schema
  - Custom Defined Origin - Destination
  - Day of Week (Average | Total | Select Dates)
  - Time of Day (Daily | Peak | off-Peak)
  - Purpose (HBW | HBO| NHB)
  - Mode (Auto | Bike | Transit)
- Characteristics
  - Standard or Custom Zone Structure (Work with Vendor)
  - Available for pre-defined configuration (see Schema)
  - Could be sourced from Raw, Transformed or Modeled data
- Use Type
  - Regional - External Analysis (To/From, Through Region)
  - Long Distance Travel Analysis
  - Internal Regional Flows (Resident | Visitor Segments)

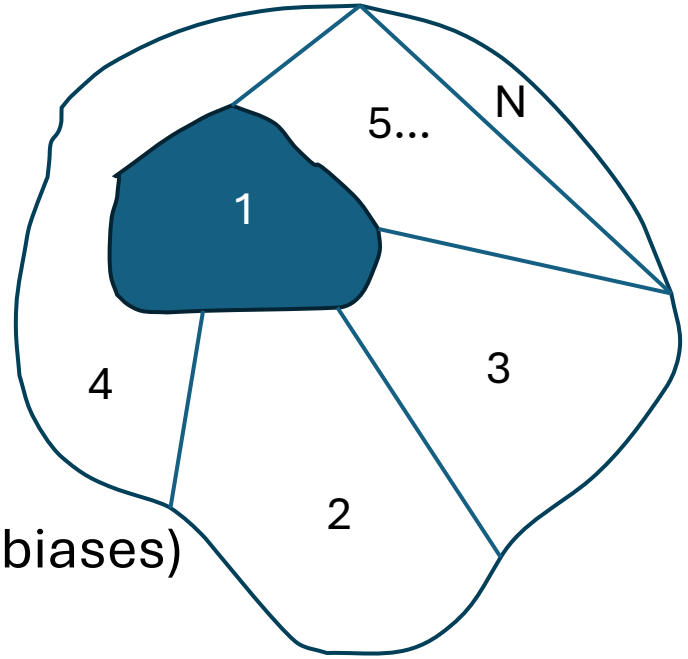


O	D	Trips
1	1	234
1	2	324
3	5	646
.	.	457
N	1	435
N	4	454
.	.	.
N	3	234


```
[
  {
    "origin": 12,
    "destination": 34,
    "purpose": "work",
    "time_of_day": "am",
    "day_of_week": "weekday",
    "mode": "auto",
    "trips": 15
  },
  {
    "origin": 56,
    "destination": 78,
    "purpose": "non-work",
    "time_of_day": "pm",
    "day_of_week": "weekend",
    "mode": "walk",
    "trips": 20
  },
  {
    "origin": 23,
    "destination": 45,
    "purpose": "other",
    "time_of_day": "mid-day",
    "day_of_week": "custom",
    "mode": "bike",
    "trips": 10
  },
  {
    "origin": 67,
    "destination": 89,
    "purpose": "work",
    "time_of_day": "am",
    "day_of_week": "weekday",
    "mode": "transit",
    "trips": 30
  },
  {
    "origin": 34,
    "destination": 56,
    "purpose": "non-work",
    "time_of_day": "pm",
    "day_of_week": "weekend",
    "mode": "auto",
    "trips": 25
  }
]
```

Select Zone : Trips Product Derivative

- Product
  - Raw : Observed Trips (Source dependent)
  - Transformed: Adjusted Sample Observations
  - Modeled Outputs: ~ Regional Model Results
- Characteristics
  - Small Sample based on (day, period etc.)
  - Transform/Modeled (adjusted to expansion – biases)
- Uses
  - New Developments in zone (before/after, trends)
  - Cautious about biases (Sample Vs. Expanded Results)
  - Check distribution with other sources (sanity checks)



O	D	Trips
1	2	234
1	3	324
1	4	645
1	5	45
.	.	646
.	.	457
1	N	435

Select Link : Trips & Trajectories Derivative

- Product
  - Raw : Observed Trips/Trajectories
  - Transformed: Adjusted Sample Observations
  - Modeled Outputs: ~ Regional Model Results (Agent)

O	D	Trips
1	2	25
3	5	5
12	24	11
1	7	17
.	.	6
.	.	19
21	79	17

- Characteristics
  - If sample, could be small based on slicing (day, period etc.)
  - Transformed /Modeled (adjusted or calibrated to regional controls)

- Uses
  - Corridor analysis (Auto, Truck, Transit Trips)
  - Caution about biases (Sample Vs. Adjusted Results)
  - Check distribution with other sources/counts

Trivia 2

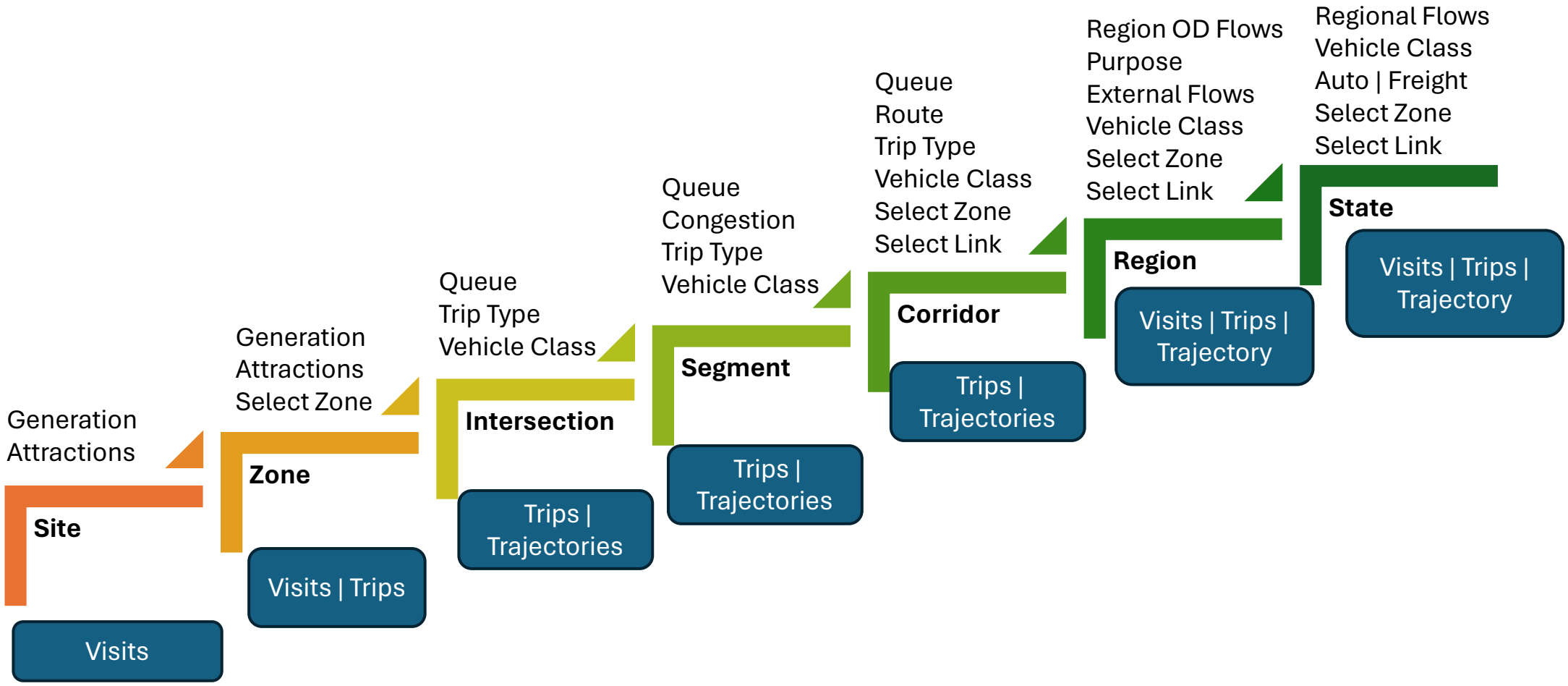
- What type of data sources can support select link analysis ?
  - Select link analysis can be produced from raw (sample trips/trajectories) or modeled data

- Which data sources provide information on resident/visitor and home location estimates ?
  - Predominately LBS and Telecom based products with Device ID persistence makes it feasible

- Are travel purposes and modes reported in passive data imputed or observed ?
  - Most data sources impute these attributes, least likely to be observed data

Applications : Part 1

Applications : A spatial perspective



Visits – Select Application

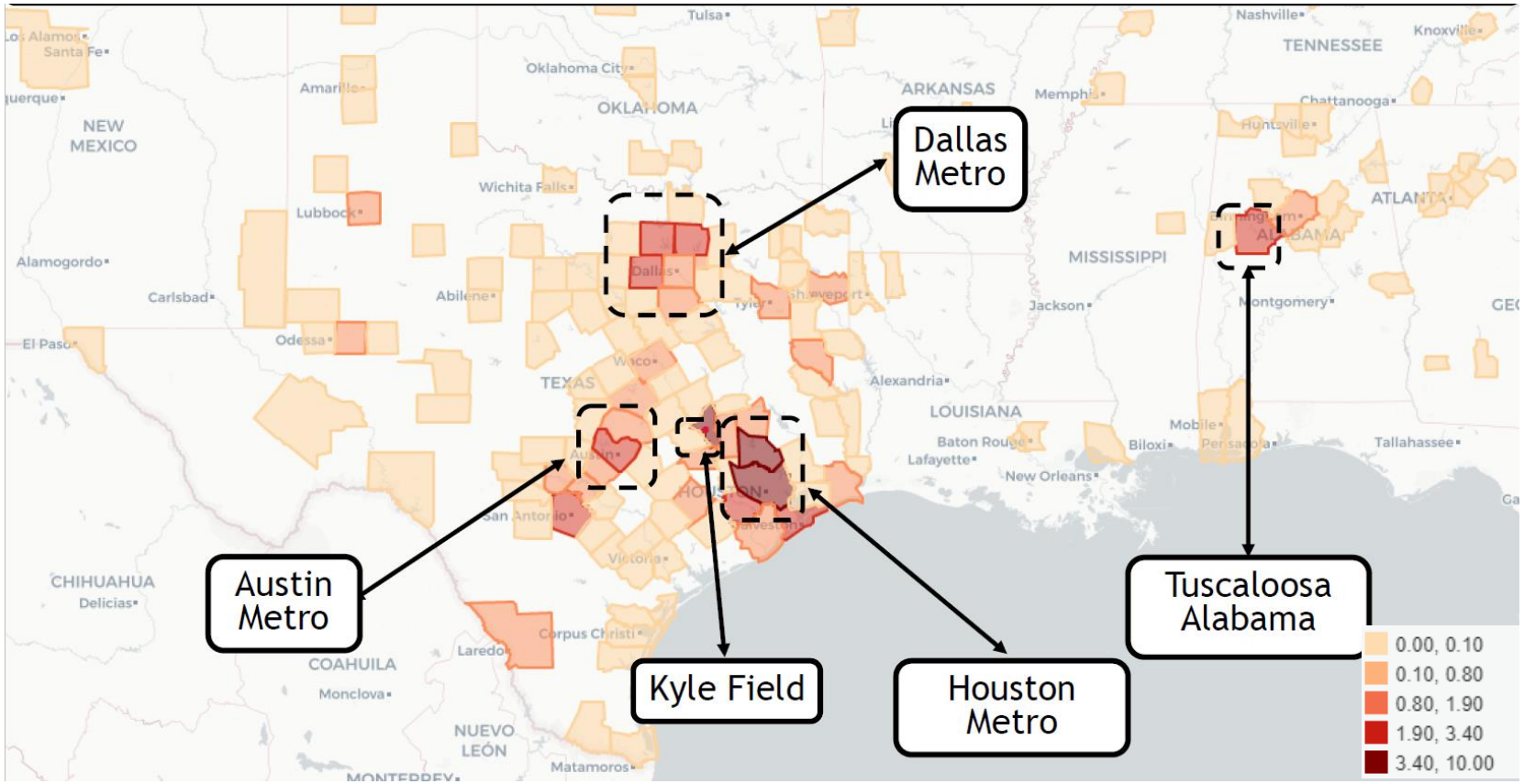
- Investigating:
  - Special Events : Gameday | Concerts
  - Inclement Weather
  - Sources : LBS Data, Demographics

