U.S. Department of Transportation
Safety Data Initiative
Request for Information

Webinar

November 15, 2018

Webinar Agenda

- **Introductory Remarks**

  Derek Kan, Under Secretary of Transportation for Policy

- **Overview of the Safety Data Initiative (SDI)**

  David Winter, Director, Office of Highway Policy Information, Federal Highway Administration

- **Overview of SDI Request for Information (RFI)**

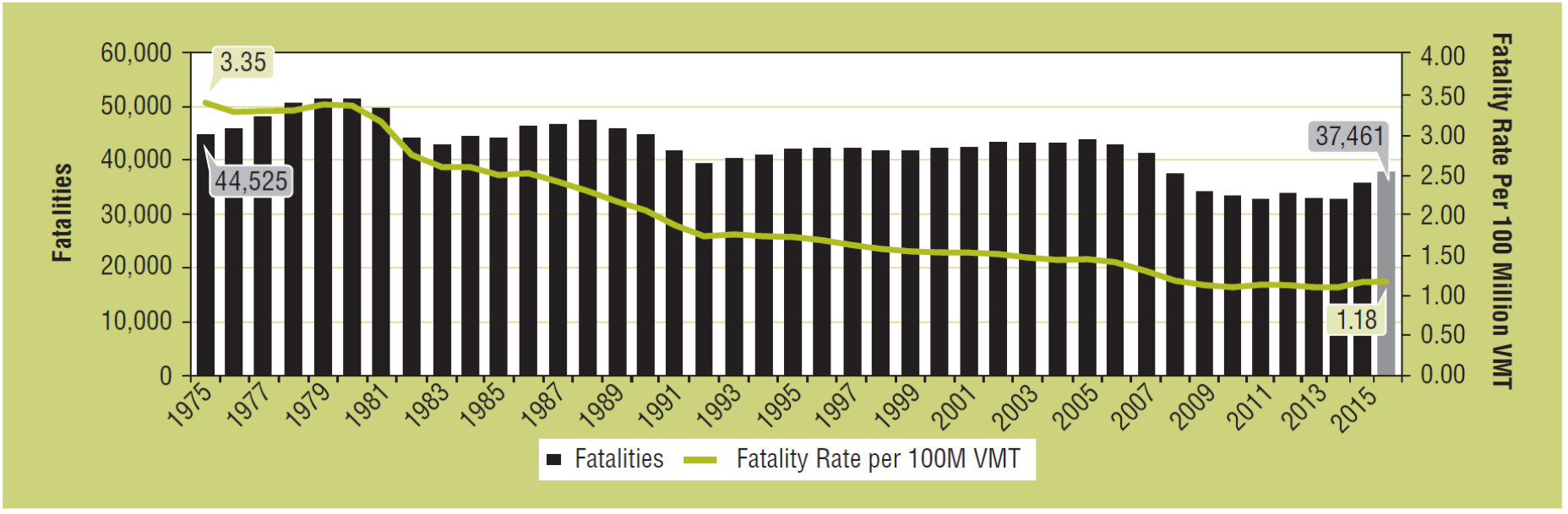
  Dan Morgan, Chief Data Officer and Acting Chief Technology Officer

- **Q&A**

Years of steady improvement in highway safety have ended

SAFETY IS #1 PRIORITY

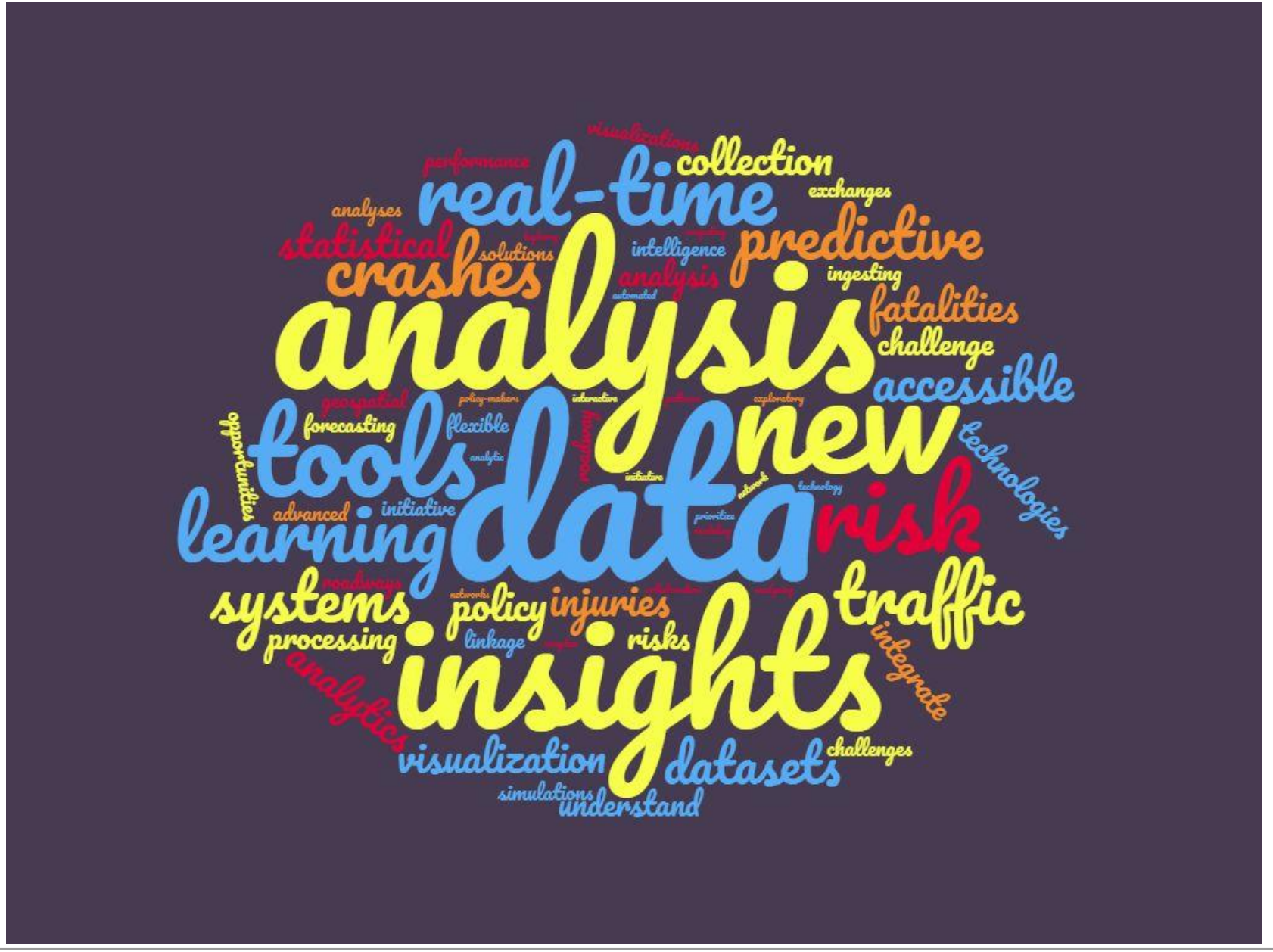
Figure 2
Fatalities and Fatality Rate per 100 Million VMT, by Year, 1975–2016



