DEVELOPING AN INTEGRATED FRAMEWORK FOR SAFETY AND MOBILITY ANALYSIS

Final Report

(Year 1)

SOUTHEASTERN TRANSPORTATION CENTER

MEI CHEN, OBAIDUR KAZI, BRYAN GIBSON, and XU ZHANG

OCTOBER 2015

US DEPARTMENT OF TRANSPORTATION GRANT DTRT13-G-UTC34

DISCLAIMER

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Developing an Integrated Framework for Safety and Mobility Analysis	5. Report Date October 2015 6. Source Organization Code
7. Author(s) Mei Chen, Obaidur Kazi, Bryan Gibson, Xu Zhang	8. Source Organization Report No.
9. Performing Organization Name and Address Southeastern Transportation Center UT Center for Transportation Research 309 Conference Center Building Knoxville TN 37996-4133	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC34
12. Sponsoring Agency Name and Address US Department of Transportation Office of the Secretary of Transportation–Research 1200 New Jersey Avenue, SE Washington, DC 20590	13. Type of Report and Period Covered Final Report: 6/2014-12/2015 14. Sponsoring Agency Code USDOT/OST-R/STC

15. Supplementary Notes:

Abstract

The growth of Big Data has begun to transform research and policy, and can empower transportation agencies to address issues of operational demand and efficiency. Advances in detection and communications technologies have led to improvements in various traffic sensors, and subsequently, have increased the amount of transportation-related data. These data contain large amounts of information on the condition and operational characteristics of infrastructure as well as knowledge of spatial-temporal variations in traveler behavior. These data enhance our understanding of the interactions between the travelers and the traveling environment, leading to better safety monitoring, improved design strategies, and enhanced traveler safety. To achieve this study's research objectives, the team examined Big Data sources and reviewed current practices and applications of Big Data. The information was used to analyze: 1) the impacts crashes have on travel time reliability/variability, 2) the variation of crash impacts during peak and off-peak periods, and 3) how different crash types produce significantly different outcomes. A spreadsheet-based visualization tool — a space-time velocity map (or heat map) — was developed to graphically represent the temporal and spatial extent of crash impacts. Heat maps enhance our visual and intuitive comprehension of the impact of crashes. The team also developed cumulative distribution functions for travel rates, which communicated a compelling story about crash impacts. Reliability analysis found that crashes negatively impacted travel time reliability. This study illustrated that Big Data can help agencies better understand crashes, their impacts, and their distribution in a spatial-temporal domain.

16. Key Words Heat map, space-time velocity map, incident, crash analysis, big data	17. Distribution Statement Unrestricted; Document is available to the public through the National Technical Information Service; Springfield, VT.

18. Security Classif. (of this report) Unclassified	19. Security Classif. (of this page) Unclassified	20. No. of Pages 78	21. Price …

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The growth of Big Data has begun to transform research and policy, and transportation is no exception. As detection and communications technologies grow more sophisticated, transportation sensors continue to generate more data. Abundant data are available on infrastructure condition, operating characteristics, and traveler behavior across temporal and spatial domains. These data present an opportunity to better understand the interactions between travelers and the traveling environment. An inventory of available data sources will identify potential applications that will ultimately improve safety outcomes, and in turn, lead to the design of better safety improvement strategies.

Big Data gives agencies an unprecedented ability to monitor real-time transportation system data, to adjust practices, and to confront changing dynamics. It can empower transportation agencies to address issues of operational demand and efficiency. A challenge for these organizations is dealing with the wide variety of data that is now available — integrating data sets to establish a more comprehensive picture of road networks. This report explores the potential application of Big Data, its uses, benefits, and future challenges. Case studies are analyzed to better understand the strength of data applications. Cataloguing current practices and sources of data will allow development of more innovative data analysis methods.

To achieve the research objectives, the team first reviewed current practices and applications; we showed that adoption of Big Data analysis has been incremental, with room for continued growth. The team next examined Big Data traffic safety sources, including traffic monitoring data, incident logs, GPS-based travel data, and connected vehicle-related data. This report lists key data sources, their contents, and offers a brief description of each data set. To generate useful applications for Big Data, it is important to understand current methods and approaches used across governments.

The team then leveraged a stationary sensor-based regional data source: an application that integrates several sources of Big Data to assess traffic flow. The Louisville-Southern Indiana

Traffic Information (TRIMARC) system records speed, volume, and occupancy data at 15-minute intervals throughout the day; incident logs are included. Using TRIMARC data, the research team performed an empirical analysis of: 1) the impacts crashes have on travel time reliability/variability, 2) the variation of crash impacts during peak and off-peak periods, and 3) how different crash types produce significantly different outcomes. The study area was a segment of Interstate 65 (I-65) in metropolitan Louisville, Kentucky. As part of the analysis, the team developed a spreadsheet-based visualization tool — a space-time velocity map (or heat map). After an incident identifier, the incident information and the corresponding traffic speed of that day were placed into the spreadsheet, and a heat map was generated. The heat map let readers visualize the spatial and temporal effects of crashes, and enhanced visual understanding of the impact of crashes. Analysis of influencing factors revealed that the cumulative distribution function (CDF) of travel rates told a better story about the impacts of crashes: it showed how crashes affected travel rates during peak and off-peak periods. Reliability analysis found that crashes negatively impacted travel time reliability. Travel time variability (TTV), travel time index (TTI), planning time index (PTI), and buffer index values significantly increased when crashes occurred. To realize the full power of this data, more data sources should be used. This study illustrates that Big Data can be used innovatively to improve the understanding of crashes, their impacts, and their distribution in a spatial-temporal domain.

1. Big Data: Current Practice and Applications

1.1 Introduction

Technological advances in recent years have generated increasingly large volumes of data, which conventional processing techniques are unequipped to analyze. The emergence of Big Data has compelled researchers to address numerous underexplored issues, including transportation safety and mobility. The output of Big Data has resulted from the convergence of many social and technological factors, such as (Bakshi 2012):

- **Mobility:** mobile devices, mobile events, and sensory integration
- **Data access and consumption:** Internet, sensors/actuates, interconnected systems, social networking, convergent interfaces and access models (Internet, search and social networking, and messaging)
- **Information model and open source:** major changes in the information-processing model and the availability of an open source framework

Both the public and private domain have contributed to generating Big Data, from government-level data on health services to consumer loyalty cards to social media to sensors (Wigan and Clarke 2013). Many stakeholders attempt to integrate Big Data into applications to enhance user outcomes. Public agencies have the power to improve performance and efficiency through data-driven performance metrics. To generate useful practice and applications from Big Data, it is important to understand current methods and approaches government organizations use to analyze these data sets. Big Data can be used to improve safety, mobility, and other dimensions of the transportation system.

1.2 Current Practices and Methods Used to Collect and Archive Data

Identifying and evaluating current practices for collecting, archiving, and analyzing Big Data will influence future research efforts. Cataloguing the current analytical methodologies employed to analyze large data sets will help researchers to develop more innovative, synergistic strategies and methods for using Big Data. This literature review focuses on the methods and findings of previous researchers who have used Big Data to address transportation safety.