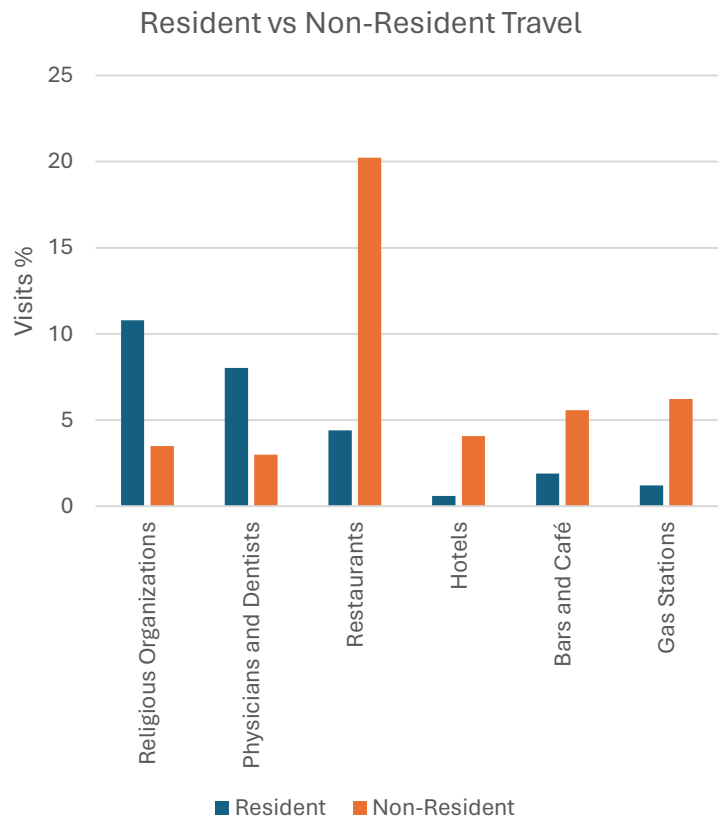
- Sources
  - Daily Visits to Business (LBS daily visits: SafeGraph)

- How to use/inferences to make
  - Regional shift in visitations
  - Time to recovery
  - Can be used in combination with trips/trajectories to changes in local travel

Select Zone/Polygon : Visits Application

Kyle Field, Bryan College Station : Alabama Vs Texas A&M (Oct 9, 2021)



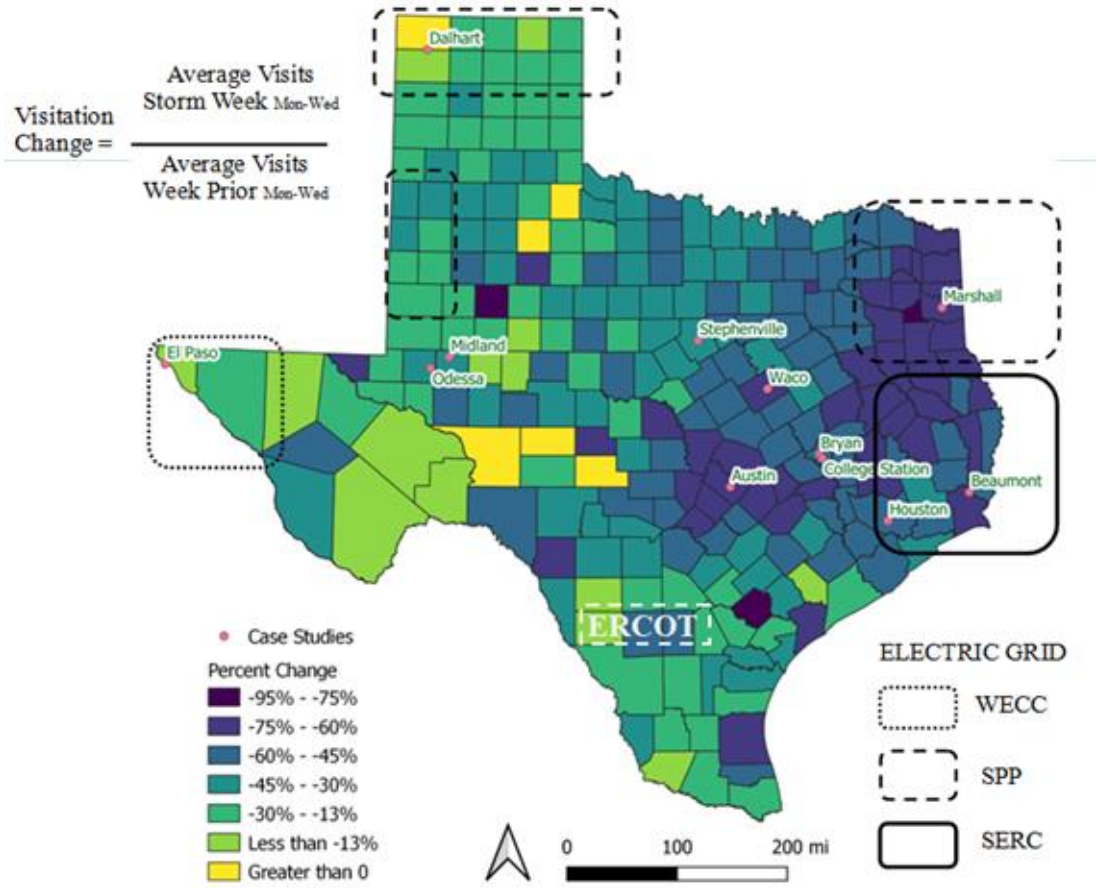


Work done as part of TxDOT - FY 23 Inter Agency Contract (IAC) evaluating potential applications and advances in using passive data

Texas Department of Transportation

Texas A&M Transportation Institute

Inclement Weather : URI 2021- Recovery



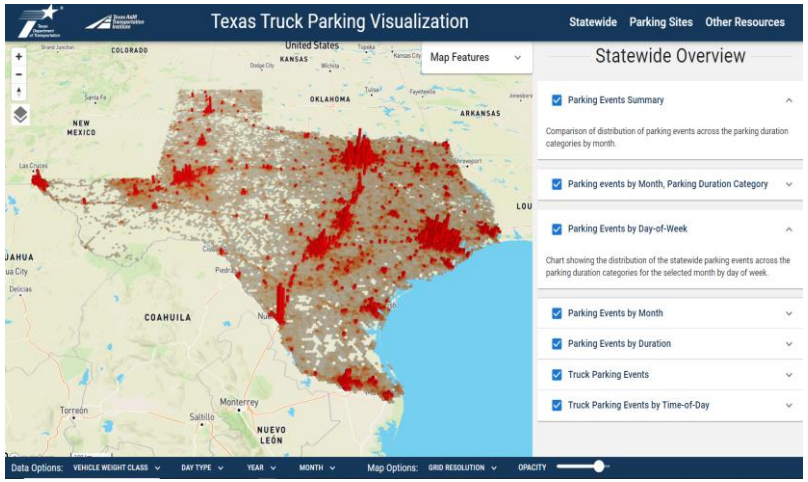
	Region	Austin	El Paso	Houston	Beaumont	Bryan College Station	Waco	Midland Odessa	Dalhart	Marshall	Stephenville
Recovery Rate	14 Feb	0.38	1.08	0.74	0.74	0.44	0.28	0.34	0.51	0.56	0.36
	15 Feb	0.29	1.21	0.28	0.20	0.36	0.30	0.49	0.94	0.21	0.43
	16 Feb	0.47	1.26	0.50	0.38	0.53	0.52	1.16	1.03	0.37	0.59
	17 Feb	0.40	0.98	0.57	0.59	0.49	0.37	0.95	1.12	0.27	0.67
	18 Feb	0.47	1.00	0.78	0.93	0.67	0.51	1.12	1.02	0.40	0.82
	19 Feb	0.83	1.03	1.04	1.14	1.01	1.01	1.17	1.23	0.58	1.08
	Days to Recover	6	0	5	5	5	5	2	2	9	5
Feb 14th—	Snowfall(in)	5.7	4.3	0.76	2.1	3.05	3.89	4.12	6.4	10.4	4.1
	Precipitation (in)	1.44	0.4	1.76	2.83	2.79	0.59	0.70	0.5	1.61	0.31
	Days below 32F(Max)	5	0	1	1	3	5	5	5	2	5
	Days below 32F (Min)	6	5	6	5	6	6	6	6	6	6
	Recovery Scale	Low				Medium			High		

Work done as part of TxDOT - FY 23 Inter Agency Contract (IAC) evaluating potential applications and advances in using passive data