Sources: FARS 1975–2015 Final File, 2016 ARF; Vehicle Miles Traveled (VMT): FHWA.

We need better tools to understand why

Data science can improve our understanding



Vision of the Safety Data Initiative

- Gain **new safety insights** with advanced analysis of traditional and new "big data" sources
- Provide decision-makers and front-line safety professionals with **new tools and near-real-time information** so interventions can be focused and effective
- Evolve from **retrospective to predictive analytics** to better target emerging risks

Vision of the Safety Data Initiative

- **Integrate** existing databases and new private sector data sources to gain new insights into safety questions;
- Create new **visualizations and tools** that allow rapid and rigorous analysis by decision-makers through clear, compelling, and interactive maps, graphs, and charts; and
- Use **advanced analytics** to identify risk patterns and develop insights that anticipate and mitigate safety risk to reduce injuries and fatalities.

Overview of the Safety Data Initiative

Safety Data Initiative: Pilot Projects and Outreach

1. New **visualizations** of **FARS traffic fatality** data

2. **Data integration** for insights on **neighborhood pedestrian fatality risk**

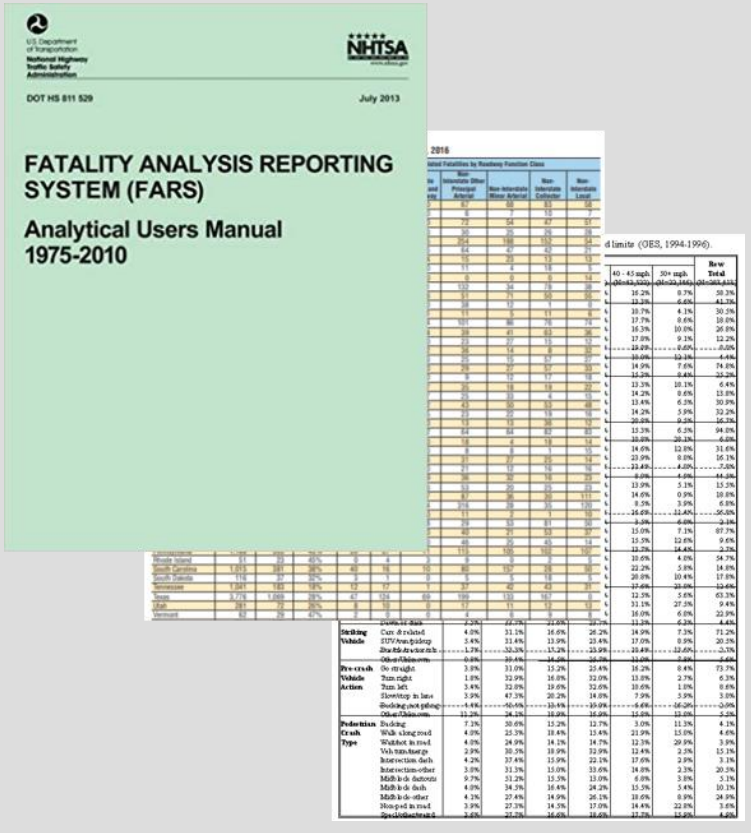
3. Potential for **advanced analysis** of **Waze data** for safety insights

4. **Integration of traffic speed, roadway attributes,** and crash data for rural safety insights