Tarko (2012) developed a novel crash-modeling paradigm that accounted for the events that precede a crash. Previous research had focused on statistical associations between traffic interactions and accident frequency. Tarko cited the lack of data used to execute previous analysis as a critical drawback. A second method, causal modeling with counterfactual evaluation of the effects demanded that researchers calculate all of the factors associated with car crashes. Both approaches can underestimate the likelihood of crashes, but Tarko improved representations of the crash occurrence process.

Tarko (2012) collected data from four drivers in a driving simulation over the course of several weeks. He found an inverse relationship between crash severity and crash frequency. He argued that precisely modeling accidents is immensely challenging. Many factors are unpredictable and cannot be reliably modeled mathematically or by using simulators. Tarko's model had four components: the count model of accident frequency, the probability model of a risky event, the probability model of a crash, and a probability model of crash severity. This model was empirically tested with driving simulators. The results demonstrated that crashes can be modeled more accurately to predict future crash events.

Giuliano et al. (2014) developed a data archive using Los Angeles area traffic data from the Los Angeles area sourced from the Regional Integration of Intelligent Transportation Systems (RIITS). This extracts data from Los Angeles's Archived Data Management System (ADMS). The ADMS includes (p. 3; see Table 1):

> …freeway, arterial and public transit data. This large, dynamic, real-time database integrates transportation data from multiple sources (public transit operators, city departments of transportation, various districts of the California State Department of Transportation, local and state police agencies).

Table 1: ADMS Data Sources

Agency	Data Type	Data Attributes	Frequency
	Freeway Detector Inventory and Real-Time Data	Routes	
Caltrans District 7, 8 and 12	Arterial Detector Inventory and Real Time Data	Cross Streets	
	Event Data	Directions	Varied: once per 30 second, minute, day; twice per year
	Ramp Metering Inventory and Real-Time Data	Occupancy	
LADOT	Arterial Detector Inventory and Real-time Data	Volumes	
Metro Bus	Vehicle and Route Data	Speeds	
Metro Rail	Vehicle and Route Data	Geo-locations	
CHP	Event Data	Clear Time	

The data archive can be used to measure the transportation system’s performance at various times. Giuliano et al. observed that historical data rich in detail was critical for modeling and simulation. Using one week of sample data, the study leveraged cluster analysis to measure average speeds and identify spatial variability in transportation system performance (e.g., congestion levels during rush hour rural areas). Future iterations of their approach could include other performance measures such as traffic volume and delays.

Wiliszowski et al. (2010) discussed the type of data that law enforcement agencies can collect and how that data may be leveraged to improve research on transportation safety. They found that data collection was not standardized across law enforcement agencies. Inconsistent data collection creates challenges for researchers who conduct large-scale studies among different states and municipalities. Random data sampling is more feasible and more consistent than performing convenience-based studies on the available data. Increasing cooperation across law enforcement agencies, governments, and research organizations can improve data quality. The aim is to create a national, standardized database that houses public safety data on crashes and traffic violations.

Parikh and Hourdos (2014) used radar to measure vehicle speeds on rural roads with dangerous horizontal curves. This comparative study documented the effects of road signs on driver behavior by contrasting average vehicle speeds before and after new warning signs were installed; these signs warned drivers of the approaching curve and of the necessity of lowering their speed. The data collection systems used battery-powered radar, and readings were decoded by software. The study validated the robustness of the collection system, and it could be used to collect speed-based data.

Muthyalagari et al. (2000) collected GPS based travel data in Lexington, Kentucky, to study driver behavior over several days. Data were drawn from a Federal Highway Administration (FHWA) program that tracked 100 households. The research found that travel patterns in Lexington generally mirrored patterns documented by previous studies. They attributed differences to socio-economic factors and to drivers being more likely to take short and/or infrequent trips.

Multiday travel measurements can provide planners with abundant information, such as the frequency of trips and average rate (Jones and Clarke, 1988). Additional research has indicated that sampling over multiple days can yield more efficient modeling efforts and increases the cost effectiveness of samples (Pas, 1986). Although Muthyalagari et al.'s (2000) effort was small, their methodology could be applied to a wider sample to examine large-scale mobility patterns.

The Highway Safety Manual (AASHTO, 2010) states methods to evaluate safety:

> Included in the HSM is a quantitative method for predicting crashes on the basis of recently developed scientific approaches. The predictive method procedure includes the use of statistically derived equations known as safety performance functions (SPFs) that were developed for a set of base conditions unique to each facility type. The estimates calculated by using the SPFs should then be modified by using crash modification factors (CMFs) to help account for various changes in the segment or intersection nonbase conditions (Xie et al. 2011, p.19).

Xi et al. (2001) applied these methods to a case study for Oregon, using crash data. They found that while the quantitative evaluations were useful for safety, challenges remained. There were data gaps in pedestrian numbers, traffic volumes in rural areas, and in the number of observations needed for underrepresented crash sites.

1.3 Current Practice and Applications

Using Big Data can benefit governments in a number of ways. StateTech (2013) lists some potential benefits of utilizing Big Data:

- Make better informed, better decisions faster
- Improve outcomes
- Identify and reduce inefficiencies
- Eliminate waste and fraud
- Increase return on investment
- Improve transparency and service provision
- Reduce security issues and crime

Big Data can also enhance security, improve service delivery, offer the public and businesses more access to data, and give agencies the ability to develop applications in collaboration with governments (Cull, 2013). Big Data can improve threat mitigation (e.g., terrorism, disease), assist with crime prevention, and increase efficiency by streamlining processes, budget reductions, and compliance ("Four Federal" 2015). In government agencies, Big Data has streamlined operations, facilitated cross-agency collaboration across agencies, and trained employees. Chief information officers (CIOs) in state and federal governments contend that e-government has improved service delivery and reduced costs (West, 2000). Desouza (2014b) interviewed CIOs at all levels of government. Many CIOs expressed concern about the perception of Big Data as a fad, even though the types and availability of data continue to expand. A pressing challenge is identifying appropriate strategies to manage and analyze data.

Big Data can foster rapid decision making because of the growth in real-time information. Federal employees have observed four ways that Big Data can influence governments: performance tracking and goal setting, cost savings, budgeting, and increasing efficiency. Big Data can empower transportation agencies to address issues of demand and efficiency in operation (Buckley and Lightman, 2015), and it can be utilized by planners to discern patterns of travel behavior. Agencies can monitor real-time activity on transportation systems and introduce modifications to deal with changing traffic dynamics. Big Data can be harnessed to develop insights into traffic problems (e.g., crashes), improve their detection, and help agencies decide where to allocate their human and financial resources (Vasudevan et al. 2015).