Texas Department of Transportation

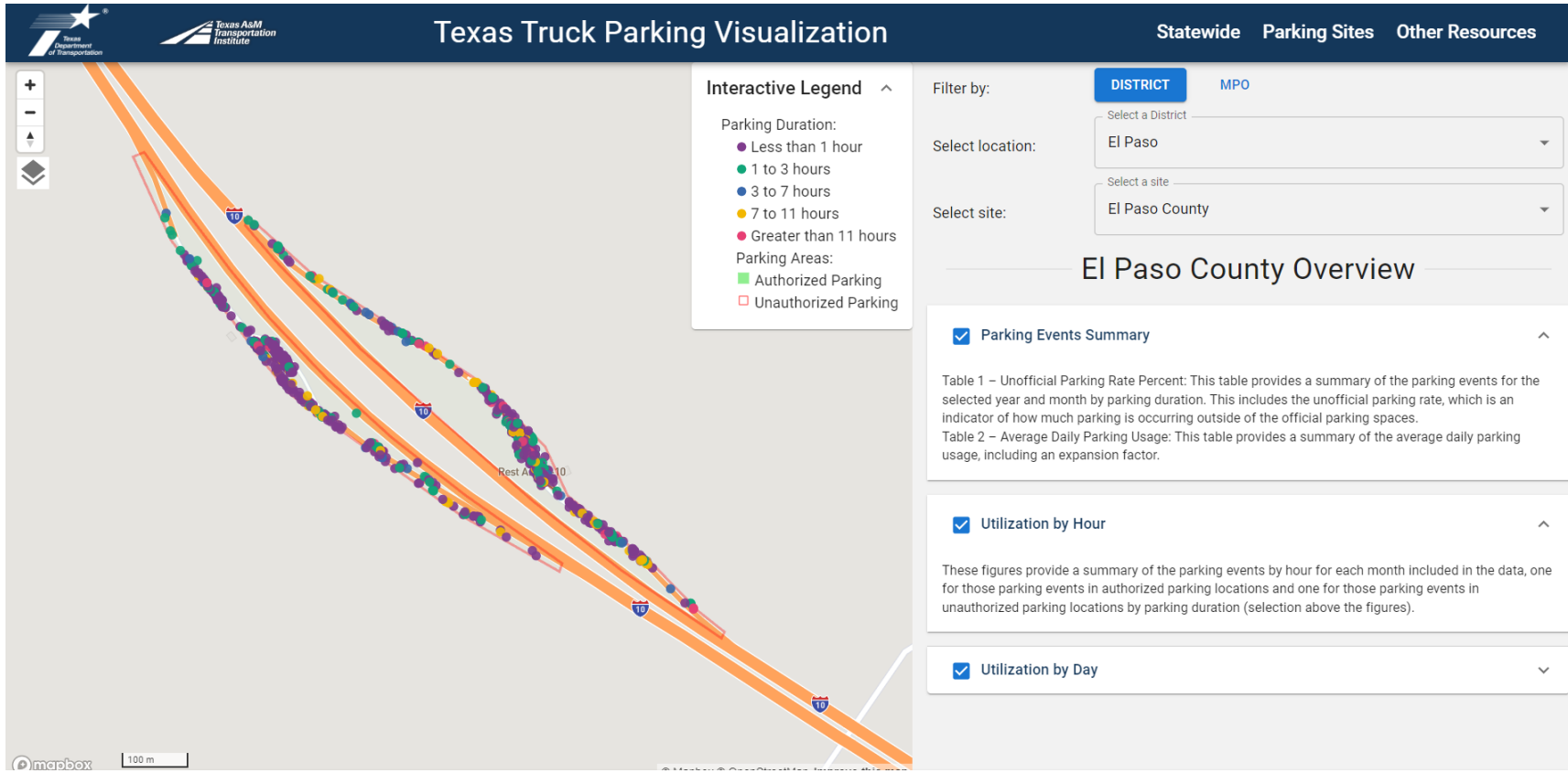
Texas A&M Transportation Institute

Sample Trips: Truck Parking



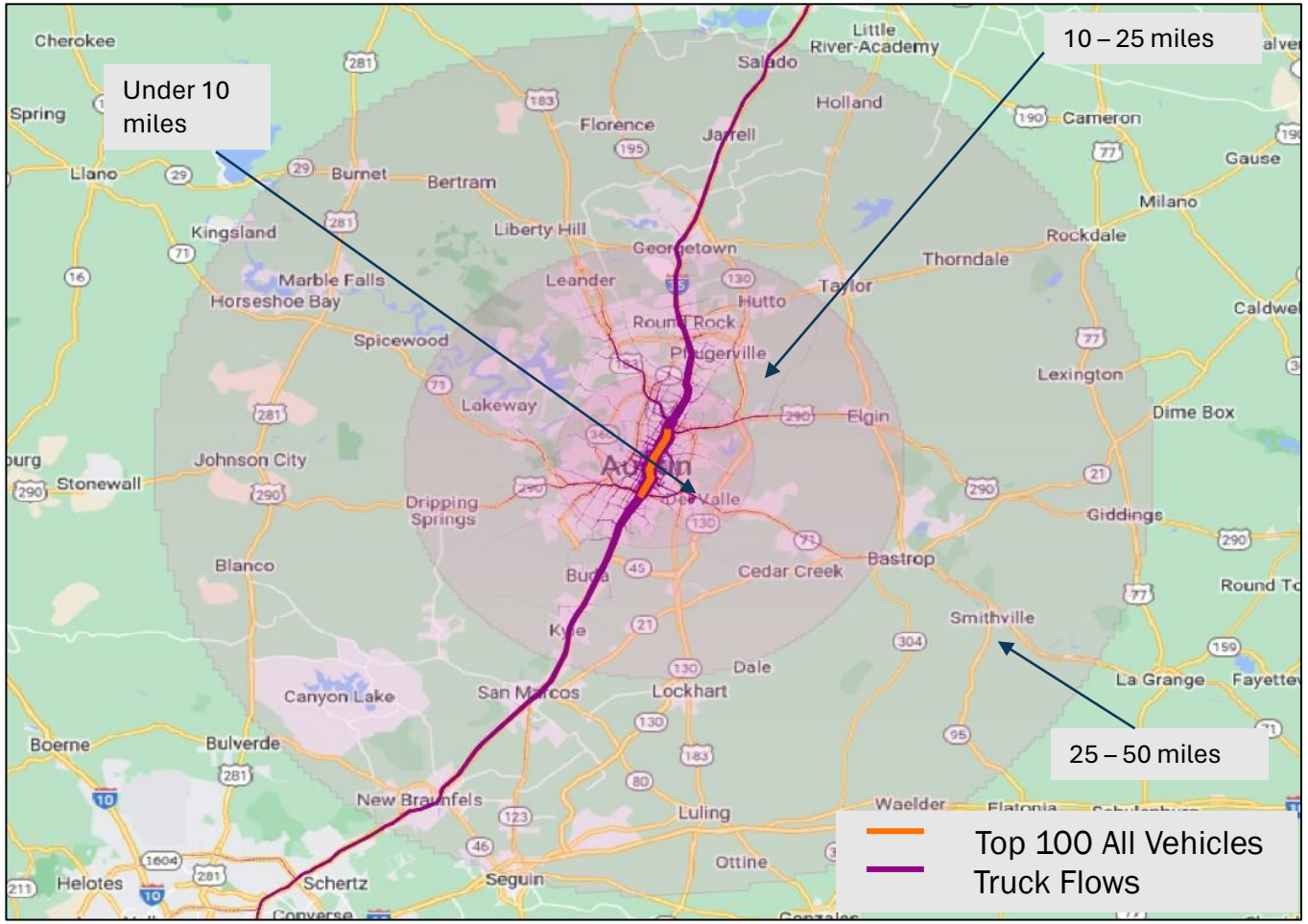
https://trkparkingtx.tti.tamu.edu/

Attribute : Non-moving trips
Data Sources: Trajectory, Waypoints

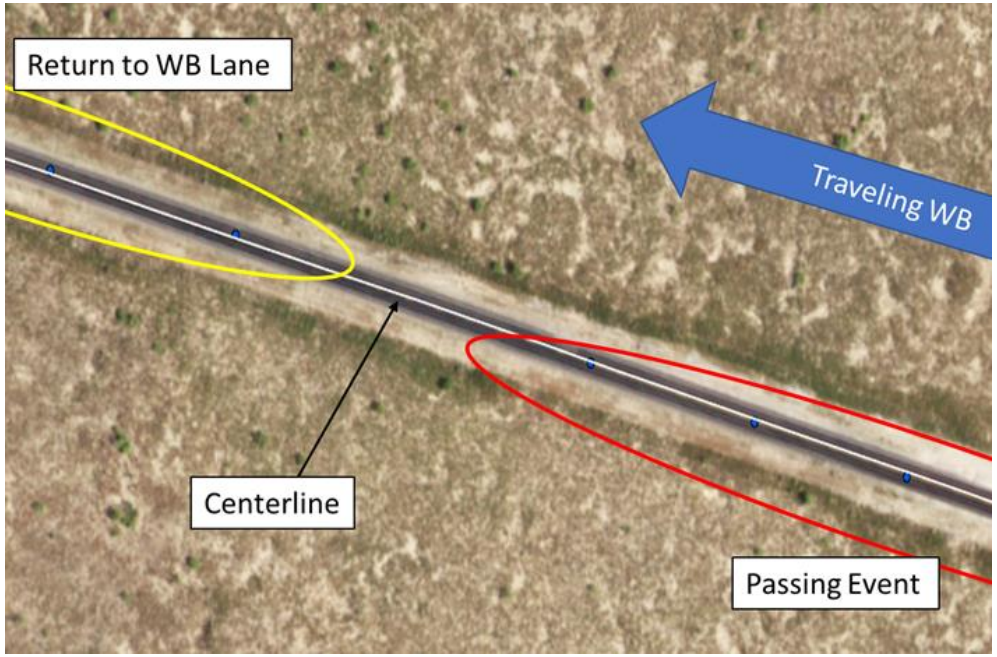


Sample Trips and Trajectories : Select Link

Range	% of Trips	
	All (Total)	Truck (Total)
Under 10 miles	80.3	17.1
10 to 25 miles	6.7 (**87.0**)	8.4 (25.5)
25 to 50 miles	5.5 (**92.5**)	29.3 (**54.8**)
Over 50 miles	7.5 (100.0)	45.2 (100.0)



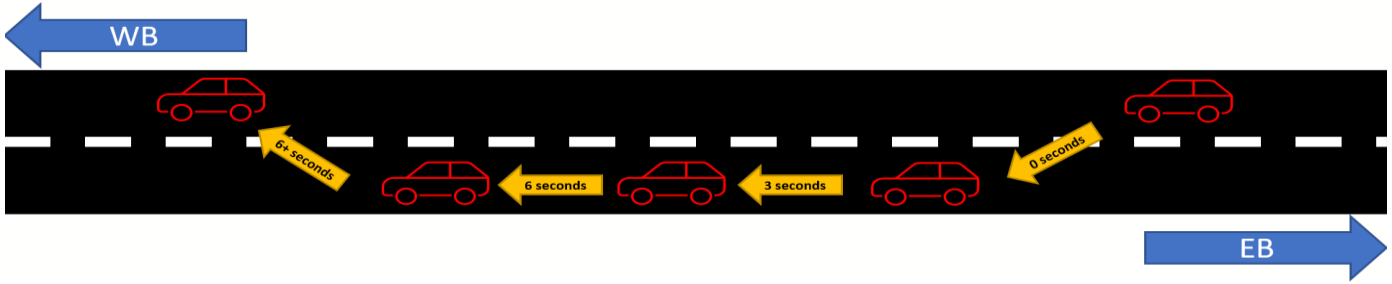
Passing Risk : Behavior Analysis Using CV Data



Daily Passing Events along State Highway 128	
60% started and ended in a passing zone ✓	30% *started* in non-passing zone but *ended* in passing zone ⚠
4% *started* in passing zone but *ended* in non-passing zone ⚠	6% started and ended in a non-passing zone ✗

Two criteria for inclusion:

- Pass from WB to EB lane for minimum of 6 seconds (before returning to WB lane)
- While passing, must increase speed a minimum of 6.21 mph (10km/h)



Bike | Ped Example - Integrating Multiple Sources

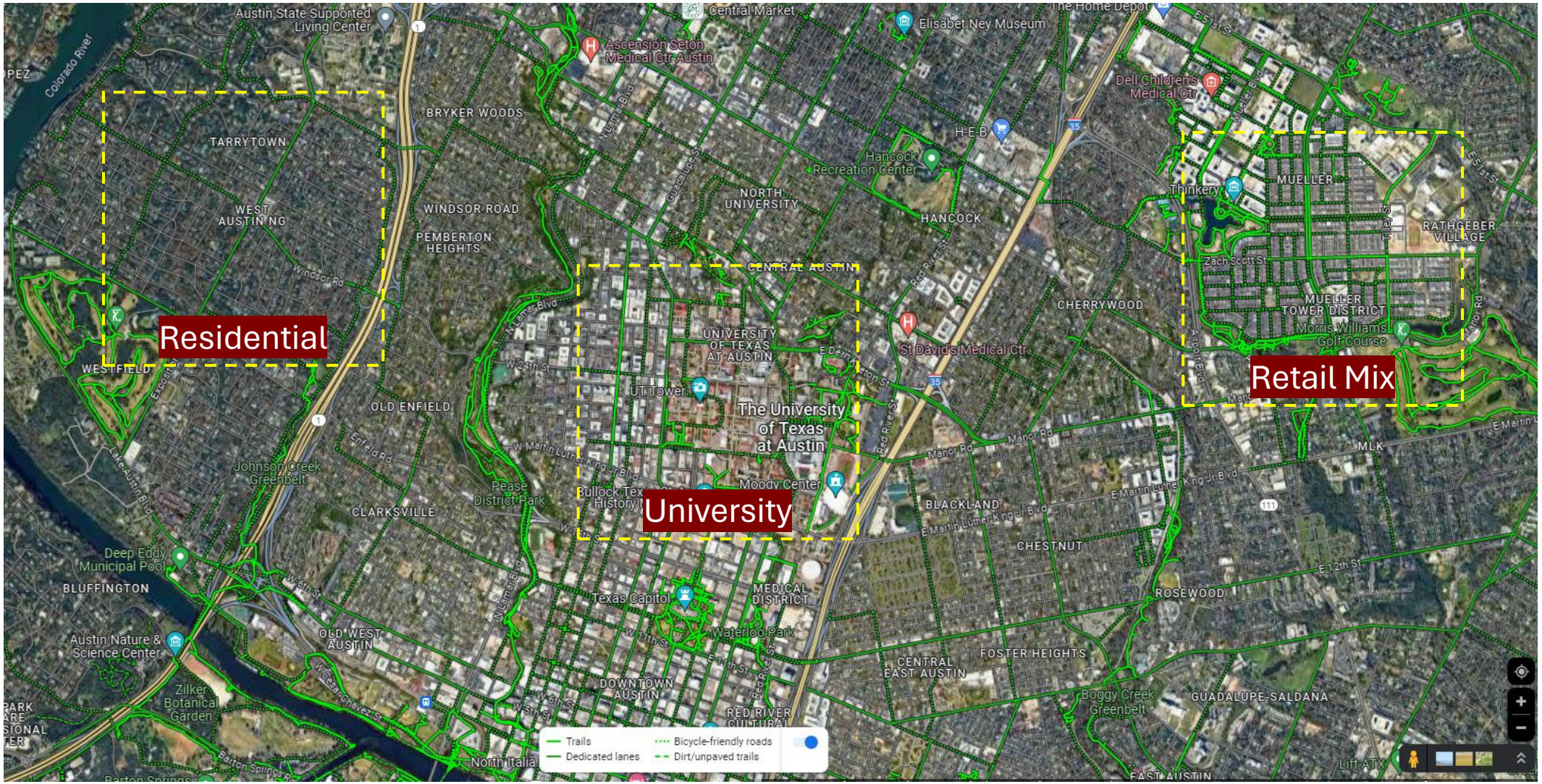
- Investigating: Latent Demand for Bike | Ped Trips
  - Mixed Use Environments
  - Potential for Conversion
  - Safety
  - Demographics

