# Safety Applications of Crowd-Sourced Traffic Data

Erika Sudderth, PhD

2019-08-13

# Safety Data Initiative (SDI)

- Launched in 2018
- Surface transportation focused
- Intended to build upon and enhance current safety efforts related to data, analysis, and policymaking

- Cross-cutting, collaborative effort:
  - Office of the Secretary of Transportation (OST)
    - Policy Office
    - Office of the Chief Information Officer
    - Bureau of Transportation Statistics
  - Federal Highway Administration (FHWA)
  - National Highway Traffic Safety Administration (NHTSA)
  - Other surface operating administrations (OAs)

# Focus Areas

SDi SAFETY DATA INITIATIVE

**Integrate** existing DOT data and new "big data" sources

Use advanced data analytics to provide **predictive insights** into safety risks

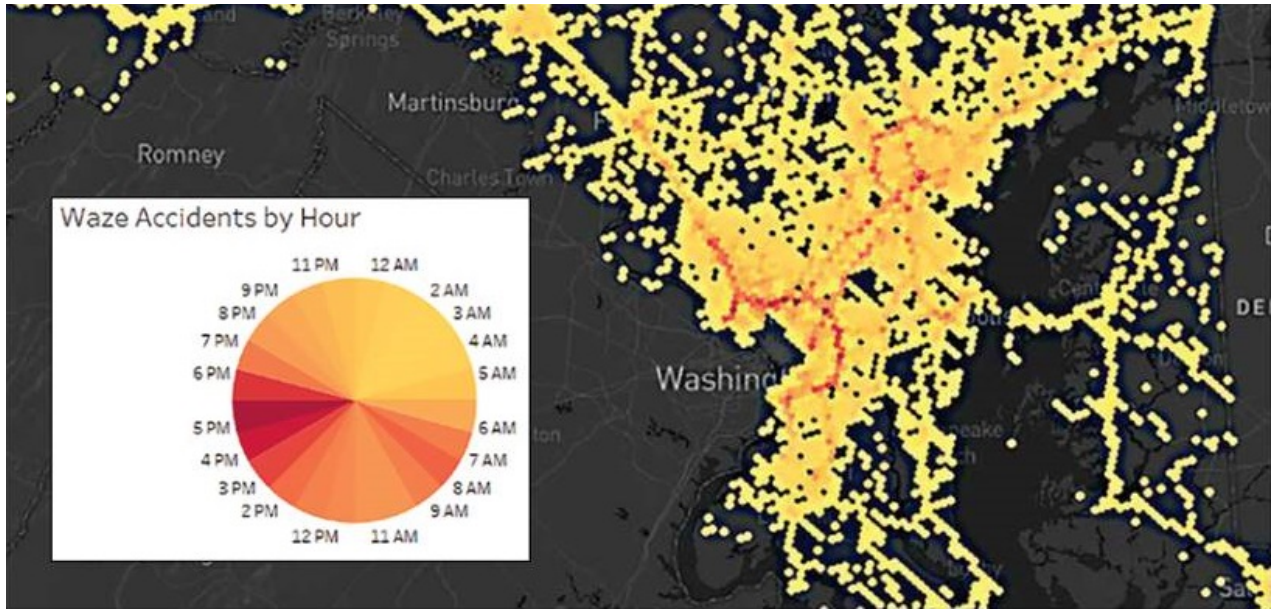
Create **data visualizations** to help policymakers arrive at solutions

# Safety Data Initiative – Waze Pilot Project

- DOT became a Waze Connected Citizens partner (data from April, 2017)
- Developed the Secure Data Commons: AWS cloud platform to process, curate, and analyze big data within DOT (Waze and other transportation data)
- NHTSA Electronic Data Transfer (EDT): daily updated crash reports for 7+ states
- Waze pilot: Integrate transportation data to develop rapid crash indictors
  - Phase 1: State-wide indicators of police-reportable traffic crashes
  - Phase 2: State and local applications of Waze analysis pipeline
    - Tennessee: Crash propensity model to target safety risk with highway patrols
    - Bellevue: Crash risk model to inform Vision Zero action plan

# Phase 1: State-Wide Crash Models using Waze data

- Assessed spatial and temporal relationships between Waze events and police-reported traffic crashes
- Integrated statewide Waze, traffic volume, job, and weather data for MD, VA, CT, and UT
- Applied machine learning to reliably estimate hourly police reportable crashes in four states
- Created interactive Tableau dashboards: when and where are model estimates accurate?




*Our Waze data integration, modeling, and visualization pipeline can support nationwide studies or state and local applications*

# Statistical Approach: Supervised Classification

## Random Forests

- Machine learning approach which minimizes overfitting
- Trained models on 70% of data using EDT reports as our labeled “ground-truth”
- Tested model performance using 30% of data to compare estimated EDT crashes with observed EDT crashes
- Rigorously trained and tested data feature combinations (50+ models)
- Best crash estimation models minimize false positives and false negatives