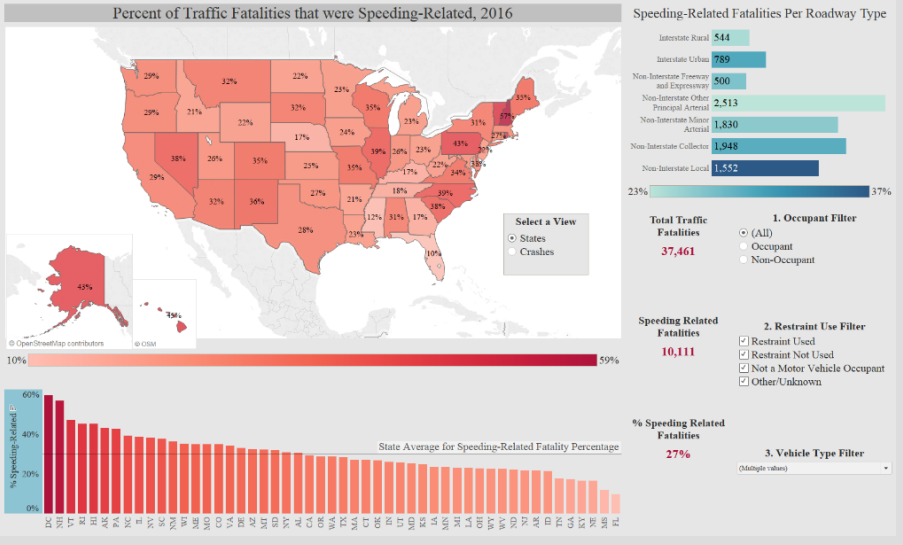
5. Convene expert **roundtables**, gather **best practices**, and launch **challenges**

I. Visualization helps discover anomalies, patterns, and trends

CURRENT WAY DATA IS REPORTED

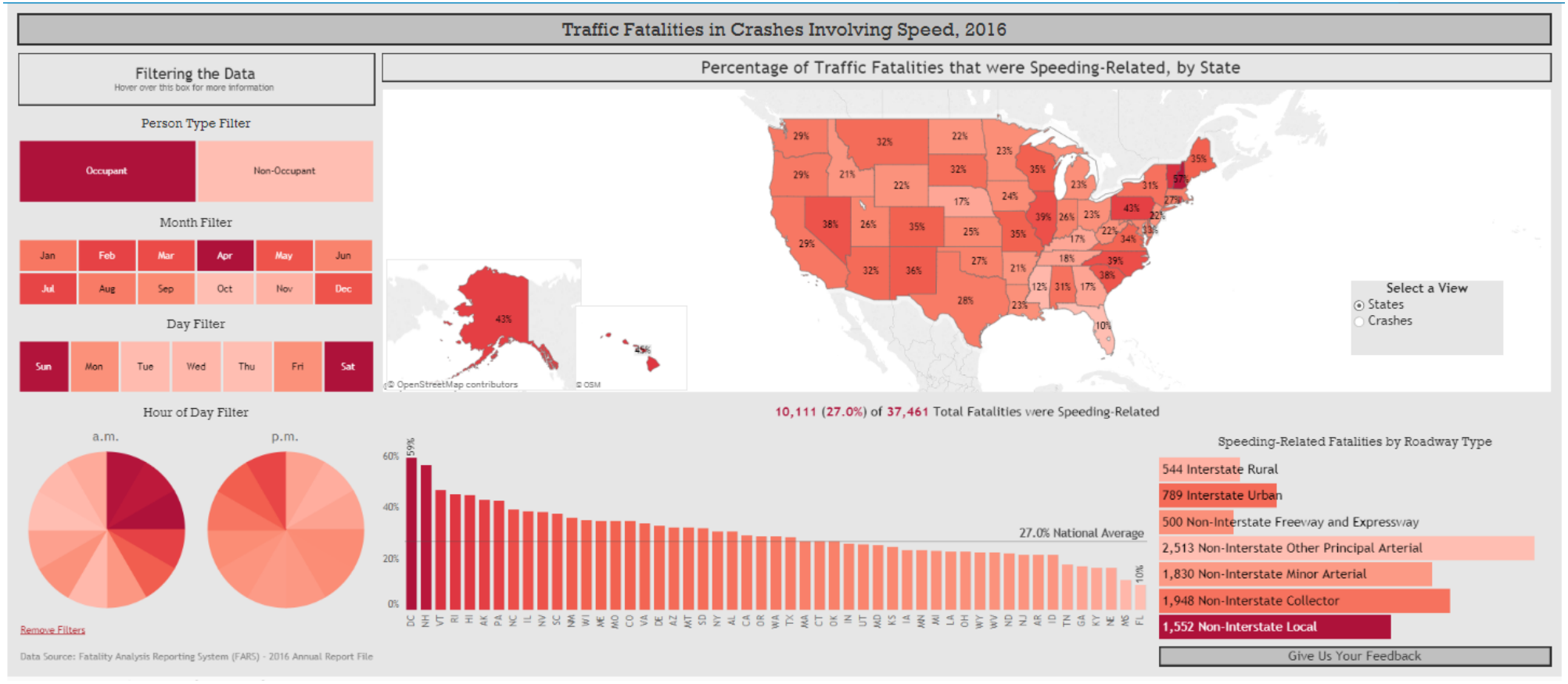


FUTURE WAY TO VISUALIZE DATA



- Visualization enables decision makers to see analytics presented visually
- Can help identify new patterns and enables interactively changing how data is presented