Effectively using Big Data requires that researchers do significant planning during the early stages of a project. They will need to resolve legal issues, develop partnerships, define opportunities, prepare for obstacles or resistance, identify performance markers, and implement risk mitigation plans (Desouza, 2014b). One challenge for organizations is the variety of data and the integration of different data sets (Desouza, 2014b). In some cases, data may be scattered among a number of networks, requiring partnerships and collaborative efforts to gain access. Planning may also be the time to institute policies related to the management of Big Data.

General issues that hinder technology in the public sector are marketing, privacy, equity, and financing (Edmiston, 2003). Many of these are salient within the context of Big Data. Asai and Akiyama (2013), during a broader discussion about data integration and analytics within the context of intelligent transportation systems, noted that Big Data can be inconsistent. There are also privacy issues that may need to be addressed during data acquisition. Data ownership and rights may inhibit the use or dissemination of data (Wigan and Clarke, 2013). Buckley and Lightman (2015) explored data sharing, outsourcing, and partnerships with the private sector. Big Data can also be of little benefit if there is a lack of expertise or tools to explore the data and conduct analysis (Gou et al. 2015). If the data quality is poor, but is still used to guide decision making, less-than-optimal outcomes may result (Wigan and Clarke, 2013). Many agencies lack the resources to store and process Big Data, or even share it across departments. In some cases, agencies are skeptical of the costs and personnel changes needed to develop and implement Big Data programs. Tene and Polonetsky (2012) wrote that: "The principles of privacy and data protection must be balanced against additional societal values such as public health, national security and law enforcement, environmental protection, and economic efficiency." (p. 67). Another quickly emerging field is Big Data visualization. Many current visualization methods, such as GIS, lack bearing and change rates (Liu et al. 2015).

Floating traveler data tracks travelers in real time, can detect factors such as location and speed, and has a number of applications (Buckley and Lightman, 2015) Table 2 is adapted from Buckley and Lightman (6) to define the purpose and potential use of Big Data.

Table 2: Possible Uses of Floating Traveler Big Data

Purpose	**Potential use**
Network performance analysis	• Can measure congestion delays and system performance to identify trouble spots
Transportation infrastructure planning and demand management	• Examine choices including route and mode of transit; generate origin and destination options • Develop applications to assist with managing demand on infrastructure
Intervention analysis	• Evaluate need for operational interventions in transportation system and the benefits of intervention
Active Traffic Management	• Develop incident detection capabilities based on real time data • Implement active traffic management strategies to mitigate the effects of incidents
Traveler information and alert systems	• Disseminate information from data via signs, media, etc. • Develop alerts via apps; allowing private developers to utilize data for same purpose
Traffic forecasting	• Forecast traffic based on current conditions and historical patterns • Adjust operations to meet forecasts

Buckley and Lightman (6) also included a number of Big Data applications that were pertinent to transportation agencies. While these opportunities were likely dependent on data availability and on the agency, the many ways to implement Big Data provides a glimpse of potential applications. Applications of Big Data to safety and mobility applications include:

- Mapping congestion and generating alerts based on historical congestion averages
- Mapping cycling and pedestrian activity
- Mapping incidents and resulting closures
- Forecasts of travel times

Traffic engineering and signal operations can also potentially benefit from Big Data. Signal timing, traffic analysis, and signal phasing can be improved through the analysis of large data sets (Gou et al. 2015). “More historical data can let engineers obtain statistical information and thus depict a clearer picture of traffic trends” (Gou et al. 2015, p. 3). Generating traffic flow and predictive traffic information is another use of Big Data (Liu et al., 2015). Big Data can also be used to enhance visualization applications. Moreover, “the rise of Big Data has made it possible to use demand data at an operational level, which is necessary to directly measure the economic welfare of operational strategies and events” (Liu et al. 2015, p. 32). Analyzing Big Data can help stakeholders implement real-time, dynamic changes in response to traffic events. This can also reshape infrastructure design and better account for each driver’s behavior.

Buckley and Lightman (2014) identified key financing areas to harness the full potential of Big Data: talent, business systems, and organizational culture. People with mathematical and statistical backgrounds are needed to manage and analyze Big Data. These skill sets are often in short supply, and private competition to land talented candidates is often fierce. Business systems, such as data management and storage will require updated IT resources and expanded networking capabilities; the latter are needed to streamline data sharing across multiple agencies. Reworking organizational culture requires less of a financial commitment; rather, it relies on stakeholders and recognizing the usefulness of Big Data and cultivating an environment that promotes innovation in data analysis. Arguing for greater investments in these areas requires stakeholders to highlight the benefits of integrating and using Big Data.

Benefit-cost analysis can illuminate the returns on investment agencies may expect from their increased engagement with Big Data.

Buckley and Lightman (2014) also recommended several steps to start a Big Data program. Transportation agencies should focus on defined projects where leveraging Big Data will provide added value and increase the likelihood of success. From here, data should be identified and collected. Agencies should take advantage of outside expertise by forming partnerships with universities and other organizations. Finally, resources must be allocated to ensure the human and organizational capital needed to complete the project is available.

The lifecycle of Big Data encompasses the processes of capturing, sorting, analyzing, and consuming it. These processes must be incorporated into enterprise systems (Bakshi, 2012). By making investments in key areas and by managing the lifecycle of Big Data, governments can understand potential benefits and move toward operationalizing it. Once an organization has internalized a thorough understanding of Big Data, it can take additional steps to manage and use it more effectively. Bakshi listed these approaches that organizations could consider:

- **Find technology enablers:** These could be new infrastructure, software applications evaluation and pilots.
- **Adopt an ecosystems approach:** Big Data is a new and emerging space, and there will be several upcoming technology options to review and select from.
- **Adopt a use case-based approach:** Data's value depends on the insight of the domain. Hence, look for use case-specific projects — for example, use cases of network-centric Big Data analytics or cybersecurity and video-based insights.
- **Invest in data-centric skill sets:** The insights provided by large data sets are only as good as the domain knowledge of the data. Therefore, skills for data analysts and scientists need to be developed and nurtured.

Drees and Castro (2014) measured the responsiveness of state governments to making data available to the public. States do this through open data policies, which allow public access and/or open data portals (e.g., when government data is stored in a repository). The authors' responsiveness score for each state was based on several factors, including their open data policy, quality of the policy, presence of an open data portal, and quality of the portal. The maximum score a state could achieve was 8. Table 3 presents the scores for each state.