- Sources
  - Short Trips (GPS raw trips: INRIX)
  - Daily Visits to Business (LBS daily visits: SafeGraph)
  - Mode Split (Modeled trips: Replica)

- How to use/inferences to make
  - Not all data and everywhere is good
  - Use a mix of sources to get a relative measure
  - Combined local knowledge and expertise with diverse data

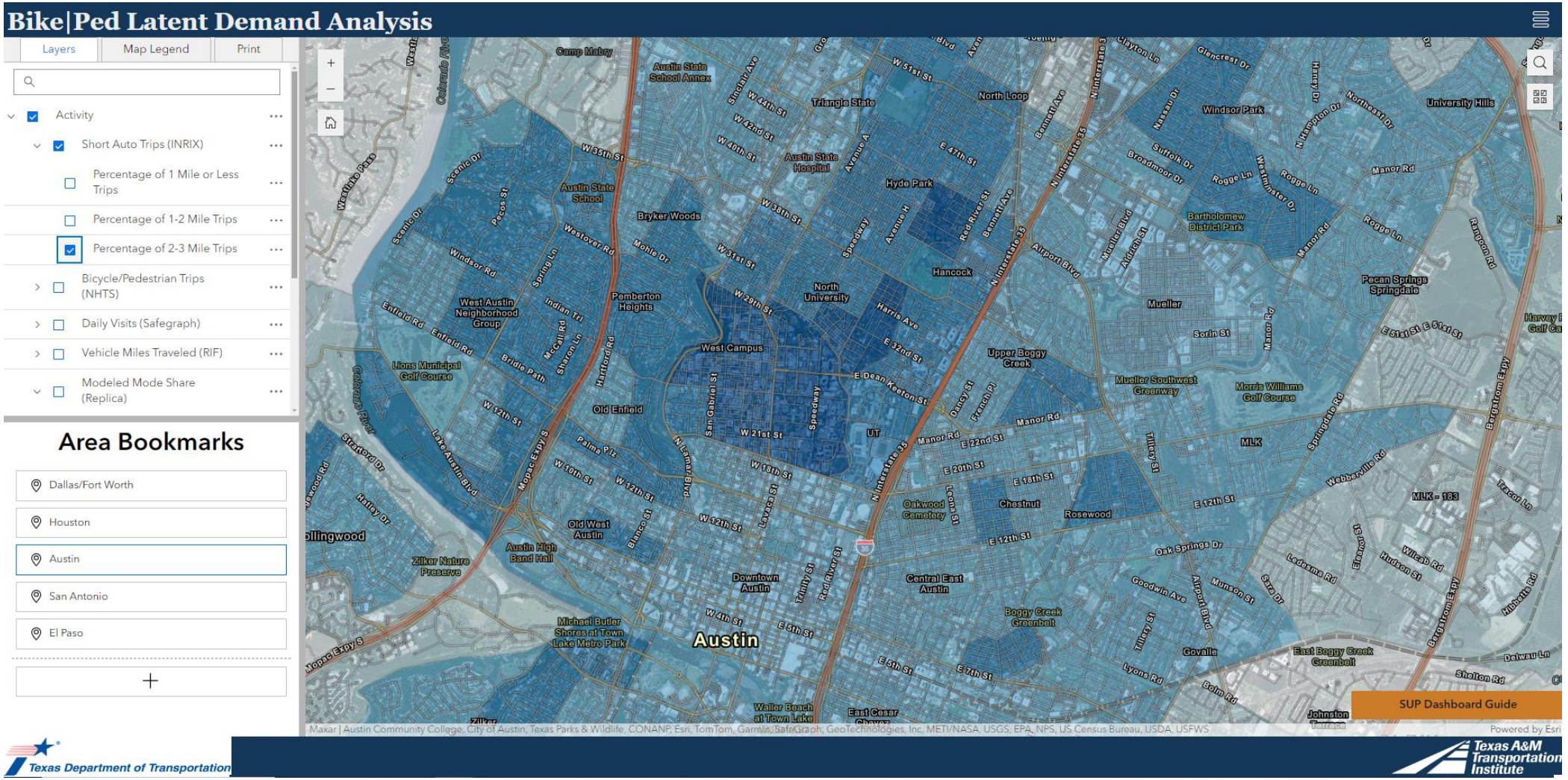
Based on an ongoing project “Estimating Latent Bicyclist and Pedestrian Demand for Shared Use Path Design”, funded by the Texas Department of Transportation (TxDOT) 0-7152. Principle Investigator: Ipek N. Sener (i-sener@tti.tamu.edu)

Austin Example – Google Bike/Ped Facilities



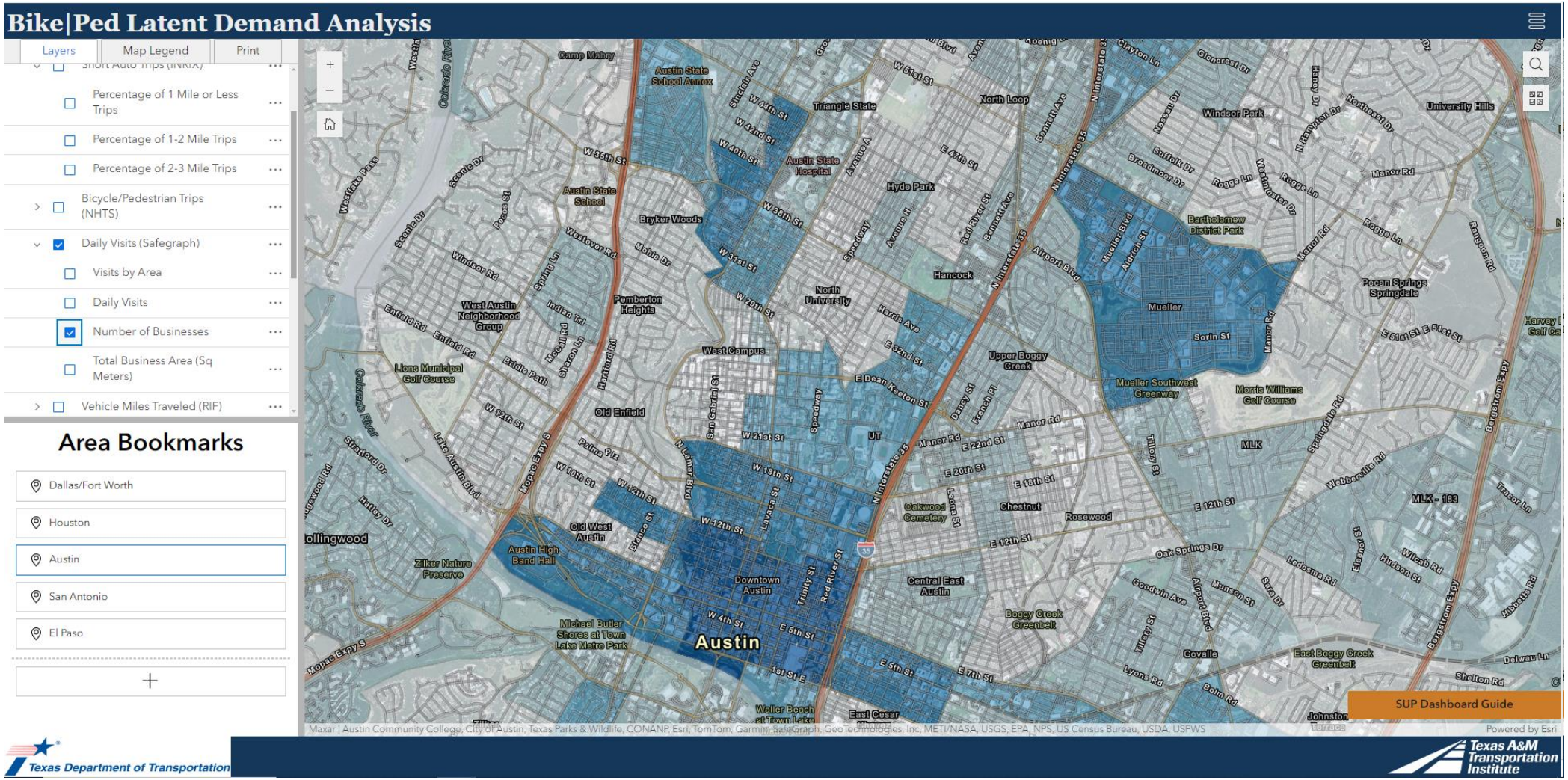
Based on an ongoing project "Estimating Latent Bicyclist and Pedestrian Demand for Shared Use Path Design", funded by the Texas Department of Transportation (TxDOT) 0-7152. Principle Investigator: Ipek N. Sener (i-sener@tti.tamu.edu)

Bike | Ped Example : Austin – Raw Short Trips



Based on an ongoing project "Estimating Latent Bicyclist and Pedestrian Demand for Shared Use Path Design", funded by the Texas Department of Transportation (TxDOT) 0-7152. Principle Investigator: Ipek N. Sener (i-sener@tti.tamu.edu)

Bike | Ped Example : Austin – LBS Visits



Based on an ongoing project "Estimating Latent Bicyclist and Pedestrian Demand for Shared Use Path Design", funded by the Texas Department of Transportation (TxDOT) 0-7152. Principle Investigator: Ipek N. Sener (i-sener@tti.tamu.edu)

Bike | Ped Example : Austin –Model Estimates

Based on an ongoing project "Estimating Latent Bicyclist and Pedestrian Demand for Shared Use Path Design", funded by the Texas Department of Transportation (TxDOT) 0-7152. Principle Investigator: Ipek N. Sener (i-sener@tti.tamu.edu)

Trivia 3

- What is the unit of reporting across raw(sample) data by sources ?
  - LBS and Cell phone are devices (persons), where GPS data is primarily vehicle units

- What is the difference between raw, transformed and modeled
  - Majority of raw data and its attributes are observations,
  - Transformed data is either scaled to population/ vehicles with contextual attributes, and
  - Modeled is representative of regional travel as in travel demand models with contextual attributes