1. FARS Data Visualization: Speed Involvement



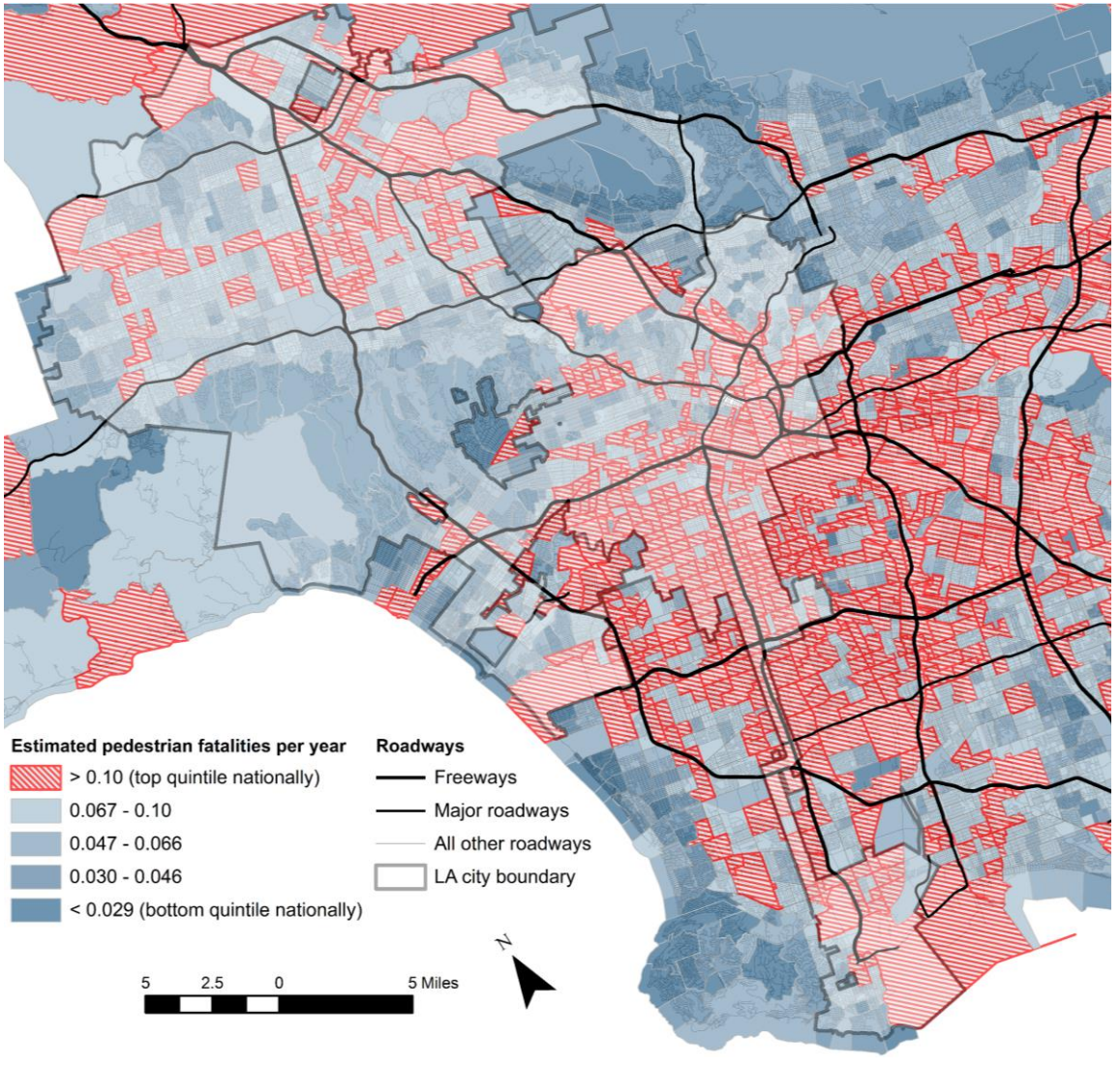
II. Data integration provides a richer understanding of an environment

DOT MANAGES ~800 DATASETS AND REPORTS; VAST MAJORITY ARE NOT INTEGRATED

5-Star Vehicle Safety Ratings · Accident Data, Reporting, and Investigations · Automated Track Inspection Program (ATIP) · Cargo Tank Facility Registration · Child Safety Seat Ease-of-Use Ratings · Crossing Inventory · **Fatality Analysis Reporting System (FARS)** · FHWA Fiscal Management Information System (FMIS) · Hazardous Material Safety Permit Program (HMSP) · HAZMAT Incidents **Highway Performance Monitoring System (HPMS)** · Highway Safety Improvement Program (HSIP) · Highway Safety Information System (HSIS) · Highway Statistics Series · Incident Reporting Large Truck Crash Causation Study (LTCCS) · Long-Term Pavement Performance (LTPP) Motor Carrier Census (MCMIS) · National Automotive Sampling System (NASS) National Hazardous Materials Route Registry (NHMRR) · National Household Travel Survey (NHTS) · **National Performance Measures Research Dataset (NPMRDS)** · National Pipeline Mapping System (NPMS) Public Map Viewer