Table 3: Data Availability Scores by State

State	Total Score
Alabama	1
Alaska	1
Arizona	2
Arkansas	2
California	4
Colorado	3
Connecticut	7
Delaware	3
Florida	2
Georgia	2
Hawaii	8
Idaho	2
Illinois	8
Indiana	3
Iowa	3
Kansas	1
Kentucky	2
Louisiana	1
Maine	3
Maryland	8
Massachusetts	1
Michigan	4
Minnesota	3
Mississippi	2
Missouri	4
Montana	3
Nebraska	3
Nevada	1
New Hampshire	6
New Jersey	4
New Mexico	2
New York	8
North Carolina	3
North Dakota	2

Ohio	3
Oklahoma	8
Oregon	4
Pennsylvania	2
Rhode Island	6
South Carolina	2
South Dakota	1
Tennessee	2
Texas	7
Utah	8
Vermont	4
Virginia	3
Washington	3
West Virginia	2
Wisconsin	3
Wyoming	1

1.4 Big Data Case Studies

While Big Data can confer benefits to all levels of government, practical applications provide the most persuasive evidence of Big Data in action. Analysis of Big Data has been popular in the areas of traffic and water management, emergency response, and public safety (Shueh, 2015). Electronic toll systems can support varying payment schedules based on measures of congestion and mobility taken in real-time (Tene and Polonetsky, 2012). These schedules assist with emissions control efforts when environmental concerns are present. Cull (2013) noted that data from transportation systems has allowed planners to coordinate smoother responses to traffic and transit disruptions.

Additionally, cities that use 311 information systems collect data that can address emergency management problems. Chicago is building a predictive analytics platform that identifies and analyzes data trends. It also generates predictions that can facilitate problem solving. Windy City, the City of Chicago's name for its live analytics dashboard, can be accessed from all city departments. The goal is to have open source data and to identify performance areas where big data and analytics can make the greatest difference. Boston is using smartphone data as part of its "Street Bump" project to better improve its mapping of road conditions. The goal of this project is to pinpoint and target areas where street repairs are needed, so they are done quickly and inexpensively ("Cloud computing" 2012). Many local governments have begun to rely on Big Data, especially police departments, which are using it to identify spatial and temporal tends in crime. As Shueh (2015) observed, "despite the conflicting signals, governments are gradually adopting Big Data tools and strategies, led by pioneering jurisdictions that are piecing together the standards, policy frameworks and leadership structures fundamental to effective analytics use." Investments in "smart government" technologies are expected to grow to over $1 billion by 2017.

Buckley and Lightman (2015) described the use of Big Data in several agencies. The City of Dubuque, Iowa, examined public transit ridership through usage data and origin-destination capture. The program tracked volunteers' use of public transit. Information about transit and routes was provided to volunteers based on their travel and ability to save money by utilizing public transit. The project helped increased public transit use. The I-95 Corridor Coalition, which includes states along Interstate 95, used cooperation among state agencies to generate real-time travel information along the corridor. The data generated has increased states' monitoring abilities while reducing costs, and improved their incident management and responses. San Francisco, California sought to improve parking management in the downtown area. The program used meters, sensors, and other tools to institute variable pricing. The program reduced rates and alleviated demand for parking. Lastly, the Oregon Department of Transportation contracted with a private vendor to receive data from a smartphone app, which tracks bicyclists. The goal was to identify locations significant bicycle usage and traffic fatalities.

Vasudevan et al. (2015) utilized a large dataset acquired from a connected vehicle pilot program in Ann Arbor, Michigan. The data included basic safety messages, which contained speed, location, and other information. Data were divided into boxes at one-minute intervals. The boxes that were highly correlated were analyzed to detect patterns. Based on this work, predictive models were developed based on similar observations across correlated boxes. Miami-Dade County, Florida, uses Big Data methods to monitor water meters and generate cost savings by identifying leaks, while Louisville, Kentucky, employs sensors to track inhaler uses and measure pollution with the aim of reducing emergency room visits and improve care (Desouza, 2014a).

DeSouza (2014b) detailed a number of Big Data initiatives in federal and state government agencies. The United States Postal Service (USPS) uses data from scanned mail in some processing centers to detect fraud or suspicious mailings before they reach their intended destination. The Internal Revenue Service (IRS) has focused on developing ways to fix errors, detect tax evasion, and collect revenue owed. The agency uses taxpayer information and data on financial transactions to conduct “robo-audits” to track financial irregularities. Massachusetts began statewide effort, coined the Massachusetts Big Data Initiative, that will encourage collaboration between state agencies and researchers to analyze a variety of issues (“Cloud computing” 2012).

2. Data Sources Review

2.1 Introduction

In the past, traffic data were collected manually. This work was tedious, time-consuming, expensive, and often unreliable. Today, massive amounts of data are collected and available in real time due to the widespread adoption of Intelligent Transportation Systems (ITS). Big Data has transformed the outlook of science and engineering, including transportation. Transportation data are available for many highway systems and roads, where it is important to view traffic flow and how crashes, congestion, and roadwork affect speed, headway, and other traffic factors. The most widely used data sources are traffic monitoring systems — they continuously generate large amounts of data and capture traffic dynamics along segments of the highway network. Big Data applications can improve traffic safety and mobility, and can establish a sound understanding of how safety and mobility influence one another.

2.2 Data Sources

The first step toward fully integrating data sources and potential applications is to take inventory of available safety data. The aim of this process is to improve safety outcomes. Data can be drawn from federal, state, and local sources. The previous chapter described data available from other sources, but this section summarizes several federal sources, including Data.gov, a federal data repository. See Appendix A for a summary of data names, summary descriptions, and links.

Federal agencies devoted to transportation safety have issued reports and statistics that may be useful, however, many sources only point toward summaries, not datasets that can be mobilized for original analyses. Governments at the state and local levels are more likely to be good sources of micro-level data. When first embarking on a new project, there are many sources of basic information available to researchers and the public. The Bureau of Transportation Statistics (rita.dot.gov/bts/) is a repository of federal transportation data, and it contains links to the National Transportation Statistics and other agency data. Topical areas are classified by mode. The Federal Highway Administration (fhwa.dot.gov) maintains statistics on traffic fatalities, road conditions, and traffic volume trends. While much of the data stored by federal agencies are highway-centric, information specific to other modes is available on the Bureau of Transportation Statistics website and federal websites accessed via the Research and Innovative Technology Administration's (RITA) U.S. Department of Transportation (USDOT) Research Hub (http://ntlsearch.bts.gov/researchhub/index.do). This page contains a search tool and links to other federal transportation agency websites.

Other data sources are the Research Data Exchange (RDE) and TRIMARC. RDE, a system that promotes the sharing of archived and real-time transportation data derived from multiple sources (including vehicle probes) and for multiple modes. This new data sharing capability will support the needs of ITS researchers and developers while reducing costs and encouraging innovation. The different sources are summarized below.

2.2.1 Data.gov

Data.gov stores unrestricted federal data on a wide variety of topics, ranging from education to finance to public safety. As of November 2014, over 131,000 data sets were available. The research team restricted queries to public safety, which yielded over 340 data sets. Summary descriptions for each dataset were examined to determine which had potential applications for transportation safety. After filtering the results, a total of 61 data sets were identified. These are listed by name in Appendix A, along with the summary description of the each data set and the links to each individual dataset. In some cases, data sets from specific agencies are found in the list of data sets from Data.gov.