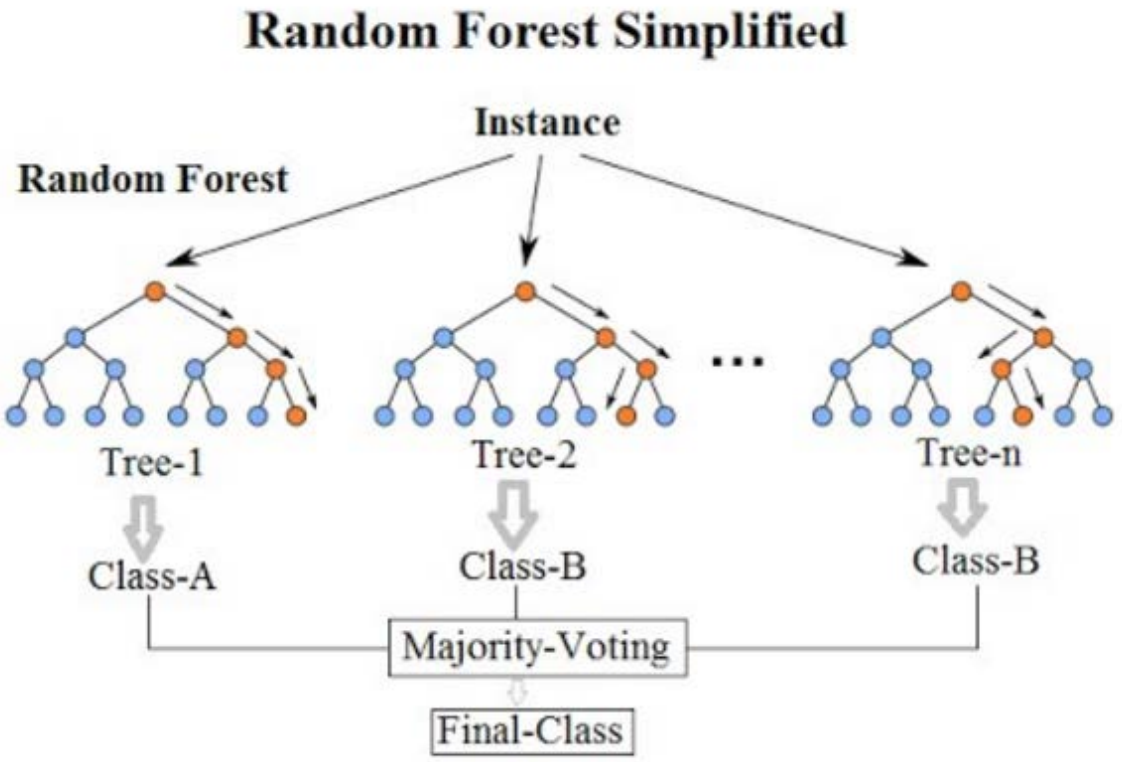
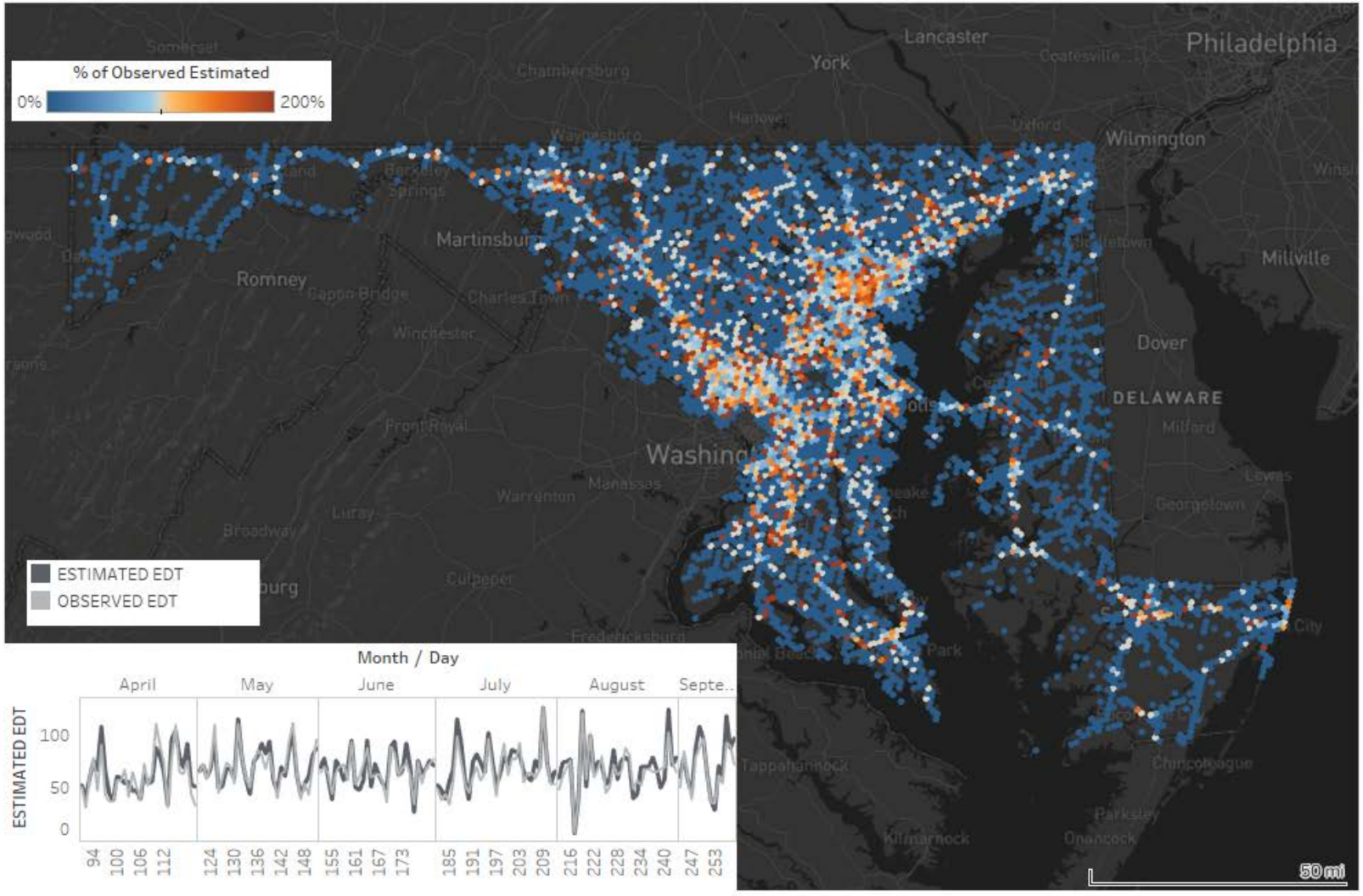