- What are broader categories of biases can one encounter with passive data
  - Spatial bias: Occurs with lack of coverage of an area or due to location estimation method/source
  - Temporal bias: Occurs as its dependent on the time of use on application (not activity reflection)
  - Demographic bias : Source data dependent

Applications : Part 2

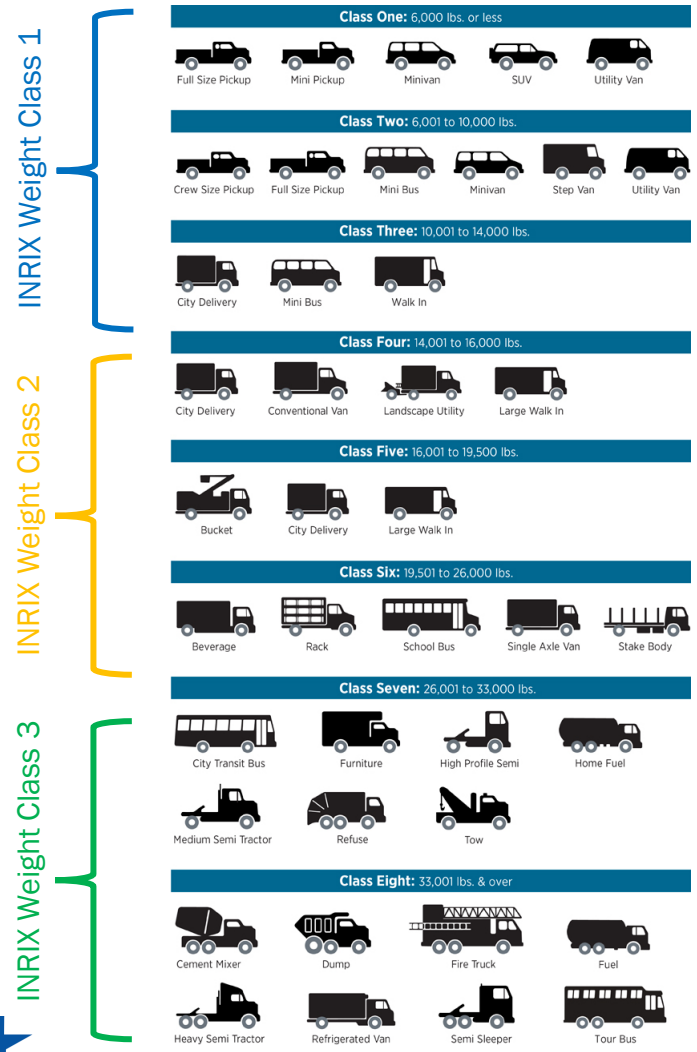
Regional Travel Demand Model Applications

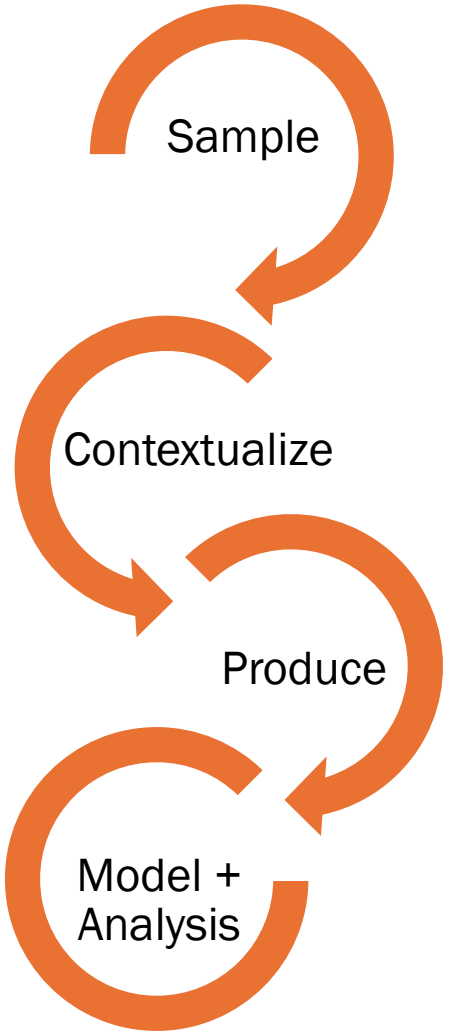
- Investigating:
  - Complement traditional commercial vehicle (cv) survey
  - Low sample size, primarily light vehicle representation
  - Need to better understand regional (cv) impact in models

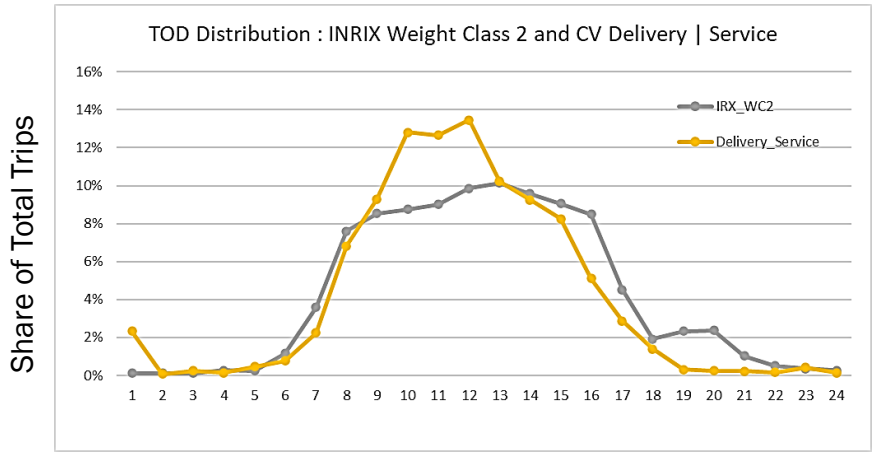
- Sources
  - INRIX sample trips and trajectories

- How to use/inferences to make
  - Develop sample daily trip tables by weight class
  - Synthesize with land use and regional VMT data to estimate CV travel
  - Use as input for regional travel demand models