2.2.2 National Highway Traffic Safety Administration

The National Highway Traffic Safety Administration (NHTSA) was established by the Highway Safety Act of 1970 and is dedicated to achieving the highest standards of excellence in motor vehicle and highway safety. The National Center for Statistics and Analysis provides a wide range of analytical and statistical support to NHTSA and the broader highway safety community. In an effort to modernize various systems and databases, a major effort is underway to devise new data collection methods. Users may comment on the future utility of current data elements, recommend additional data elements and attributes, and describe their anticipated data needs. A brief overview of this data system is given in Table 4.

Table 4: National Center for Statistics and Analysis Data

National Automotive Sampling System (NASS): contains Crashworthiness Data System (CDS) and General Estimates System (GES). CDS uses injury mechanisms to improve vehicle design. GES is a national sample of police reported motor vehicle crashes.
National Driver Register (NDR): a database of information on drivers with suspended or revoked licenses or convicted of traffic violations.
Special Crash Investigations (SCI): crash investigation data that includes basic police reports for all crashes and comprehensive reports on selected crashes.
State Data System (SDS): a collection of files from police reports (34 participating states).
Not-in-Traffic Surveillance (NiTS): virtual data collection on non-traffic crashes and non-crash incidents with injuries or fatalities.
Crash Outcome Data Evaluation System (CODES): data which links crash, vehicle, and driver behavior to outcomes of motor vehicle crashes.
Model Minimum Uniform Crash Criteria (MMUCC): recommended data elements for states to include in reporting crash forms and state level databases.
Fatality Analysis Report System (FARS): yearly data on fatal injuries from crashes.
Vehicle Crash Test Database: compilation of data from research, New Car Assessment program results, and compliance crash tests.
Biomechanics Test Database: experimental data used for developing Anthropomorphic Test Devices and associated Injury Criteria.
Component Test Database: engineering data using during various research projects.
Crash Injury Research (CIREN): data on crashes, including reconstruction and medical injuries.

There is also research information for the following areas: Biomechanics and Trauma, Behavioral, Crash Avoidance, Crashworthiness, Driver Simulation (NADS), Enhanced Safety of Vehicles (NADS), Event Data Recorder (EDR), Human Factors, Child Seat Research, Public Meetings, and Vehicle Research and Testing.

2.2.3 Research Data Exchange

RDE is a system that promotes the sharing of archived and real-time transportation data derived from multiple sources. RDE provides a variety of data-related services that support the development, testing, and demonstration of multi-modal transportation mobility applications being developed under the USDOT ITS Dynamic Mobility Applications (DMA) Program. Also included are other research activities focused on connected vehicles. Data accessible through the RDE includes documentation and is freely available to the public. The DMA Program strives to enhance current operational practices and transform the management of future transportation systems through the acquisition and provision of integrated data from infrastructure, vehicles, and travelers.

Basic information, including the list of data environments, is available at: https://www.its-rde.net/home. Eleven data sets collected from different location across the U.S. are available from this database (Figure 1).

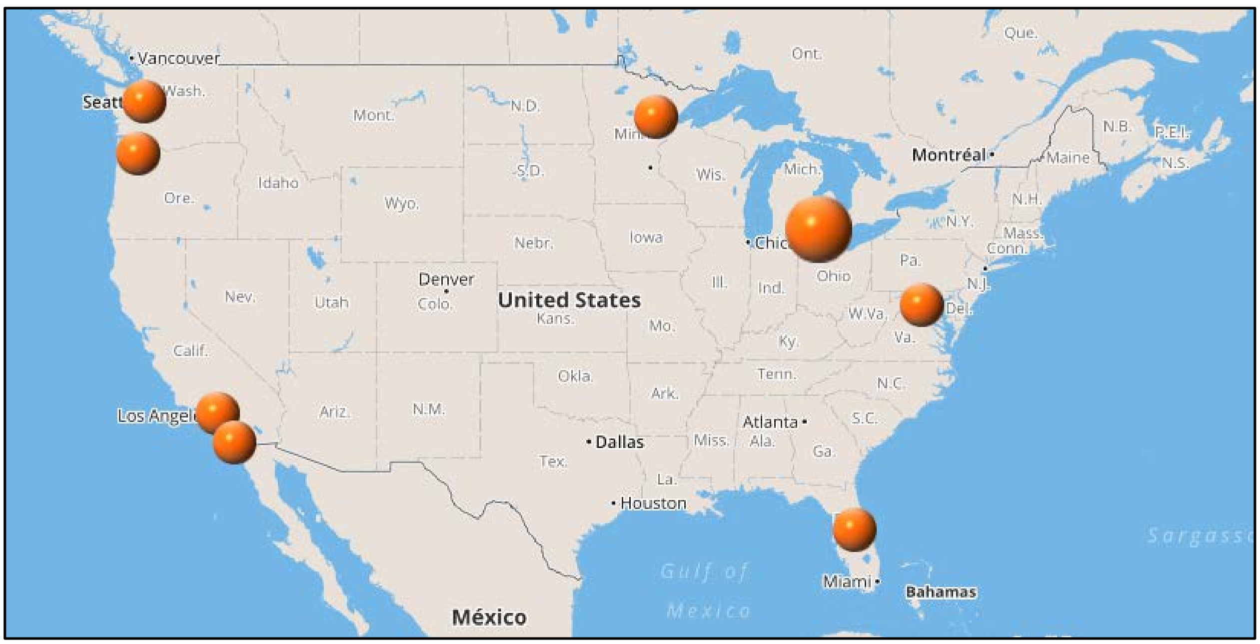


Figure 1: Location of Collected Data

Available data include: GPS data, vehicle trajectory, speed, acceleration data, connected vehicle Probe Data Message (PDM), and Basic Safety Message (BSM) information. Below, there is a brief description of each data set, which are split into two categories: (1) data related to connected vehicles, and (2) traditional traffic data. Further information regarding the data is available in Appendix B.

2.2.3.1 Connected Vehicle Initiative

Six sets of data are included in this category. They are listed below with brief descriptions.

(a) Vehicle Infrastructure Initiative Proof of Concept

The Vehicle Infrastructure Initiative Proof of Concept (POC) contains data on the first major set of trials conducted at the Michigan Test Bed in 2008. The POC trials featured fifty-two roadside equipment (RSE) points within 45 square miles, 27 vehicles configured with onboard equipment (OBE), and a Dedicated Short-Range Communications (DSRC) network. The testing program had three major phases: subsystem test, system integration and test, and public and private applications test. The data consist of RSE and OBE data collected over six days. These six days were selected because the first and last days had much higher number of duplicate records and questionable data values. OBE trajectories data contained speed, longitude, latitude, and time stamps with individual ids (OBE_ID), while RSE data contained speed, longitude, latitude, elevation, heading, engine status, and time stamp with individual ids (RSE_ID).

(b) National Center for Atmospheric Research 2009

The National Center for Atmospheric Research (NCAR) houses data from a second set of trials that were conducted at the Michigan Test Bed to validate the Vehicle Infrastructure Initiative Proof of Concept. These trials used fewer vehicles and focused on collecting data during rainy or snowy weather. The data consisted of RSE and OBE data for the six days with the best data.