Image credit: https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

# Model Performance (April-Sept 2017 in MD)

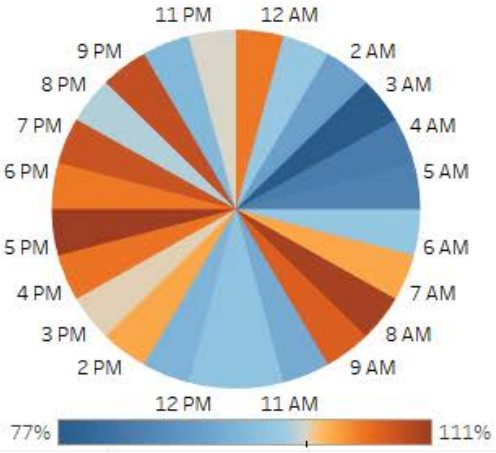
Model estimates highly accurate overall; miss some precise patterns




## Crashes by Day

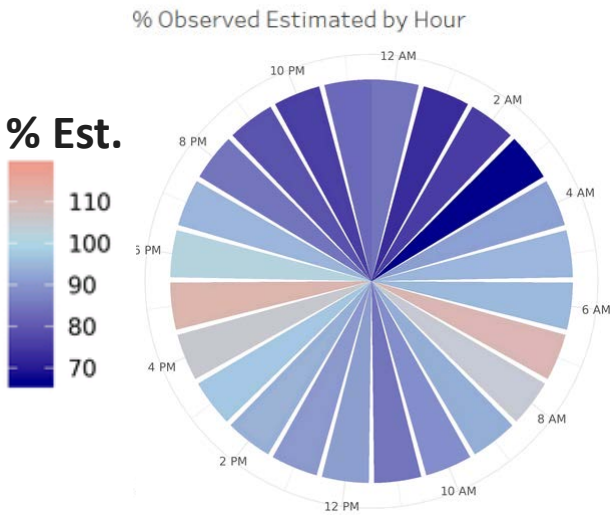
| Day Of Week | ESTIMATED EDT | OBSERVED EDT | PRCT OBSERVED |
|---|---|---|---|
| Monday | 1,089 | 1,099 | 99.09% |
| Tuesday | 1,623 | 1,602 | 101.31% |
| Wednesday | 1,788 | 1,709 | 104.62% |
| Thursday | 1,768 | 1,694 | 104.37% |
| Friday | 1,922 | 1,840 | 104.46% |
| Saturday | 1,945 | 1,869 | 104.07% |
| Sunday | 1,390 | 1,413 | 98.37% |