2. Data Integration to Analyze Pedestrian Fatalities

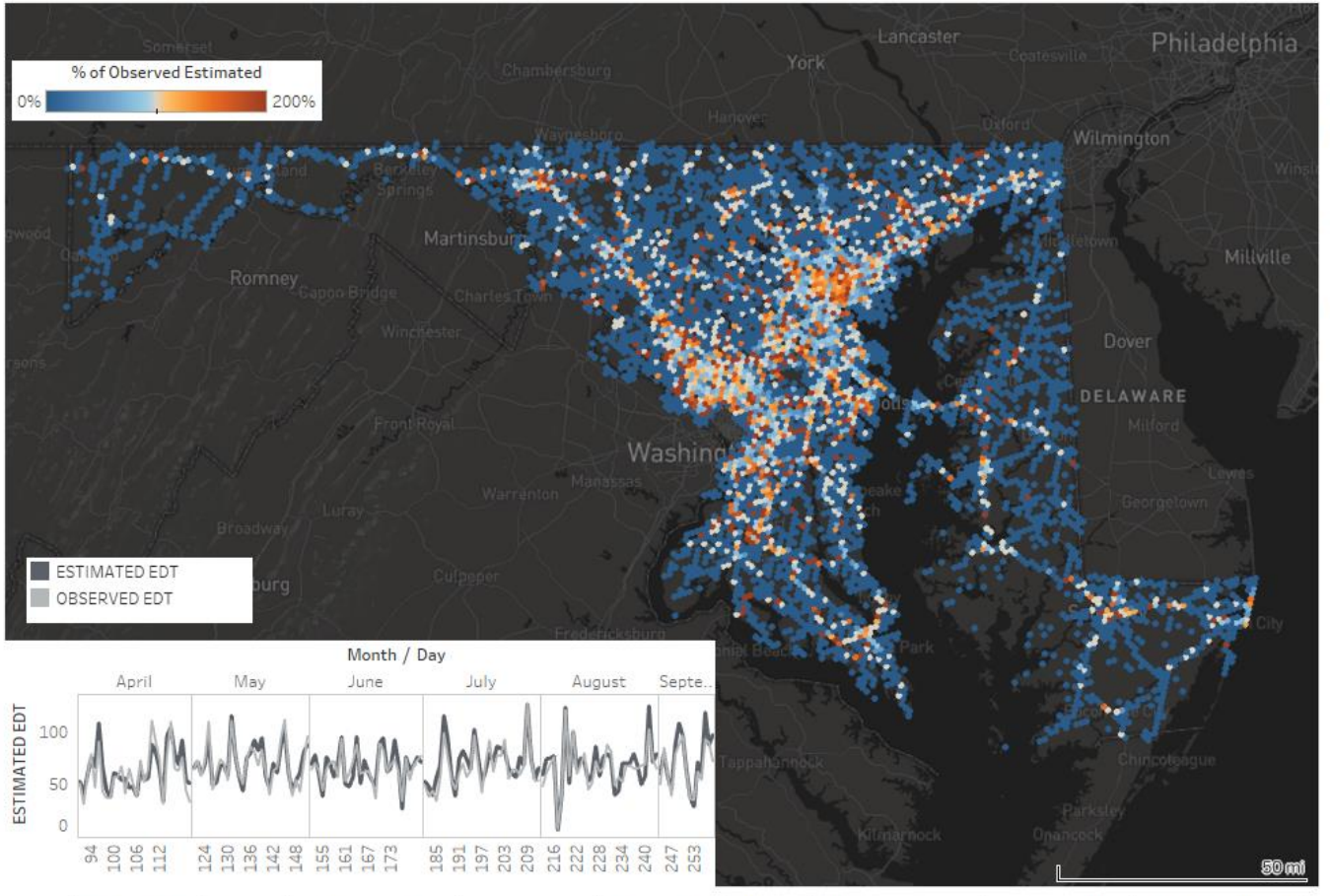
LOS ANGELES



5-Year Pedestrian Risk Estimate

3. Using Crowdsourced Data and Machine Learning for Safety

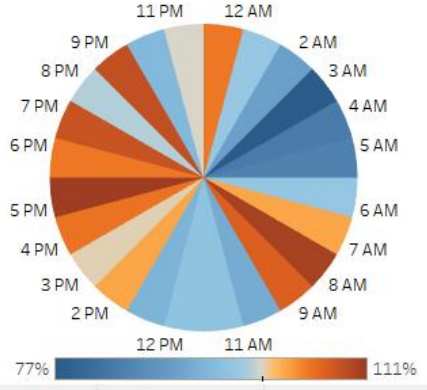
Percent of Observed Estimated EDT Crashes - *Output of Model #30 with 30% Test Sample)*