(c) National Center for Atmospheric Research 2010

The National Center for Atmospheric Research conducted a third set of trials similar to the 2009 trials at the Michigan Test Bed. In this case, the tests compared atmospheric data from vehicle-mounted sensors to data from a nearby fixed weather observing station.

(d) Safety Pilot Model Deployment - One Day Sample

There were three objectives of the Safety Pilot Model Deployment (SPMD): 1) exploration of the real-world effectiveness of connected vehicle safety applications in multi-modal driving conditions, 2) evaluation of how drivers adapt to the use of this connected vehicle technology, and 3) safety benefits of the connected vehicle technology. The SPMD — One-Day Sample data from April 11, 2013 contained sanitized mobility data elements that were collected from over 2,700 vehicles equipped with connected vehicle technologies traversing Ann Arbor, Michigan. The mobility data were intended to support continued advancements in the connected vehicle domain as well as continue the development of applications to support improved transportation operation. Figures 2 and 3 (Booz et al. 2014) represent the SPMD's content and framework.

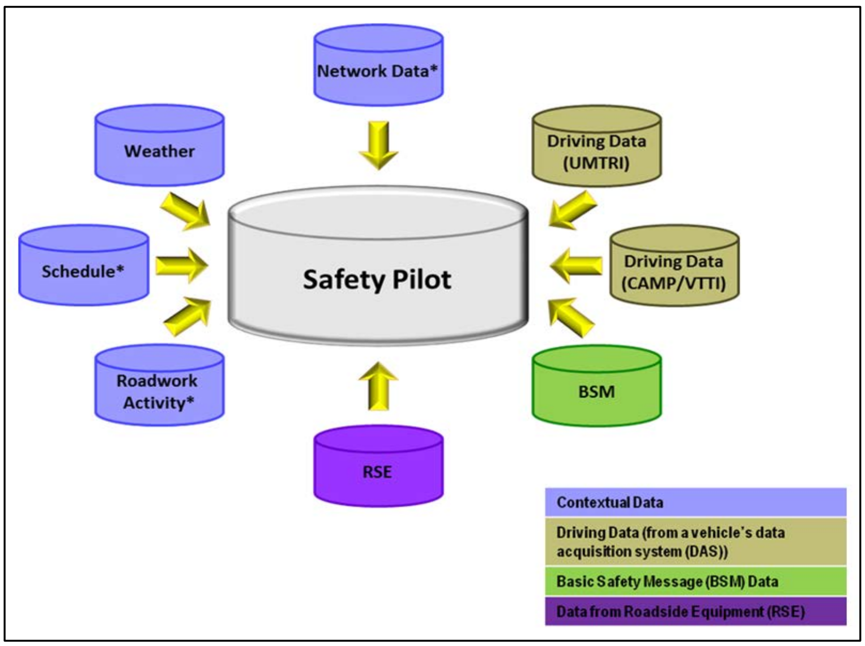


Figure 2: Content of Safety Pilot Data Environment

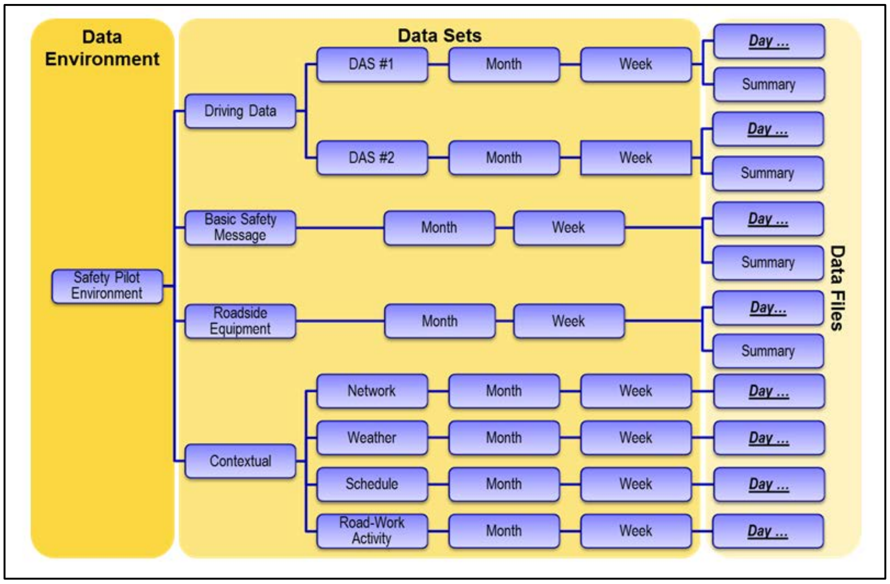


Figure 3: Generalized Data Framework of SPMD Data Environment

Error! Not a valid bookmark self-reference. presents each dataset and a list of the accompanying files from the SPMD.

Table 5: List of Datasets

SPMD Environment						
Driving Data				Contextual		
DAS1	**DAS2**	**Basic Safety Message**	**Roadside Equipment**	**Weather**	**Network**	**Schedule**
AudioTimes	HV_Radar	BrakeByte1Events	BSM	Weather/ climatic data	Pointer to Resources	Pointer to Resources
DataDas	HV_Primary	BrakeByte2Events	Geometry			
DataFrontTargets	RV_Rx	BsmMD	Lane			
DataGpsDas	DAS2_Trip_Summary	BsmP1	LaneNode			
DataLane		ExteriorLightsEvents	MAP			
DataRv		PosAccurByte1Events	Packet			
DataTc		PosAccurByte2Events	PCAPFile			
DataWsu		PosAccurByte3Events	SPAT			
DAS1_Trip_Summary		PosAccurByte4Events	SPATMovement			
TrnBytes		SteerAngleEvents	TIM			
		ThrottlePositionEvents	TIMRegion			
		TransStateEvents	TIMRegionNode			
		WiperStatusFrontEvents				
		BSM_Trip_Summary				

(e) Leesburg, Virginia Vehicle Awareness Device

A Vehicle Awareness Device (VAD) was installed on a single test vehicle over a two-month period to generate continuous data. Data collection occurred during numerous trips around Leesburg, VA, and on one long road trip from Ann Arbor, MI, to Leesburg, VA, by way of eastern Indiana. The VAD installed in the test car was identical to the VADs installed in about 2,800 vehicles that participated in the Safety Pilot Model Demonstration in Ann Arbor, MI. The data set provided researchers an early sample of the more extensive data collected as part of the Safety Pilot Model Deployment. All of the Basic Safety Messages produced by the originator's vehicle were recorded.

(f) Minnesota Department of Transportation Mobile Observation Data

Mobile observation data were collected by the Minnesota DOT's maintenance vehicles enrolled in the Minnesota Integrated Mobile Observation (IMO) project. Data were derived from mobile, vehicle-based observations of road conditions; included were vehicle engine status and weather conditions. Transmission via cellular media took place in approximately real-time from vehicles to the Minnesota DOT. This method of Vehicle-to-Infrastructure (V2I) reporting may be an important component of the connected vehicle program.