## % Observed Estimated by Hour

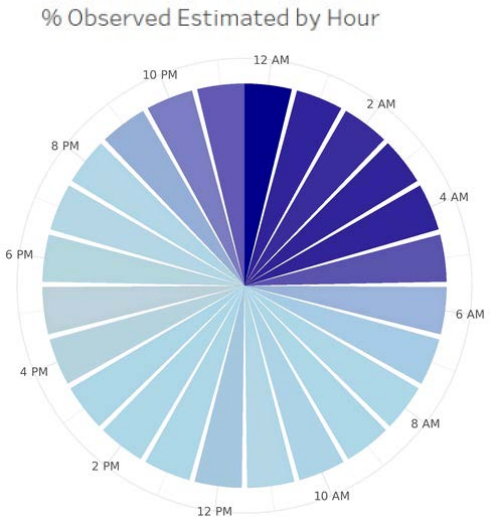


# Models perform well across multiple states
# Variation by hour and location related to Waze coverage

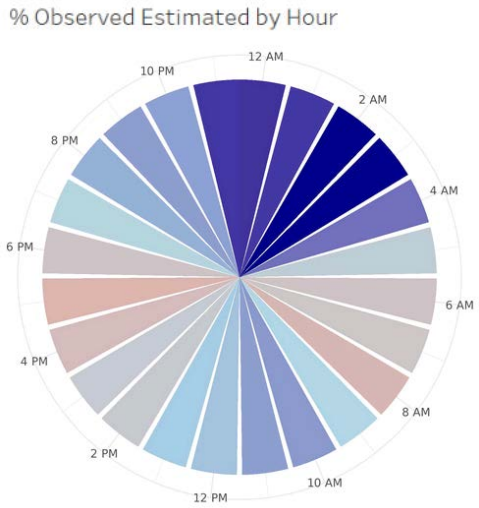
## Maryland



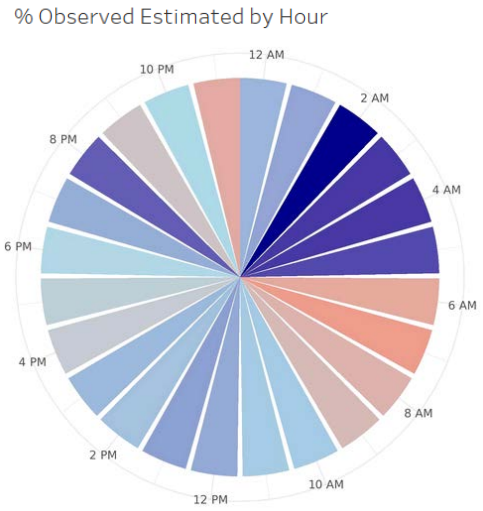

## Connecticut

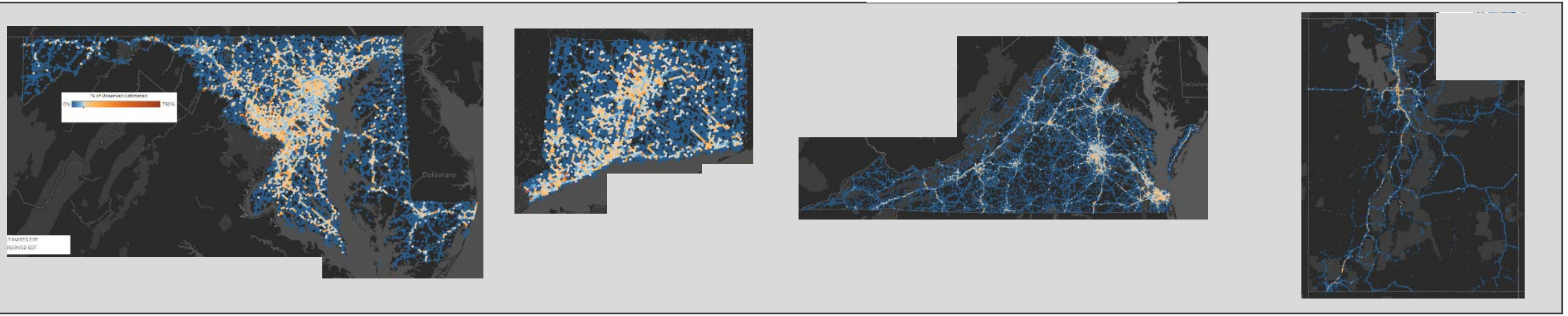



## Virginia




## Utah

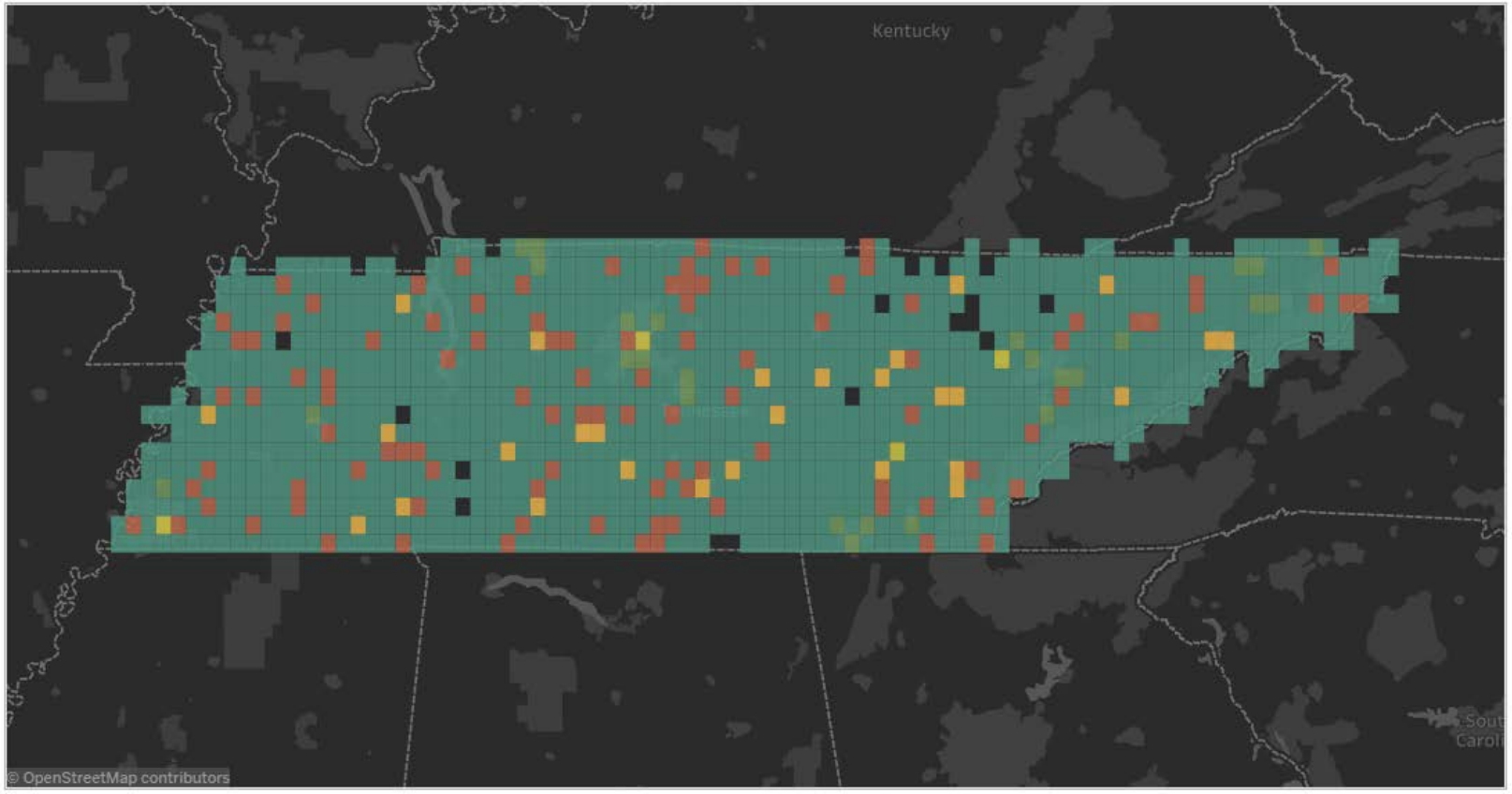


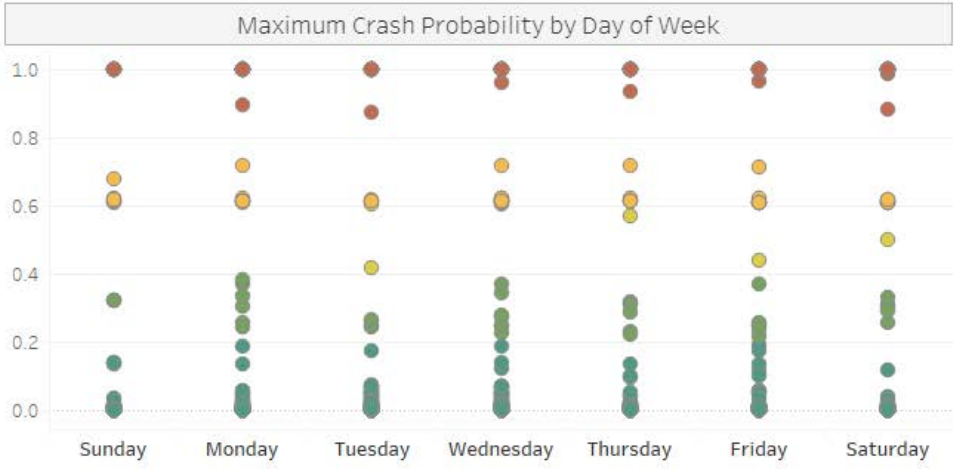



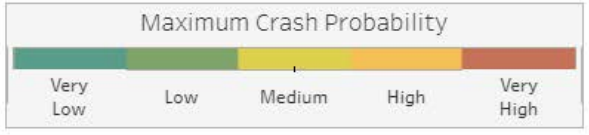
# Phase II: Tennessee Case Study

- Highway Patrol uses machine learning to predict crash propensity and target patrols
- Integrating Waze data with existing grid models improves estimates
  - Spatial resolution: 42 to 1 sq mile
  - Temporal resolution: 4 hrs to 1 hr
- Results will help HP better target high crash risk locations and times

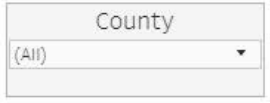
# Maximum Crash Probability - *Model 05 , May 6, 2019 - May 13, 2019 in Tennessee*