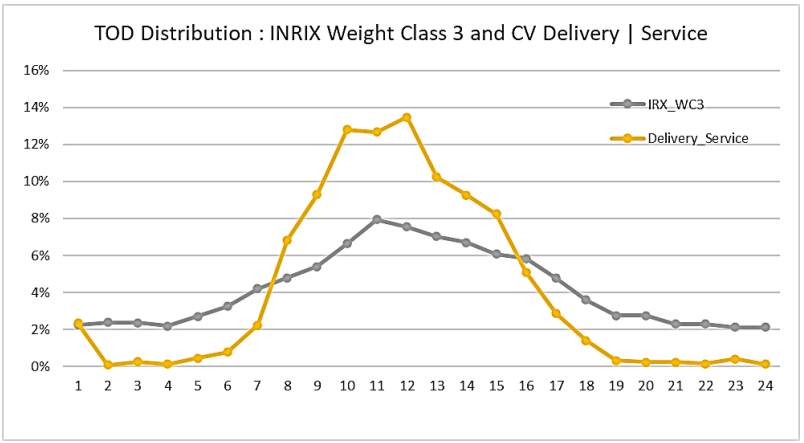
Trips Samples by Weight Class 1





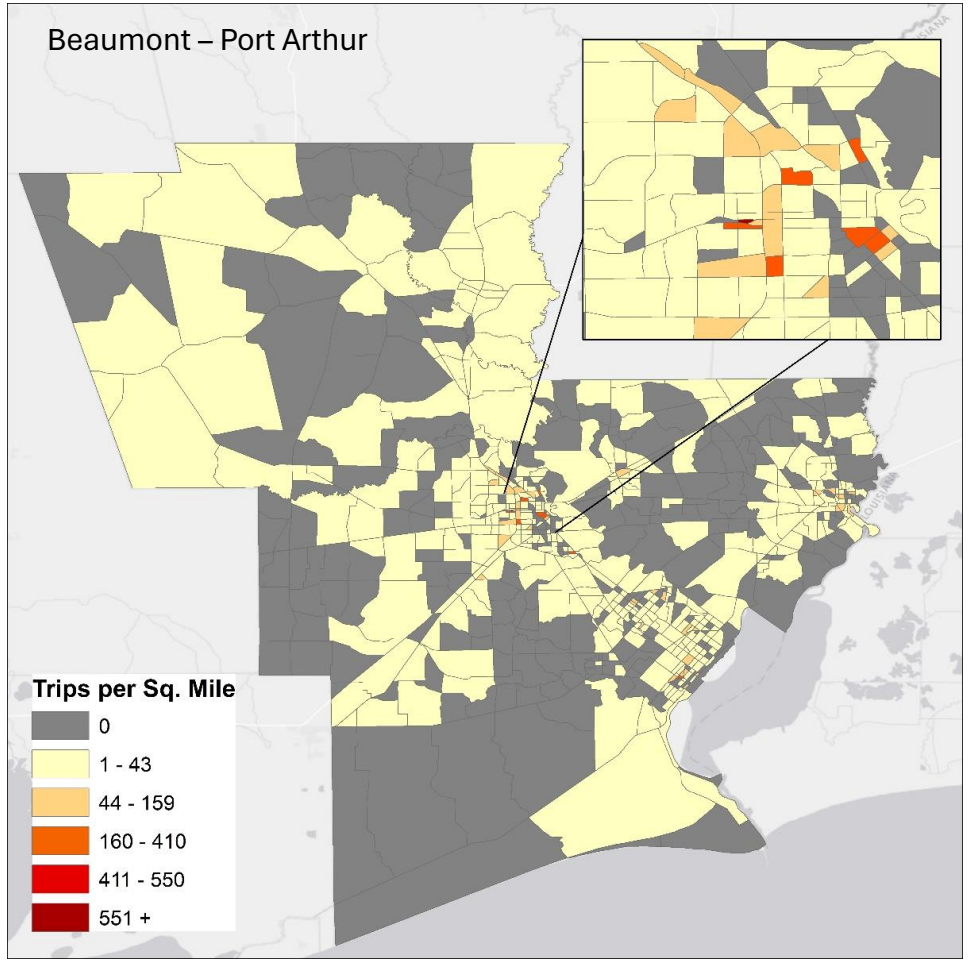


INRIX Weight Class 2 (14,000 – 26,000 lbs.) : Delivery | Services

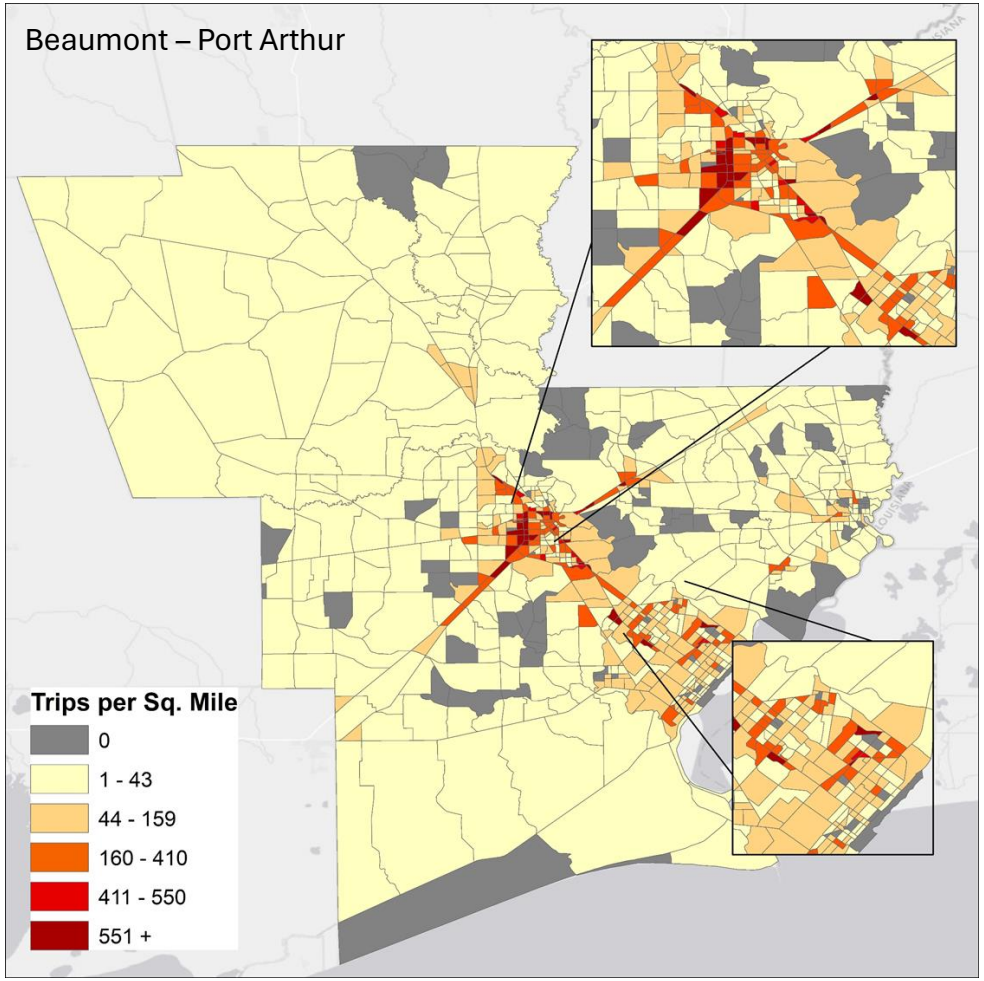


INRIX Weight Class 3 (> 26,000 lbs.) : For Hire Truck | Fleet

Survey Vs. Passive Data Daily Commercial Trips



CV Traditional Survey Data



Passive Data

Regional Quick Response Model (RQRM)

- Objective:
  - Standalone web-based passive data based RQRM Model
  - Mimics traditional travel demand model (auto only)
  - Application to low-growth areas, sub-model or new model area (no survey data)

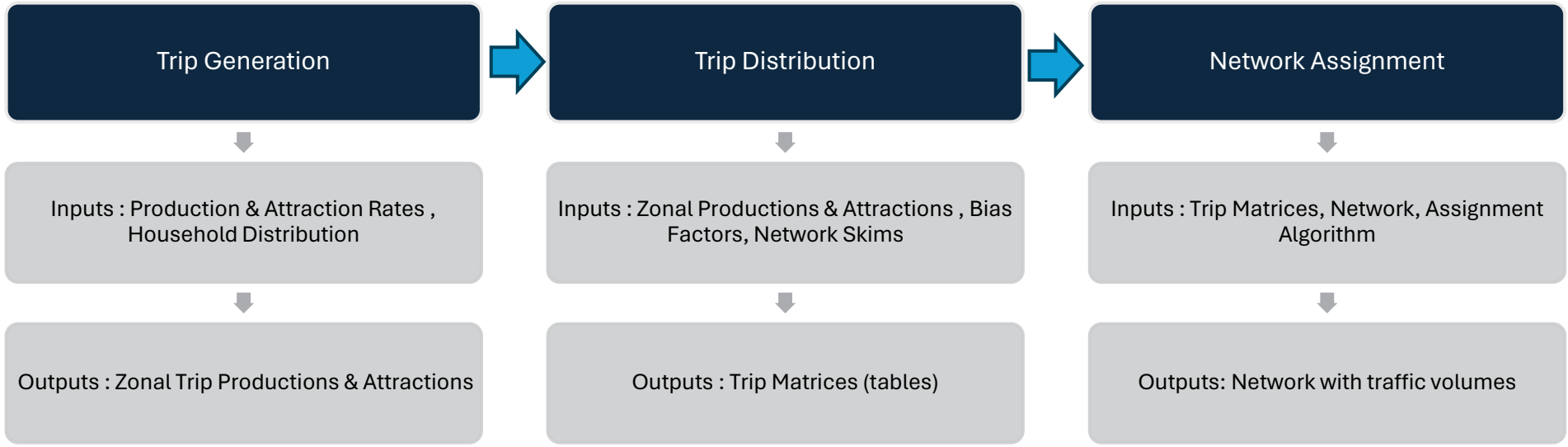
- Sources
  - LBS (Smartphone Applications) and Vehicle GPS Data

- How to use/inferences to make
  - Understand one or more growth scenarios
  - Assess the effect of roadway widening projects or addition of new roadway

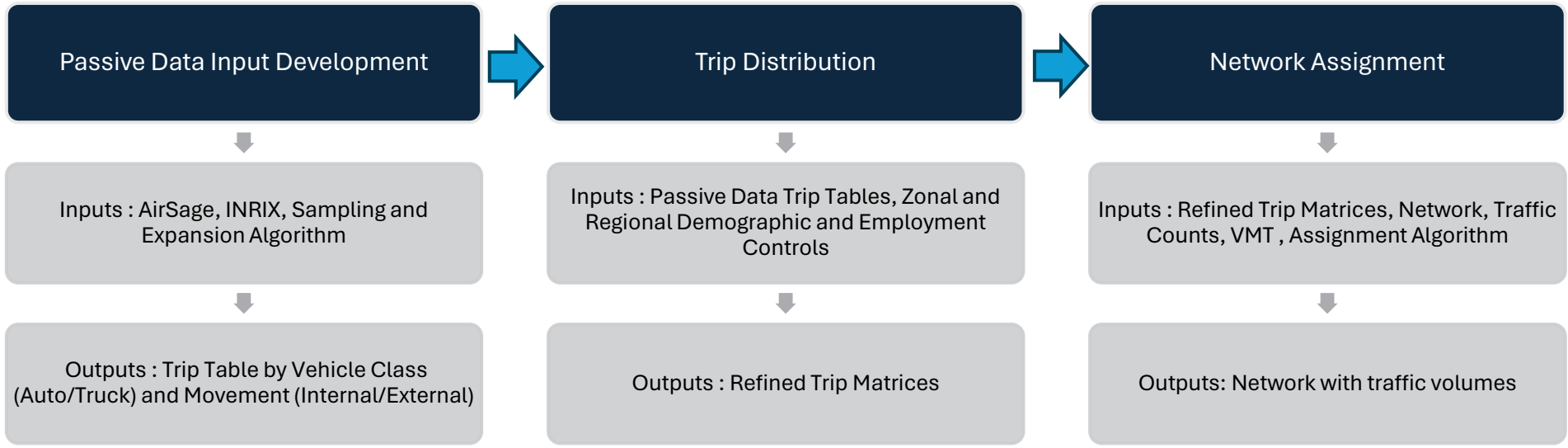
Ongoing work as part of TxDOT - FY 21 Inter Agency Contract (IAC) implementing a Quick Response Model using passive probe data

TexPACK TDM Vs. QRM

TexPACK Trip Based



QRM



RQRM – Travel Market Segments

Example – Regional Trip Table

1	2	3	4	....	323	324	340
2							
3							
..							
..							
323							
324							
...							
340							

Trip Movement Type

- Internal – Internal
- External - Internal (EXLO)
- External – External (THRU)

Trip Table Travel Market Segment Reference

Blue bins represent within region trips, includes

- Noncommercial resident auto trips for work & other
- Noncommercial visitor auto trips for business/recreation
- Local commercial (delivery/service travel)

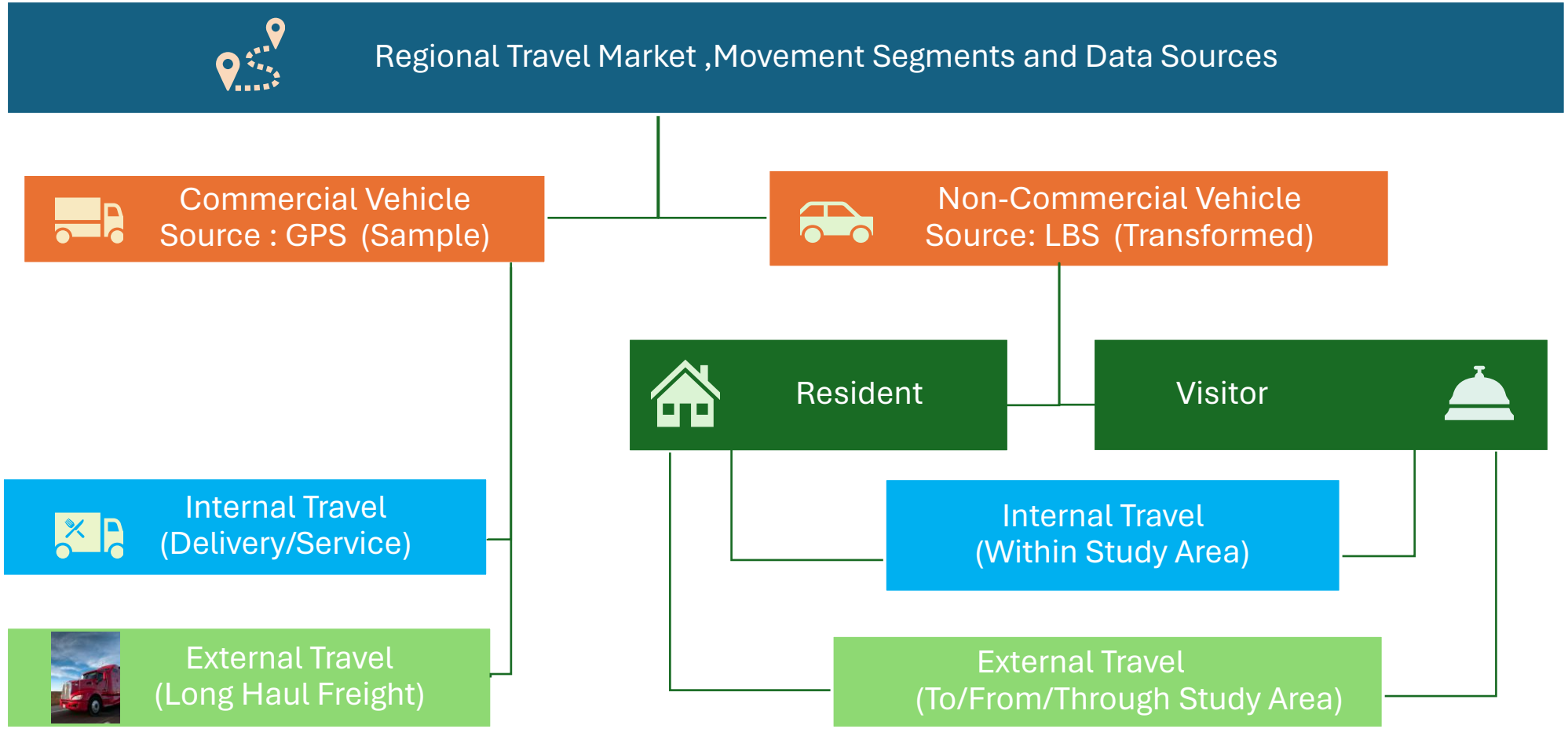
Purple bins includes:

- Long haul commercial truck trips entering/exit region
- Noncommercial visitor auto trips entering/exit region
- Noncommercial resident auto trips enter/exit region