2.2.3.2 Traditional Traffic Data

Five data sets are included in this category. They are listed below with brief descriptions.

(a) Florida Department of Transportation Orlando ITS World Congress

The Florida Department of Transportation (FDOT) recorded data using Vehicle Awareness Devices (VADs) installed on Lynx transit buses in Orlando, Florida. The VADs became operational in September 2011 and continued operation during the ITS World Congress in October 2011. Recorded data included the required components of the J2735 Basic Safety Message (BSM). This data collection tested the capability of Vehicle Awareness Devices to capture and store data in the form of the J2735 Basic Safety Message. This served as a prototype for larger scale tests, such as the Basic Safety Model Deployment. Available data contained a time stamp, longitude (deg), latitude (deg), elevation (m), heading (deg) and speed.

(b) San Diego, California

The San Diego data environment contained one year of raw and cleaned data for over 3,000 traffic detectors that were deployed along 1,250 lane miles of I-5 in San Diego. The recorders captured and aggregated data at varying intervals, including 30-second raw reports and 5-minute, hourly, and daily aggregations. Each point in the data set contained time stamp, average speed, average occupancy, total flow, and direction with individual station ID. The data set also included the average speed, occupancy, and flow of each lane.

Additionally, this data set contained georeferenced data for over 1,500 incidents and lane closures for the two sections of I-5 that experienced the most incidents during 2010. This freeway incident file contained a number of fields, including individual incident identification, start time, duration, location, ALK grid, link identification, and a description of the incident. Complete trip (origin-to-destination) GPS "breadcrumbs" contained latitude/longitude, vehicle heading and speed data, and time for individual in-vehicles devices that were updated at 3-second intervals for 10,000 trips taken during 2010. Weather data were collected from seven weather stations in the San Diego area. Observation points held time stamps, temperature, visibility, wind direction, wind speed, snow (inches), precipitation, and relative humidity.

(c) Pasadena, California

The Pasadena data covered the roadway network in and around the City of Pasadena, California. Data were collected during September and October in 2011. The data environment included network data (highway network file), demand data (trip tables), network performance data (link volumes, turn volumes, speeds and capacity), work zone data, weather data, incident data, Closed Circuit Television (CCTV) camera data, and Changeable Message Sign (CMS) data. Data from simulations were included where there were no sensors, and to provide forecasts. Data were documented in three different formats: Structured Query Logic (SQL) format, plain text format, and VISUM format.

(d) Seattle, Washington

From May 2011 to October 2011, extensive data were collected in the city of Seattle. This included raw and cleaned data from traffic detectors deployed by the Washington Department of Transportation (WSDOT) along I-5 in Seattle. These data included 20-second raw reports and 5-minute aggregations. For very detailed freeway performance evaluations, the 20-second data set provided freeway performance data for individual loop detectors. The 20-second data were divided among four different data tables. These data tables described the location of the cabinets that contained the loop electronics, each specific loop, and correction factors developed to account for loop sensitivity issues. WSDOT aggregated 5-minute data at its traffic management center and reported them independently from data collected at 20-second intervals. This level of aggregation provided sufficient detail to identify the onset of congestion, yet limited the amount of data handling required to develop those performance measures. The 5-minute data was split into three tables; Cabinets, Loops, and Loop Data. The Loop Data table contained loop identification, time stamp, location (longitude/latitude), volume, and occupancy data.

The City of Seattle and WSDOT have placed numerous automatic license plate readers (ALPRs) at intersections around the city and on some state routes. Matching license plate reads from ALPRs at different intersections enabled the direct calculation of arterial travel times from one intersection to another by subtracting the time of passage at the upstream location from the time of passage at the downstream location. The ALPR data were stored in two related tables. The first table listed the locations of ALPR cameras. The second table provided specific travel time segment information, such as unique identifier date, time, travel time, direction, and failed pass to check data quality.

Incident data records came from the WSDOT's Washington Incident Tracking System (WITS). These data were collected from the reports of incident response teams. It contained individual incident identifier, incident type, location (milepost), incident notification time, arrival time, and the amount of time needed to clear the event. GPS breadcrumb data from commercial trucks was not available to the public because of data ownership and privacy issues.

(e) Portland, Oregon

The Portland data included several different data sets. There were freeway data, which consisted of two months of data gathered from dual-loop detectors stationed on the main line and on-ramps of I-205, including speed, volume, direction, occupancy, travel time, and delays aggregated over different time intervals — 20 seconds, 5 minutes, 15 minutes, and 1 hour. All data were in the same format

Incident data from the Oregon Department of Transportation's (ODOT) Advanced Traffic Management System database and planned event data from the ODOT Trip-Check Traveler Information Portal information web site were available. These data sets contained individual incident identifiers, latitude, longitude, location information, incident start time and end time, duration, and last update time. Weather data were collected from two sources: National Oceanic and Atmospheric Administration (NOAA) data and Remote Weather Information System (RWIS) stations. These data contained station id, report time, temperature, wind speed, precipitation and humidity.

Three types of arterial data were collected: (1) volume and occupancy data from four single-loop detectors on 82nd Avenue, (2) signal phase and timing data for 32 signals along the 82nd Avenue corridor, (3) travel times on 82nd Avenue (calculated using data collected by two Bluetooth readers). Bluetooth data contained location (longitude and latitude), time stamp, and individual travel times.

Arterial detectors, intersections, and stations data files contained the location (longitude and latitude) of each detector, intersection and station, identification, and description. Arterial, signal phase, and timing data contained signal phasing information and the time stamp of different intersections. Transit data were collected by TriMet, the Portland-metro area transit agency, including schedule, stop event and passenger counts data for both bus and light rail.

2.2.4 TRIMARC

Traffic Management Centers (TMC) around the country house archived traffic monitoring data. TRIMARC is a local example, a regional TMC in the Louisville metropolitan area. It is an Intelligent Transportation System (ITS) designed to improve the performance of the freeway system in the metropolitan Louisville and Southern Indiana area. TRIMARC was designed with ITS in mind to improve the freeway system's performance in metropolitan Louisville and Southern Indiana.

TRIMARC sensors collect data on volume, speed, and lane occupancy at 15-minute intervals at each detector section. Data come in 30-second slots, after which the TRIMARC server aggregates them for each 15-minute period and provides sample size information that how many 30-seconds slot have been aggregated. This offered a convenient way to check data quality. These data also contained incident data, encompassing incident location, direction, time, types of incident, and its duration. Figure 4 is a map of TRIMARC's coverage.