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

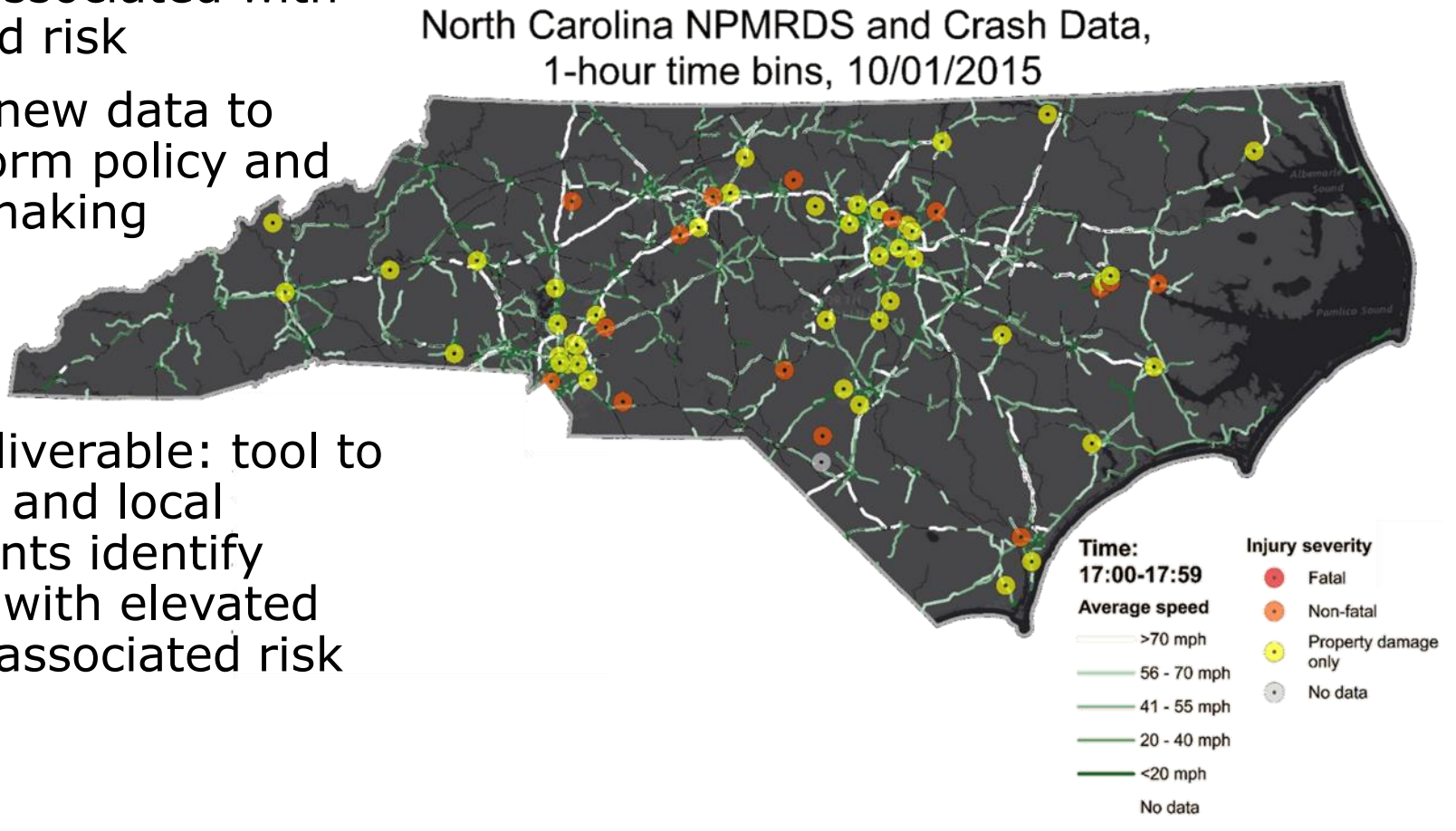


asonality | est - rose plot | est - map | est - day of week tbl | est - seasonality | diff - rose plot | diff - map | diff - day of week tbl | diff - seasonality | prct - Dashboard | est - Dashboard | diff - Dashboard

Model estimates highly accurate overall; miss some precise patterns

4. Rural Speeding Risk Using Near Real-Time Speed Data

- Identify traffic speed patterns associated with heightened risk
- Leverage new data to better inform policy and decision-making
- Dream deliverable: tool to help state and local governments identify roadways with elevated speeding-associated risk



5. Outreach and Solving for Safety

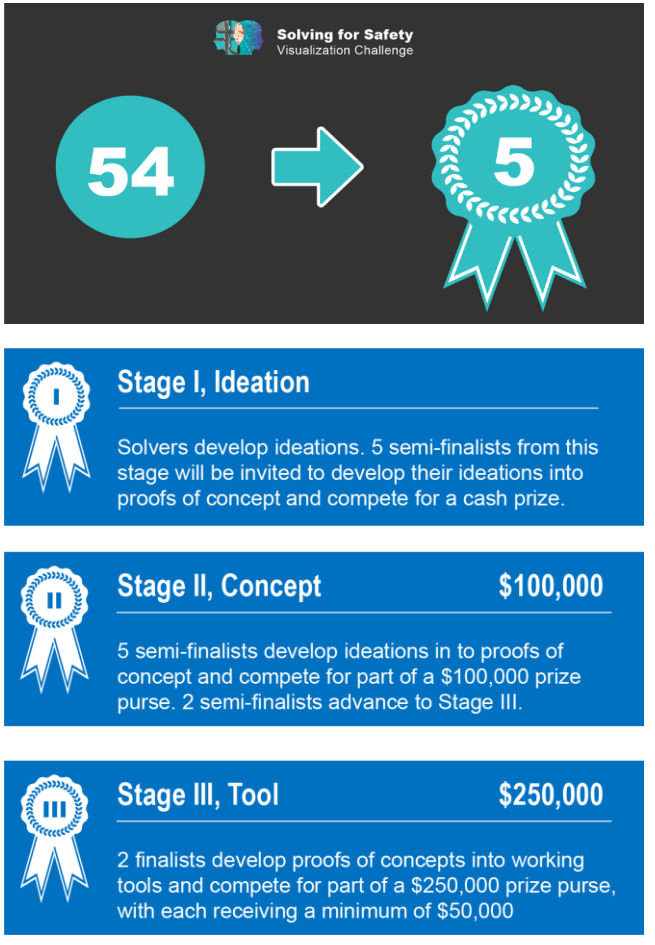
Solving for Safety Visualization Challenge