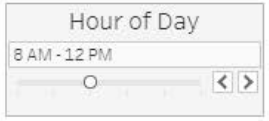


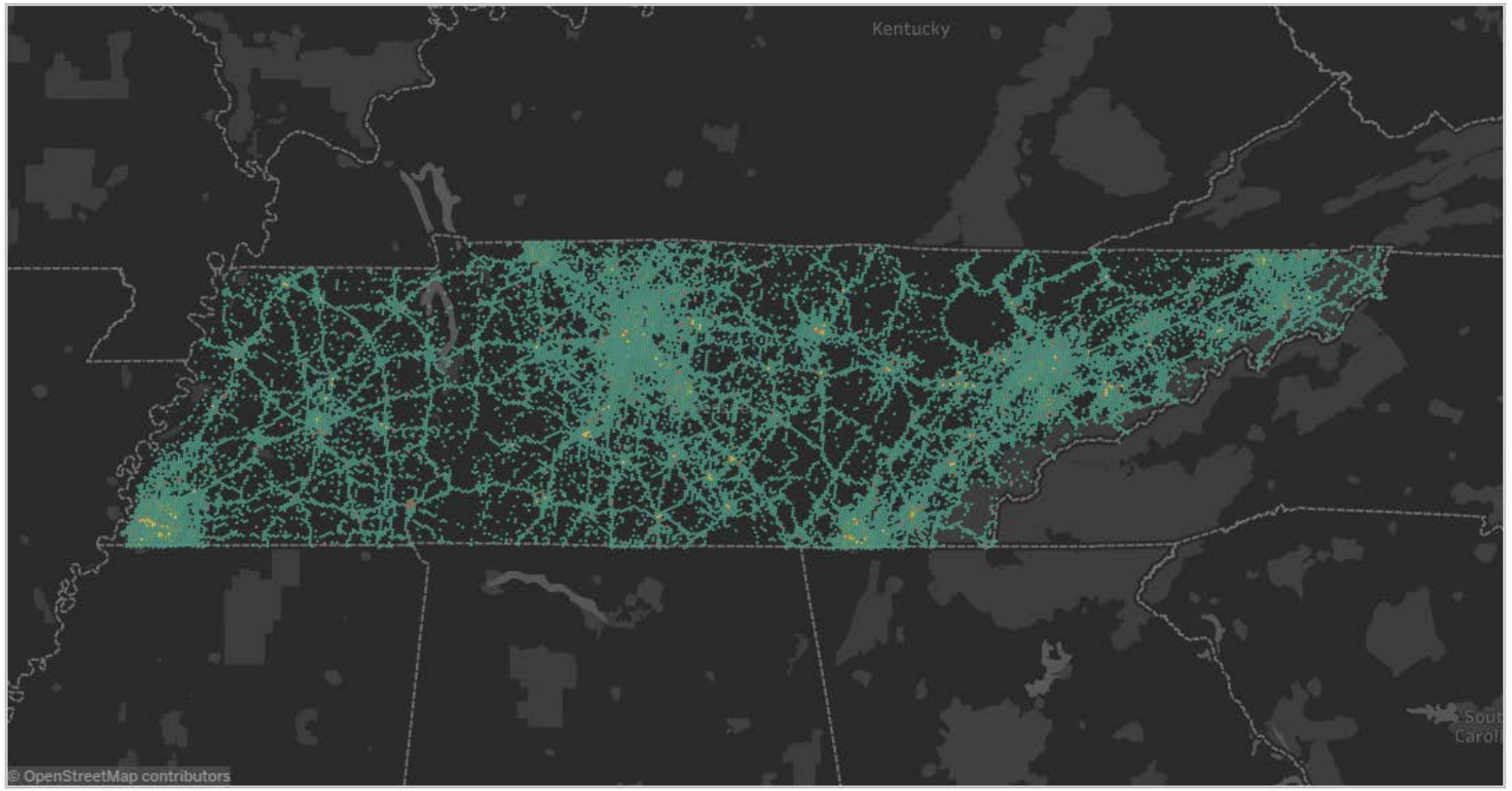



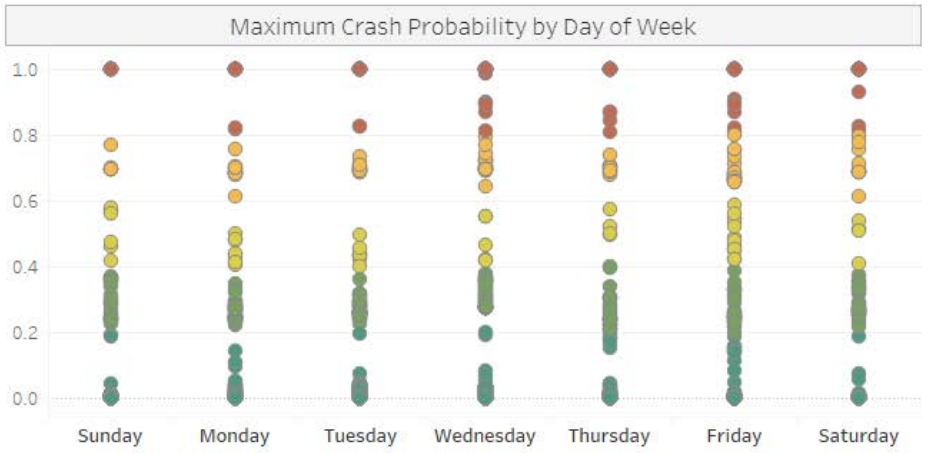


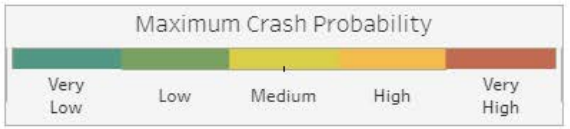


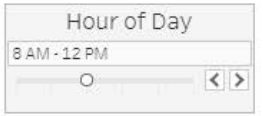
# Maximum Crash Probability - *Model 05 , May 6, 2019 - May 13, 2019 in Tennessee*






# Phase II: Bellevue Case Study

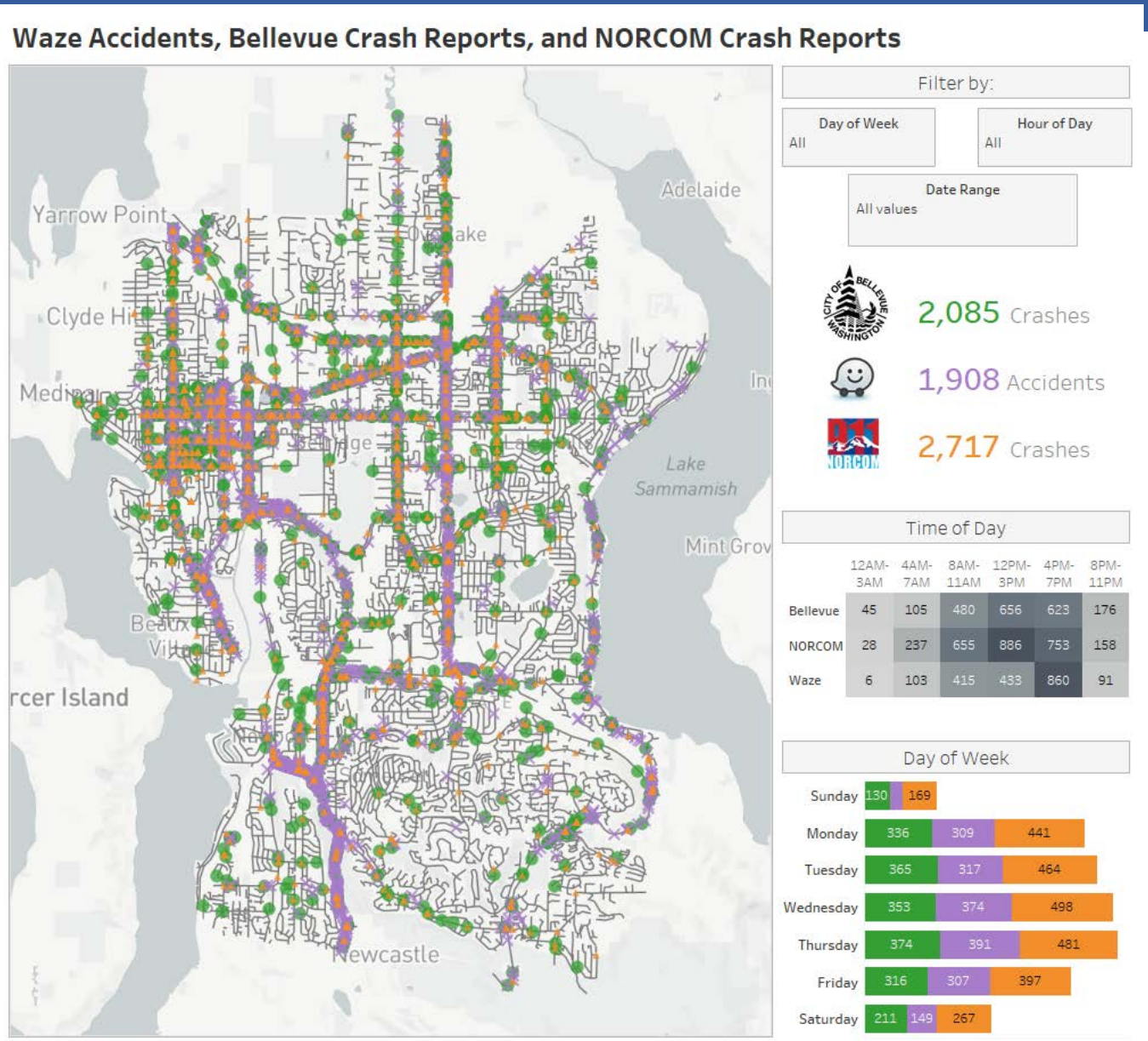
**Crowdsourced traffic incident data to improve traffic safety management**