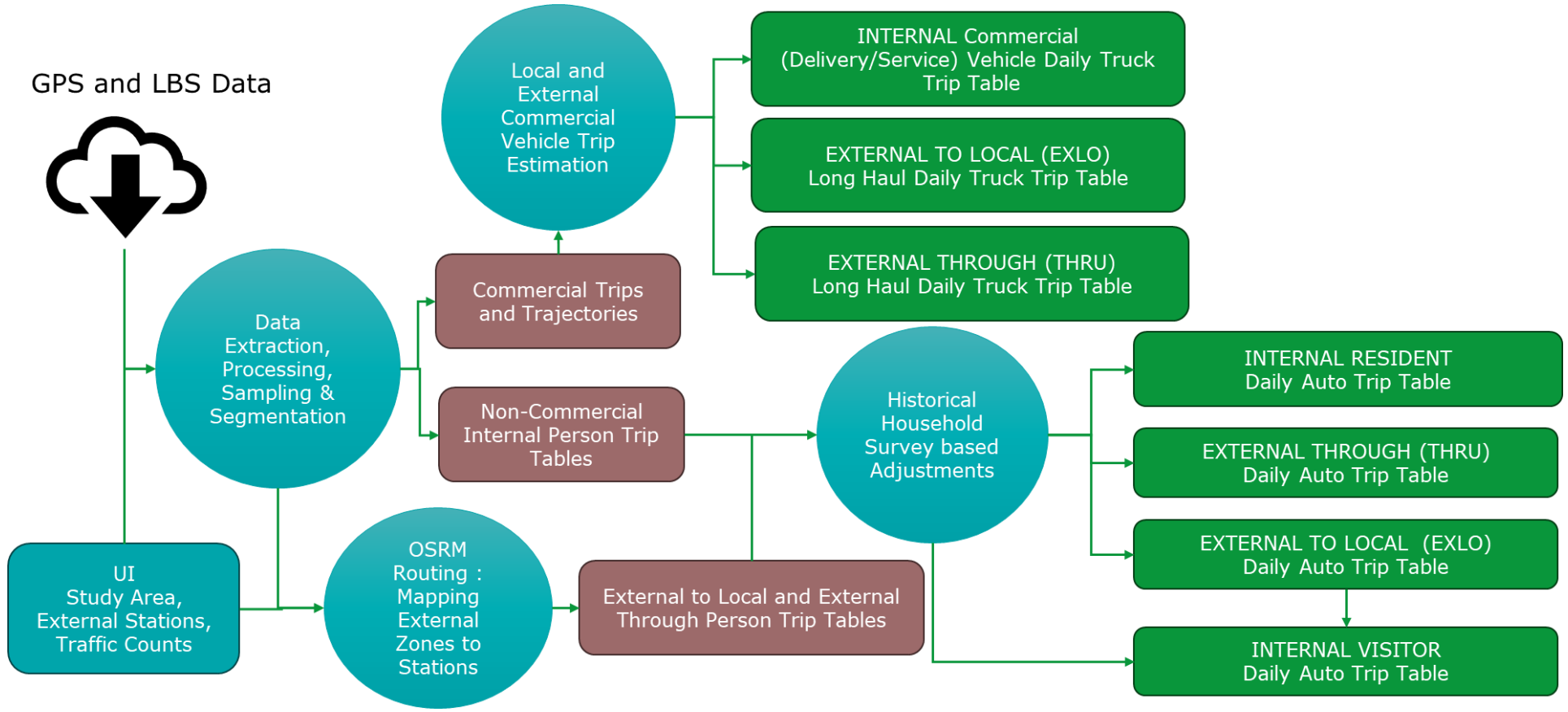
Lavender bins includes:

- Long haul Commercial truck trips passing through the region
- Non-Commercial auto trips passing through the region