- Objective
  - Engage safety groups, universities, and tech firms to develop tools that use innovative analytical visualizations to advance traffic safety
- Propose tools to tackle one of the following needs:
  - Discover insights to reveal trends and patterns
  - Inform policy and decision-making by visualizing simulated results
- Tools to address one of the following safety focuses:
  - Vulnerable system users (e.g., pedestrians, bicyclists)
  - Conflict points (e.g., intersections, highway-rail grade crossings)
  - High risk factors (e.g., impaired drivers, speeding)

5. Solving for Safety Status

Solving for Safety Visualization Challenge

- **Stage II: Proof of Concept Development**
  - **Arity (Allstate)** - use connected vehicle and driver behavior data to explore the relationship between driving behavior and road design
  - **Ford Motor Co.** - combine crash data with connected vehicle and driver behavior data to determine crash risks, test solutions, and evaluate results
  - **Uber** - combine Kepler.gl, its web-based tool that visualizes geolocation data sets, and historical speed data collected from Uber trips with NHTSA's FARS data to better visualize traffic safety data
  - **University of Central Florida** - integrate real-time and static traffic data and use predictive analytics to help diagnose real-time safety conditions
  - **VHB** - use pedestrian avatars and apply game theory techniques to help state and local transportation professionals "see" potential safety improvements from the pedestrian's perspective



Outreach and Stakeholder Engagement

- Convened Diverse Stakeholders for Safety Data Forum (June 2018)
  - -Technology and data firms
  - -Transportation safety organizations
  - -State and local safety officials
- Identify Best Practices
- Collaborate with State and local governments

Request for Information

Current State

- Objectives
  - Take a systematic approach to safety
  - Use data to identify risk patterns, prioritize risks, and develop effective solutions
  - Improve collection, management integration, analysis, and use of data

- Environment
  - Transportation safety involves a multitude of stakeholders and data sources
  - Transportation risk management is an increasingly multi-model challenge that requires action across organizational boundaries

The Opportunity

A changing data environment …

- New technologies, capabilities and tools
- Proliferation of data available from new sources
  - Increasingly timely and contemporaneous
  - Increasingly granular and fast-moving
- Diverse and evolving technology environment

…presents new challenges

- Capacity to harness emerging analytics, integration, and visualization to drive predictive safety insights
- Data integration can be time-consuming
  - Data linkage, mapping, and merging are common issues
  - Complex data lineage with missing documentation
- Ability to ingest, process, and manage new data sources

Ideal End State

Current Challenges	Potential Procurement Solutions
Data integration is a time-consuming and difficult task, which reduces our ability to achieve quick results	Data integration processes that permit the repeatable integration of data for analysis. The integrated data will be useful both internally and externally
USDOT has limited capacity with regard to statistical methods, data analysis, model development, and staff capacity for predictive safety insights	Access to vendors that can perform quick analyses using innovative techniques to answer research questions that broaden our understanding of transportation safety through predictive insights
Acquiring new data sources and ingesting, processing, managing, and using new data – especially "big data" – is ad hoc and is dispersed throughout DOT	Creation of the necessary data systems and procedures to ingest, process, manage, and use new and more real-time data sources, especially "big data"

Assumptions & Objectives – Data & Tools

- Enable a more effective, frequent, and rapid exchange of data and information between USDOT and the rest of the transportation community; data connections and linkages will be created and shared amongst the transportation community
- USDOT will not be the sole source of data, nor the sole entity able to ingest, process, and manage data into usable information
- Data ownership will vary depending on the data and their use
- Data integration will build on previous research in order to reduce the time it takes to perform innovative analysis
- Data governance, use, accessibility, and technology standards that are flexible and appropriate for data with a spectrum of restrictions
- Balance appropriate data stewardship and access with encouraging open, available, and transparent data

Assumptions & Objectives – Task Management

- Sufficient flexibility to enable a multitude of on-demand projects that span scope and purpose, safety activity and safety topic, participants, datasets, and analytical tools used

- Flexibility for active participation, collaboration, and exchanges between USDOT and other members of the transportation community

- Effectively manages and leverages complex relationships that may go beyond the traditional government to vendor relationship

Assumptions & Objectives – Outcomes

- Build USDOT's capacity to convert data and research into information that proactively and effectively informs policy and decision making across surface transportation modes and for all users of the system

- Expand USDOT's capacity to perform prospective analysis focused on risk and predictive insights through a mix of traditional and new, innovative analytical tools and statistical methods

- Identify data quality and integration challenges, and initiate both top-down and bottom-up solutions to data use barriers

- Develop and deploy shared data service models with adequate security and privacy protections to foster a robust exchange of information among transportation community members