**Approach:**

- Integrate data sources and create dashboards
- Develop crash estimation models: conditions, times, locations with high propensity
- Transfer methods to Bellevue (CC partner)

**Outcomes:**

- First integrated view of 3 traffic crash datasets highlights unique contributions of each by time and location
- Segment-level crash models will guide transportation safety investment decisions



# Estimated Crash Counts

City of Bellevue, WA

Time of Day

(All)

(All)
Weekday
Weekend

Functional Class

(All)

High Injury Network

(All)

Priority Neighborhoods

(All)

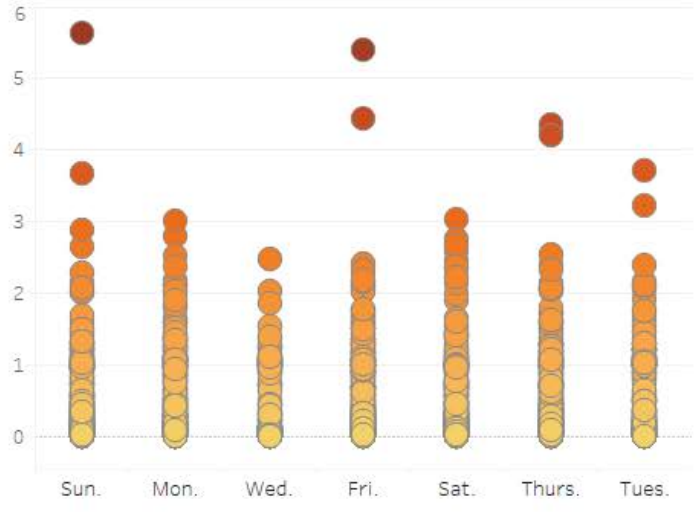
Estimated Crash Counts

Low — High


Street with Highest Estimated Crash Count

| Functional Class | Street Name | |
|---|---|---|
| Principal Arterial | 148TH AVE NE | 77 |
| Minor Arterial | 156TH AVE NE | 45 |
| Collector | MAIN ST | 13 |
| Local Access | FOREST DR SE | 5 |

Estimated Crash Counts by Day of Week




## Dashboard summary

- Estimated crash counts for modeled segments
- Tool-tip shows observed and estimated number of crashes
- Can filter by Time of Day, Functional Class, HIN (on/off), priority neighborhoods
- Summary table shows street with highest *estimated* crash count

U.S. Department of Transportation Volpe Center

# Weighted-Estimated Crash Counts

City of Bellevue, WA

Time of Day
(All)

(All)
Weekday
Weekend

Functional Class
(All)

High Injury Network
(All)

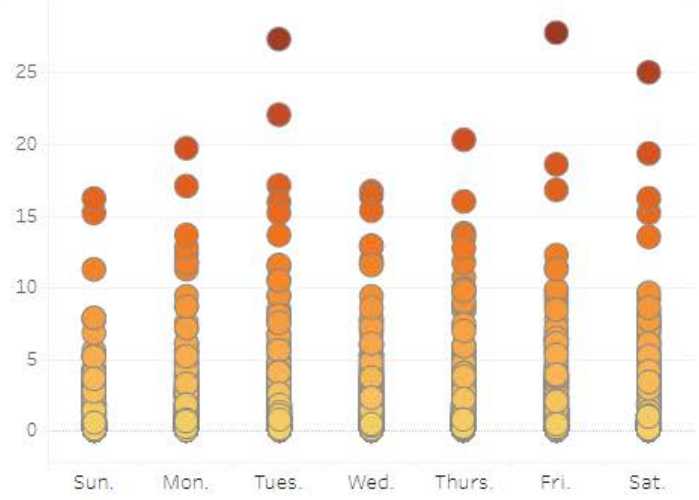
Priority Neighborhoods
(All)