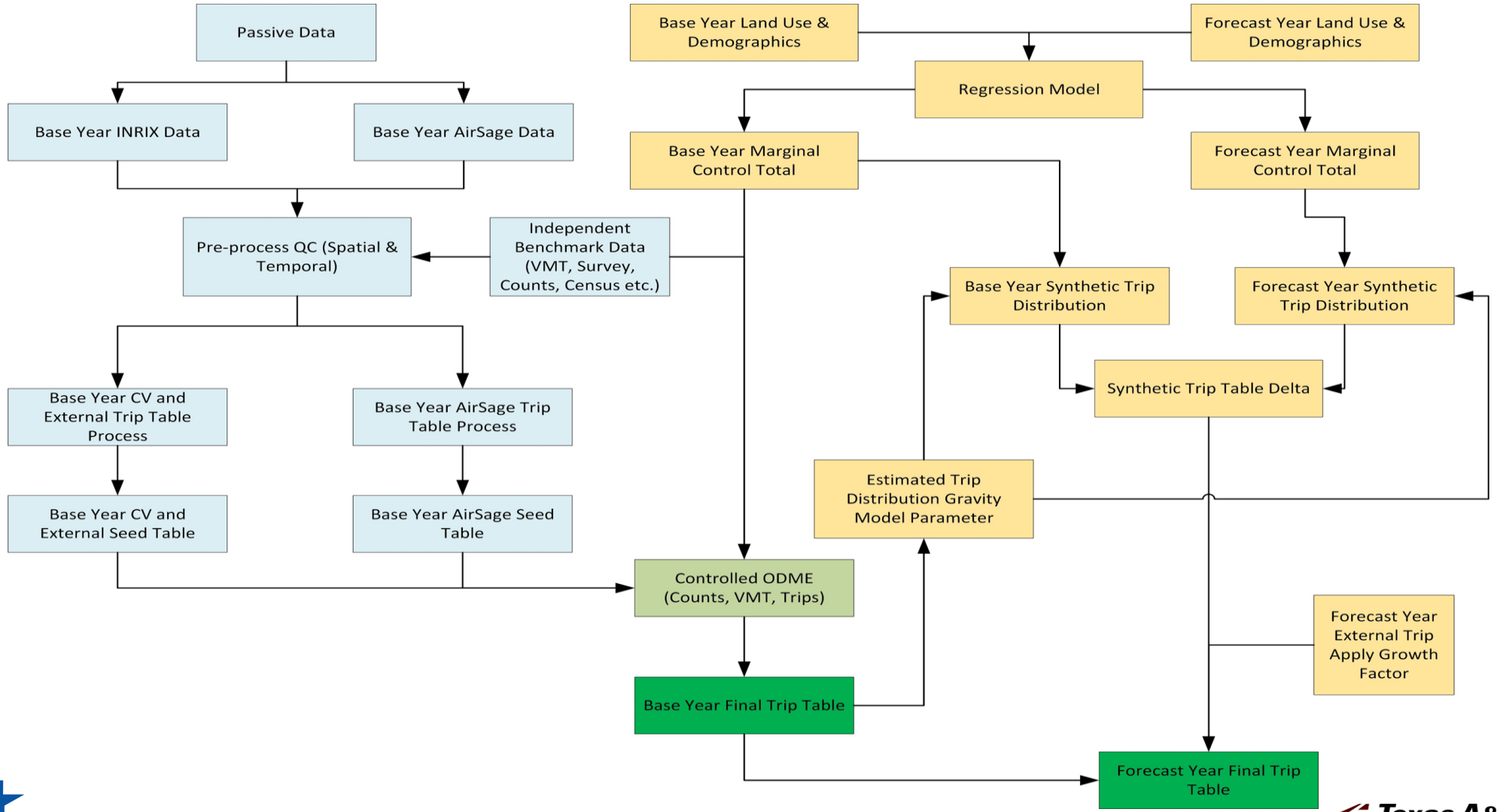
RQRM – Travel Market : Data Sources



RQRM – Travel Market : Data Processing

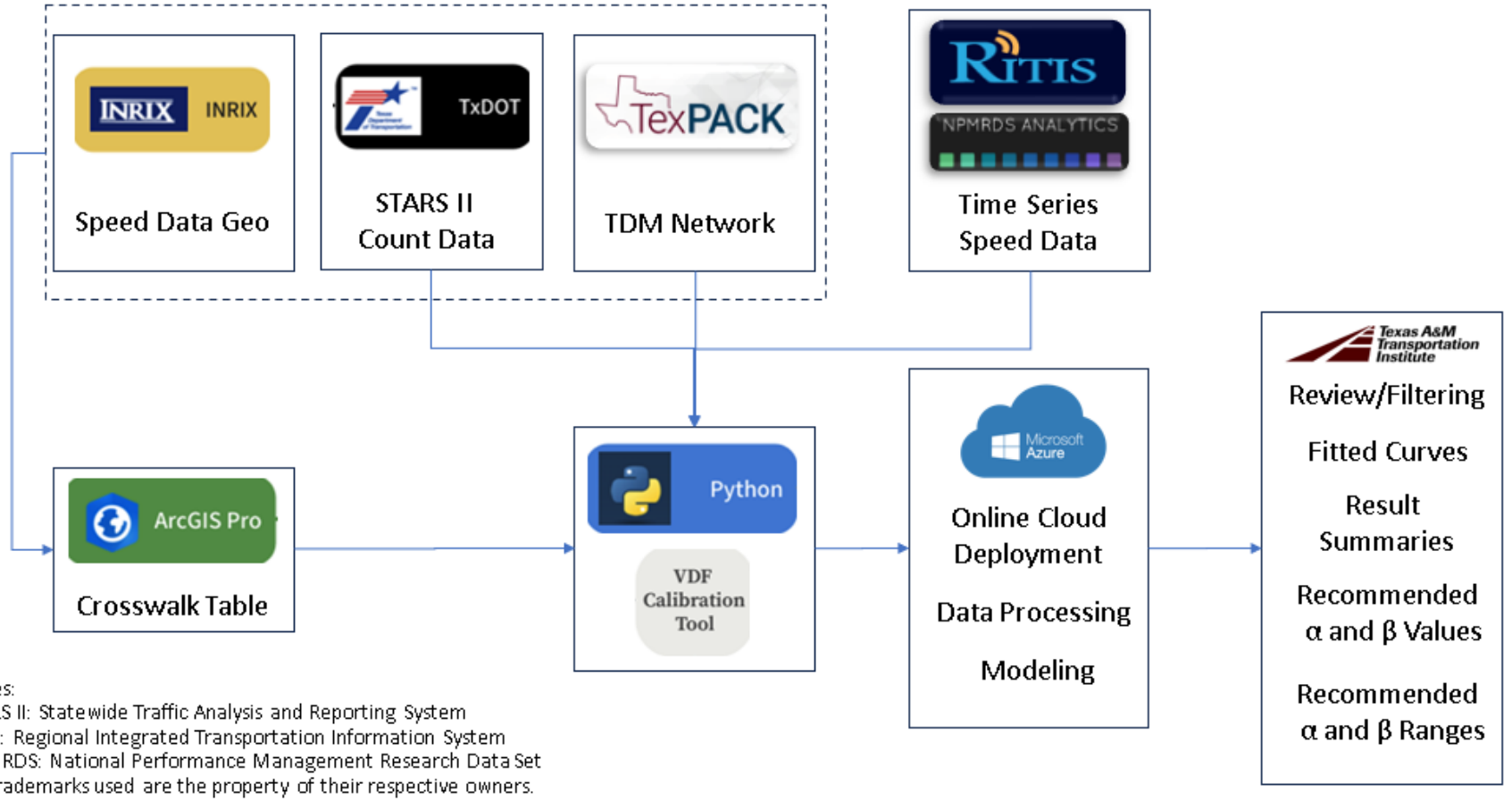


RQRM – Data and Model Framework



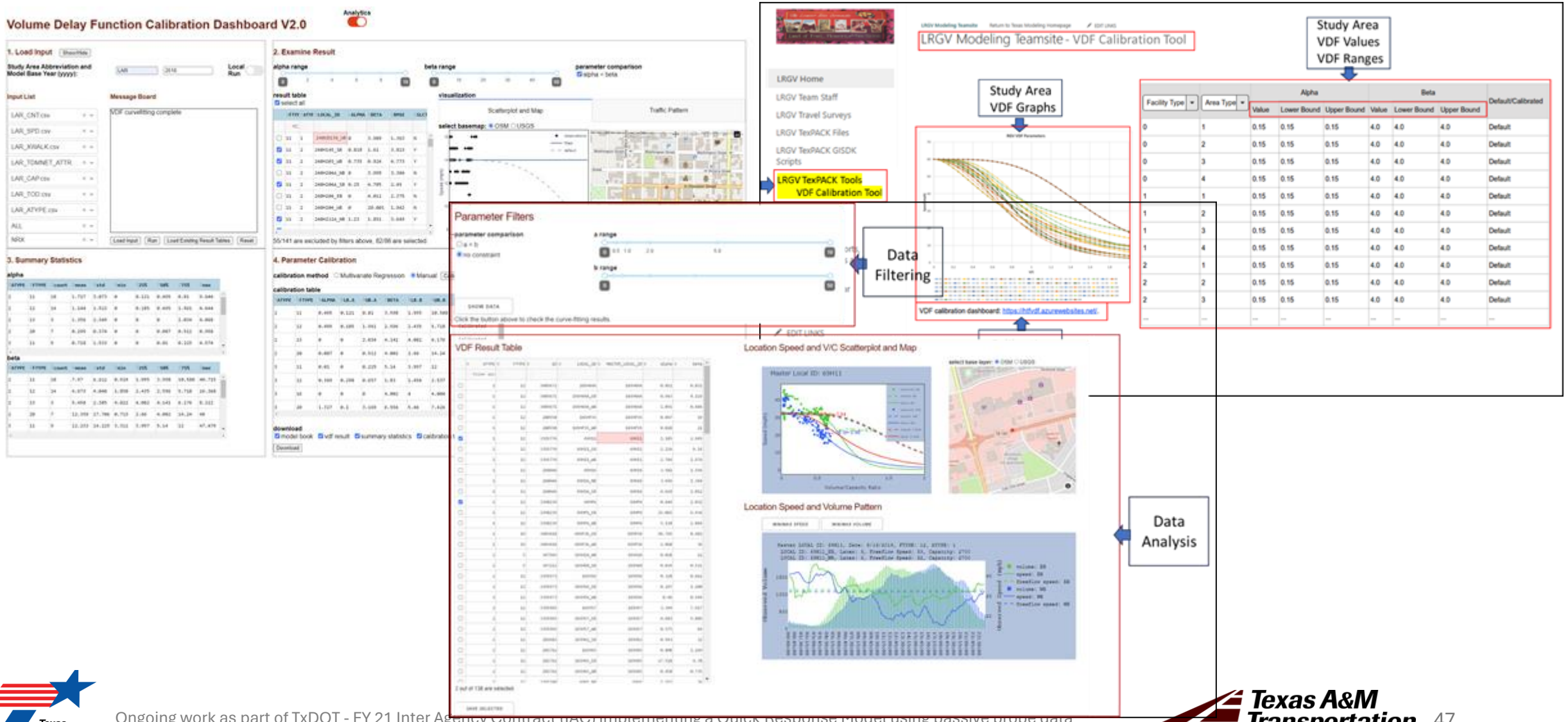
INRIX – VDF Applications 1

BPR VDF Calibration Process ($T_f = To \times (1+\alpha \times [v/c]^{\beta})$)

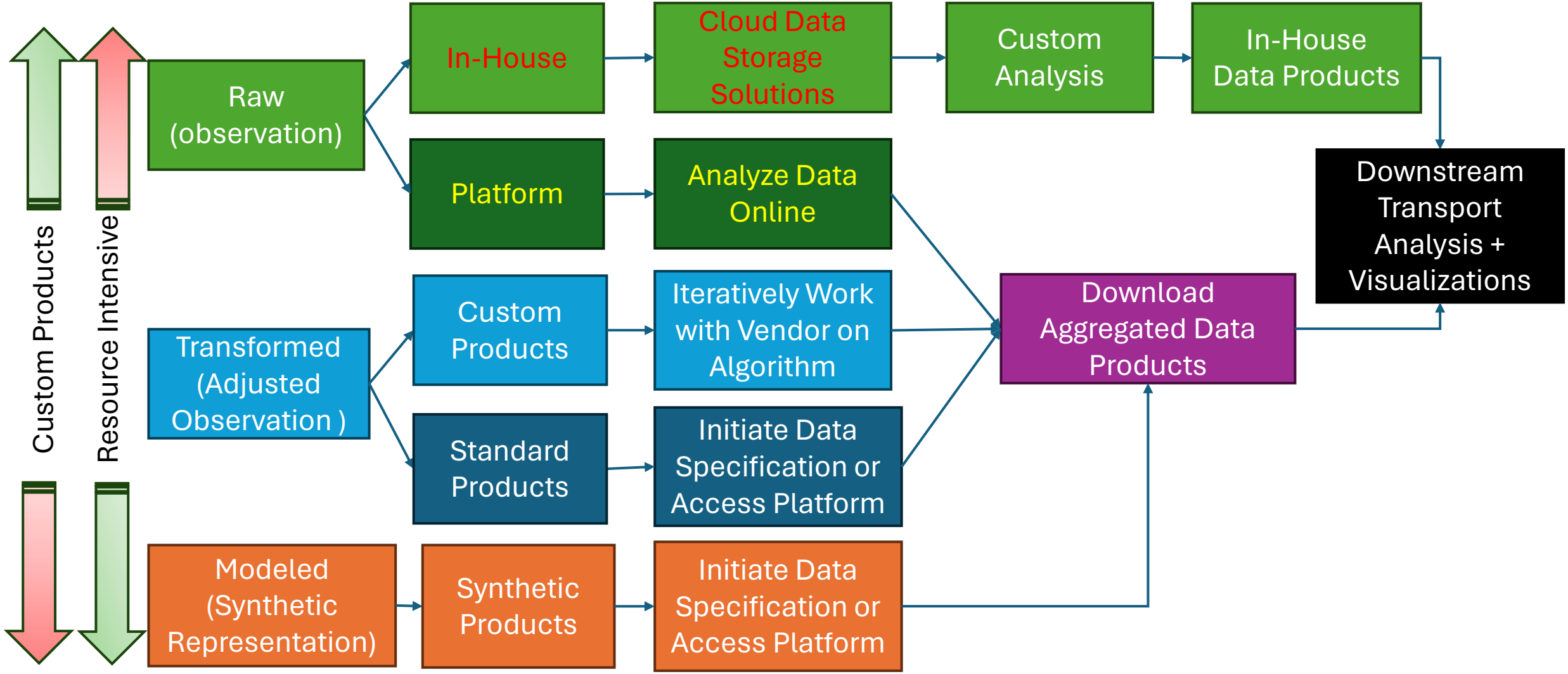


Notes:
STARS II: Statewide Traffic Analysis and Reporting System
RITIS: Regional Integrated Transportation Information System
NPMRDS: National Performance Management Research Data Set

INRIX – VDF Applications 2

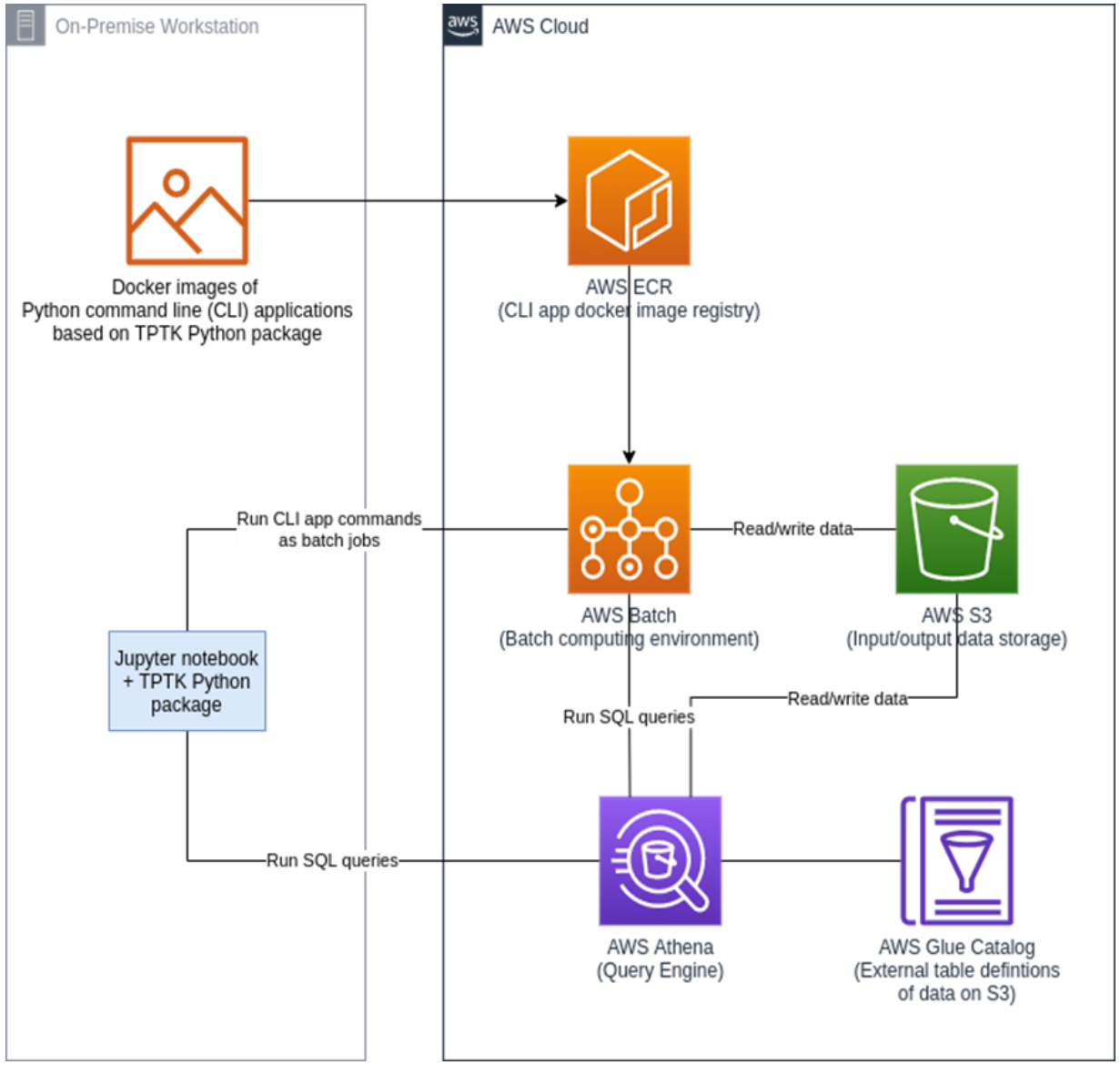


Data decision making pathways



In House Application Development

- Data Lake Management
  - Traditional surveys and data
  - Big Data (Points, Trips, Trajectories)
  - Schemas & Documentation
- Data Pipeline Development
  - Scale methods, notebooks to production
  - QA /QC process Integration
- Transportation Science Platform
  - Access to non-traditional users (SQL)
  - Intermediate products for analysis
  - Platform for quick visualization and reports
- Documentation | Code Repositories
  - GitHub code version control and collaboration
  - Confluence : documentation and analysis review



Data Governance & Management

Skills

- Python | R | SQL | Spark, Cloud Technologies (AWS, Azure, GCP)
- Reporting, Visualization , Quantitative, Service Oriented

Staff

- Data Engineer, Cloud Administrator, Program Manager
- Web App Developer, Data Scientist, Data Analysts , Subject Matter Experts
- Transportation Planners, Traffic Engineers, and Analysts

Governance

- Ingestion, Transformation, Storage and Archiving Mechanism and Policies
- Simplified Cloud Access and Audit, and Retrieval Medium by User Type
- QA/QC, Resource Monitoring, Analysis Tools for End Users

Acknowledgements

Thank you – Questions ?

Source : https://datasciencedojo.com/blog/data-science-memes/

Texas A&M Transportation Institute

Vijay Sivaraman
(813) 389-8224
V-sivaraman@tti.tamu.edu