Anticipated Deliverables

- Identifying, collecting, and analyzing new safety data, data sets, data systems, and tools
- Integrated, connected, and/or fused data sets from multiple sources that are ready to be used for analysis, which could include data quality control, data cleaning, data standardization, and readily available and accessible data for individual projects or recurring safety data needs
- Facilitate the establishment of data management and sharing agreements amongst transportation community partners, both with and without USDOT as a direct partner, and vendors for safety data analysis
- Descriptive and statistical analysis using analytical tools and techniques for short and long term information needs
- Conduct exploratory safety analysis on available data sets using analytical tools and techniques
- Complex analysis using a variety of tools, including inferential statistics, modeling and simulations, artificial intelligence and machine learning, and non-statistical approaches
- Data visualizations that translate analysis and research results into useful information
- Data visualizations that permit the exploration of data by both the sponsoring entities and the wider transportation community unaffiliated with the project
- Continual shared services as the lynchpin of a data exchange network over a period of time

USDOT Questions & Requests for Information

Overarching Questions

- What frameworks and strategies should USDOT consider based on our focus areas (integration, visualization, and insights), strategies, and ideal end state?

- Are there additional deliverables that USDOT should consider based on our focus areas, strategies, and ideal end state?

Data Integration

- Describe and provide examples of your capabilities and past performance in integrating, connecting, and/or fusing multiple data sources.

- Describe the data integration methods used, including deterministic and probabilistic linkage techniques, the use of data attributes such as time and space, or other integration methods. If applicable, describe how you have made integrated data accessible and repeatable for others to use in future analyses and projects.

- Describe your demonstrated approaches to providing services that employ data science platforms, cloud-native data analysis products and services, and ongoing governance/management of such platforms and the resulting analytics.

Data Exchange

- Describe successful models that enable the exchange and use of information amongst a number of contributing parties in an open or controlled-access system, and the role(s) vendors could play in establishing and maintaining such as system.

- Describe and provide examples of your demonstrated approaches to establish and maintain a successful, effective data exchange system. Propose new, innovative approaches to establishing data systems and data exchanges that are flexible and can be adapted over time, if applicable.

- Describe successful voluntary arrangements that are facilitated by vendors and enable effective data integration and analysis between Federal, State, and/or local government entities. Highlight the participant benefits of using third-party vendors for intragovernmental information exchanges.

Data Analysis

- Describe your demonstrated approaches to performing the types of analysis below. If applicable, provide example(s) for each of the four types of analysis.
  - Inferential statistics. Linear models; non-linear models; spatial regression; Bayesian statistics; econometrics; and survey sampling.
  - Modeling, simulations, and forecasting. Commodity, freight flow, and market forecasting; probabilistic risk analysis and actuarial sciences; transportation related modeling, forecasting, and simulations.
  - Artificial intelligence and machine learning. Big and unstructured data analytics; neural networks and deep learning; qualitative content analysis and language processing; video, photo, and image analysis.
  - Non-statistical approaches. Geospatial analysis, including analysis that involves linear referencing systems; root cause investigation; content and survey questionnaire design; audio and video digitization and analysis.
  - Any other type of data analysis that was not mentioned prior that would be applicable to transportation safety analysis.

- Describe and provide examples of your approach to exploratory analysis of data sets, especially big data sets, to determine their suitability for a particular analytical application or research question. Address your approach for discovering and correcting for bias and other factors that might impact a dataset's fitness for purpose.

Data Visualization & Management

- Describe your data visualization capabilities, including tools and techniques applied. If applicable, describe your experiences visualizing: multiple data sources into a cohesive narrative; very large data sets and "big data" (due to volume, variety, and/or velocity); and advanced analytical techniques described in part D. Provide examples of data visualization performed, if applicable.

- Describe and provide examples of your capabilities in ingesting, cleaning, storing, managing, and maintaining data. If applicable, describe how your data management led to a flexible, accessible data environment.

- Describe your demonstrated ability to work with "big data" (primarily high variety data. You may also discuss high-volume and/or high velocity data).

- Describe your experiences in ensuring privacy, business confidentiality, and how security sensitive information is managed in a data system environment used by multiple government, private, and non-profit entities. What systems and protections would need to be in place to build trust and encourage the exchange of information?

Partnerships

- Describe and provide examples of successful vendor consortiums for data integration, analysis, and data visualization that leverage the capabilities of multiple vendors to have a holistic set of competencies to perform a wide variety of work.

- Describe successful vendor consortiums that involve a mix of entities, which could include large for-profit companies, small and disadvantaged businesses, non-profit entities, universities, quasi-public entities, research consortiums, and other applicable vendor types.

Procurement Method

- Describe government and private sector arrangements used by other government entities that successfully facilitated data analysis, integration, and visualization objectives similar to the descriptions provided above. For example, Broad Purchase Agreements, Indefinite Delivery/Indefinite Quantity contracts, etc.

- Describe and provide examples of vendor experiences with successful business model approaches that would provide mutual value to the vendor and USDOT, and permit the level of on-demand, flexible services for data systems and services described in the RFI.

Questions?

Safety Data Initiative

https://www.transportation.gov/policy/transportation-policy/safety/safetydatainitiative

Responses to RFI Due:

December 7, 2018, 2:00 pm Eastern

USDOTSafetyDataInitiative@dot.gov