Estimated Crash Counts
Low High


Street with Highest Estimated Crash Count

| Functional Class | Street Name | |
|---|---|---|
| Principal Arterial | NE 8TH ST | 192 |
| Minor Arterial | FACTORIA BLVD SE | 106 |
| Collector | 156TH AVE NE | 66 |
| Local Access | FOREST DR SE | 35 |

Estimated Crash Counts by Day of Week




## Dashboard summary

- *Weighted* Estimated crash counts for modeled segments
- Weights: 25 KSI, 10 injury, 1 PDO
- Different streets have highest counts

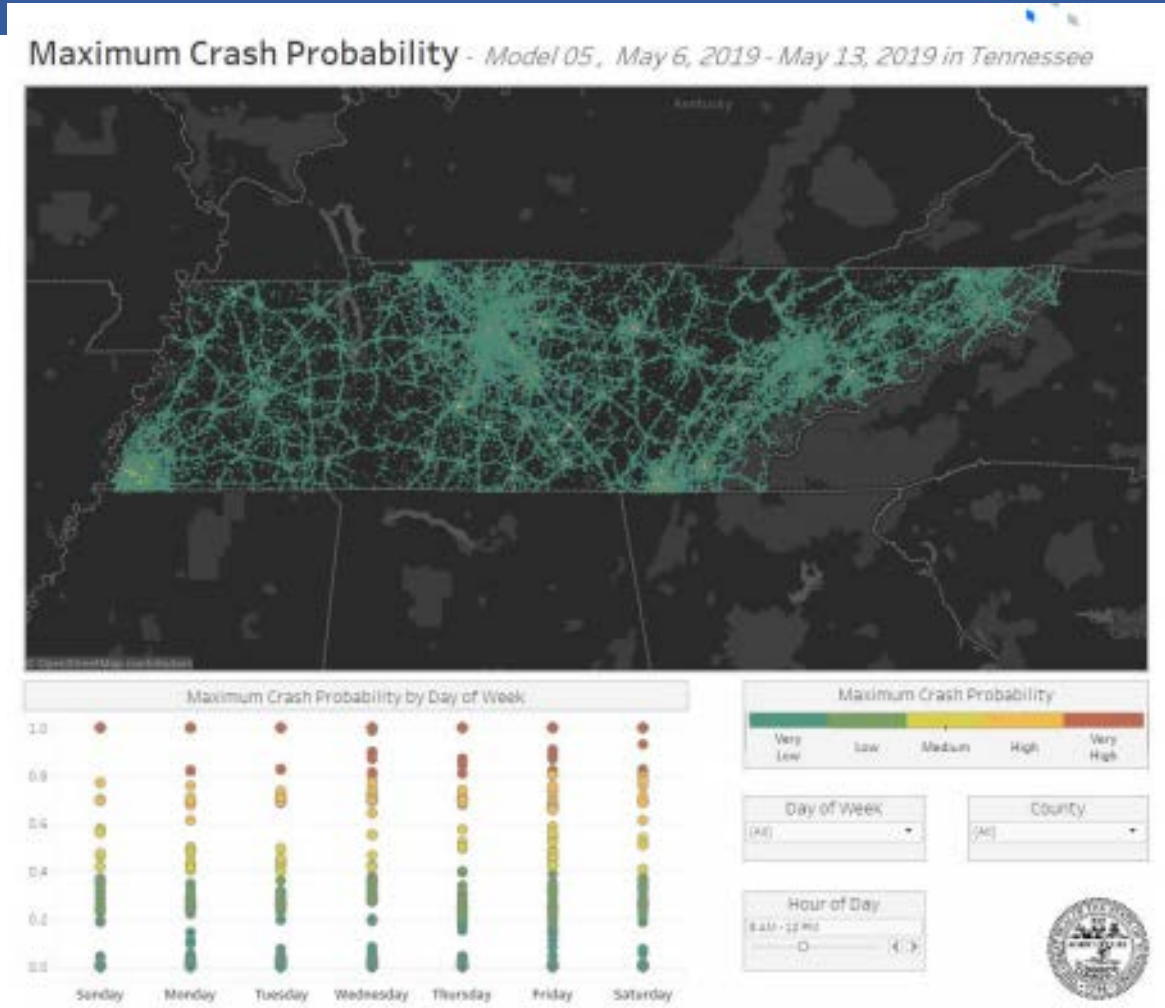
U.S. Department of Transportation Volpe Center

# Insights – What Have We Learned?

Waze data provide important contextual information to inform state and local safety applications

- Crash models using integrated Waze, traffic volume, job, and weather data give reliable estimates
- Tennessee Highway Patrol will more effectively target high-risk times and areas
- Crash propensity models will guide city-wide safety investment decisions

*Crowd-sourced traffic data can enhance other roadway data to illuminate safety risk patterns and inform decision making*



# Waze Pilot: Next Steps


- Transfer data integration, modeling, and visualization approaches to state and local case study partners (grid and segment models)
  - Tennessee: Deploy updated crash propensity models with Waze data at finer spatial and temporal resolution
  - Bellevue: Transfer analytical methods and dashboard development process.
- Explore safety applications with other state and local partners
- OST SDI Procurements: https://www.transportation.gov/content/safety-data-initiative

https://www.volpe.dot.gov/news/using-crowdsourced-data-estimate-crash-risk

https://www.wired.com/story/waze-data-help-predict-car-crashes-cut-response-time/