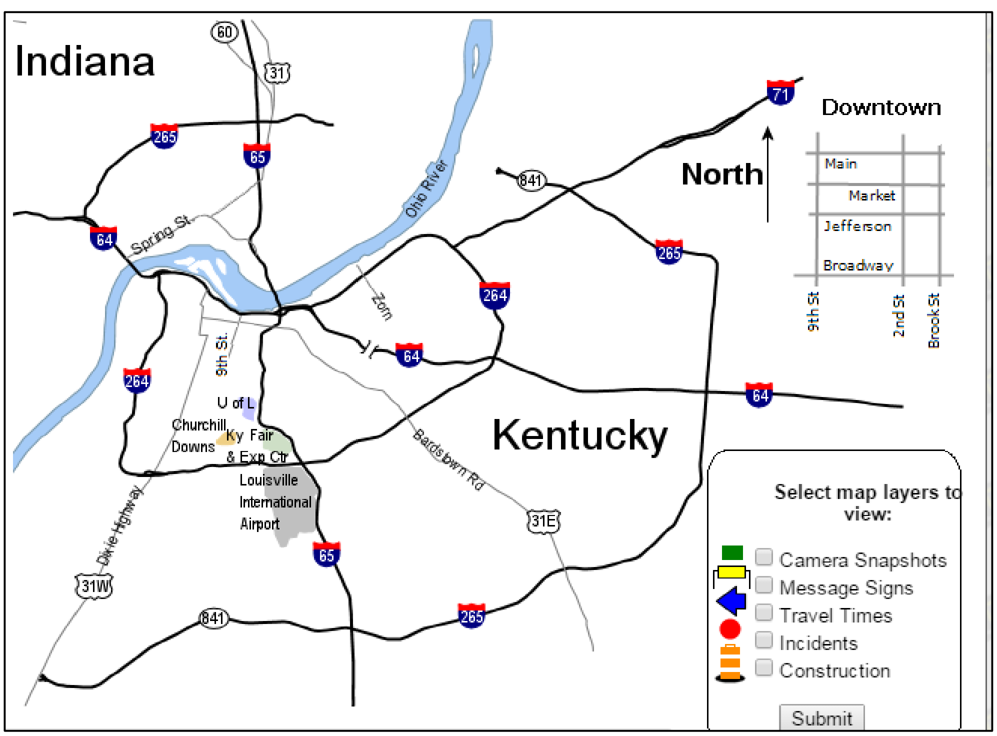


Figure 4: TRIMARC Interactive Map

3. Application of Integrating Big Data

3.1 Introduction

Highway incidents are a significant issue for travelers in the United States. According to the FHWA, traffic incidents are one of the two major causes of traffic congestion in the United States. Among traffic incidents, crashes are the origin of abnormally heavy traffic congestion. Crashes also diminish the reliability of transportation networks.

Crashes and associated traffic congestion are a major concern to the public. To measure highway system performance, average travel time has been used as a principal indicator. But because of the highly dynamic nature of traffic, average conditions do not capture the entire story. Travel time reliability was recently proposed as a measure of the variability and degree of congestion. This concept has gained traction among researchers and practitioners, because it accurately compares an individual driver's experience against average travel time. Many factors (e.g., road geometry, demand, capacity, weather, etc.) impact the travel time reliability of freeway facilities. Crashes affect travel time reliability because they reduce the capacity of a roadway segment, create a temporary bottleneck, and thus significantly delay travel. Very few studies have comprehensively examined the impact of traffic crashes on freeway travel time reliability.

The incident data used for this study is rich with information; it includes incident type, location, start and end time, duration, and lane/shoulder block information. The primary objective of the research was to visualize the impact of crashes and investigate how factors influence the overall consequences. The research team examined the relationship between accidents and reliability on freeway routes. To accomplish these objectives, statistical analyses were conducted using one year of speed, travel time, and incident data collected on I-65 northbound and southbound in the Louisville, Kentucky, metropolitan area.

The next section recent academic work and describes the study sites, data sources, visualization tool and influencing factor analysis, existing travel time reliability measures, and the methods used to derive them. The concluding sections describe the analysis, summarize the findings, and make recommendations based on the study.

3.2 Literature review

Transportation researchers have recently turned their attention toward travel time reliability. Using traffic crash and empirical traffic flow data collected from the Netherlands, Tu et al. (2012) presented an empirical travel time reliability analysis. One limitation in their research was that the duration and severity of each accident were unknown, so they assumed each accident had a duration of three hours. Yu et al. (2013) used reliability analysis to assess freeway crash risks and to evaluate hazardous freeway segments. Reliability analysis accomplishes this by integrating traffic flow parameters and real-time crash occurrence risk at the disaggregate level with weather parameters. Yu et al. (2013) found this method provided more accurate crash predictions than logistic regression. Zhong et al. (2011) used data on rural roads in Wyoming to model and predict crashes. The data they used included accident records, traffic volume, speed, and other factors, from 36 roads over a 10-year period. Negative binomial regression and Poisson regression were used to examine the causes of rural crashes. Multiple regression approaches have attempted to analyze the relationship between crash rates and geometric roadway features. However, multiple studies have found linear regressions are unsuitable (Miaou et al. 1993; Okamoto and Koshi, 1989). Zhong et al. (2011) demonstrated that roads with higher speeds and traffic volumes elevated crash rates at certain higher risk locations. Wright et al. (2015) showed that incidents produce higher values in all reliability measures. They also examined how incidents affect the probability of traffic congestion on freeway segments. Compared to the normal condition, they found that shoulder incidents significantly increased the probability of freeway segment traffic breakdown, while incidents spread across multiple lanes resulted in the most significant increases in travel time variability and in the buffer index.

Few studies have looked at the interactive effects of traffic and weather factors and roadway geometry on different crash types. Investigations of incident duration and identification of the contributing factors have been scarce, especially in research that uses different data types.

Among these, Yu et al. (2013) attempted to explore the use of microscopic traffic and weather indicators to differentiate between crash types and to analyze the crash type propensity at the micro-level for three major crash types — rear-end, sideswipe, and single-vehicle crashes. Ahmed et al. (2012) investigated the effect of the interaction between roadway geometric features and real-time weather and traffic data on the occurrence of crashes on a mountainous freeway. They found that geometric factors were significant in all seasons. Crash likelihood could double during the snowy season due to slick pavement conditions and steep grades, and when combined, produced a hazardous road surface On the other hand, Hojati et al. (2012) presented a framework to exhaustively mine traffic-incident data and directed subsequent analysis toward an incident delay and travel-time reliability model.

Though there are several proposed models that are highly efficient, they cannot be applied to other cases because different studies call for the use of different variables. As such, results may not be transferable across different locations. Data collection and reporting process have also been incommensurate. While the findings of previous studies will not reduce the number of crashes/incidents, they will reduce their effects.

3.3 Overview of Study Corridors and Data Sets

3.3.1 Description of Study Corridors

This study examines the impacts of crashes along northbound and southbound I-65 in the Louisville metropolitan area. The study segment is 5.6 miles long in the northbound direction and 5.4 miles long in the southbound direction (Figure 5). The data environment from two sources contains one year of data (2012). Data are available for each day, logged at 15-minute intervals.

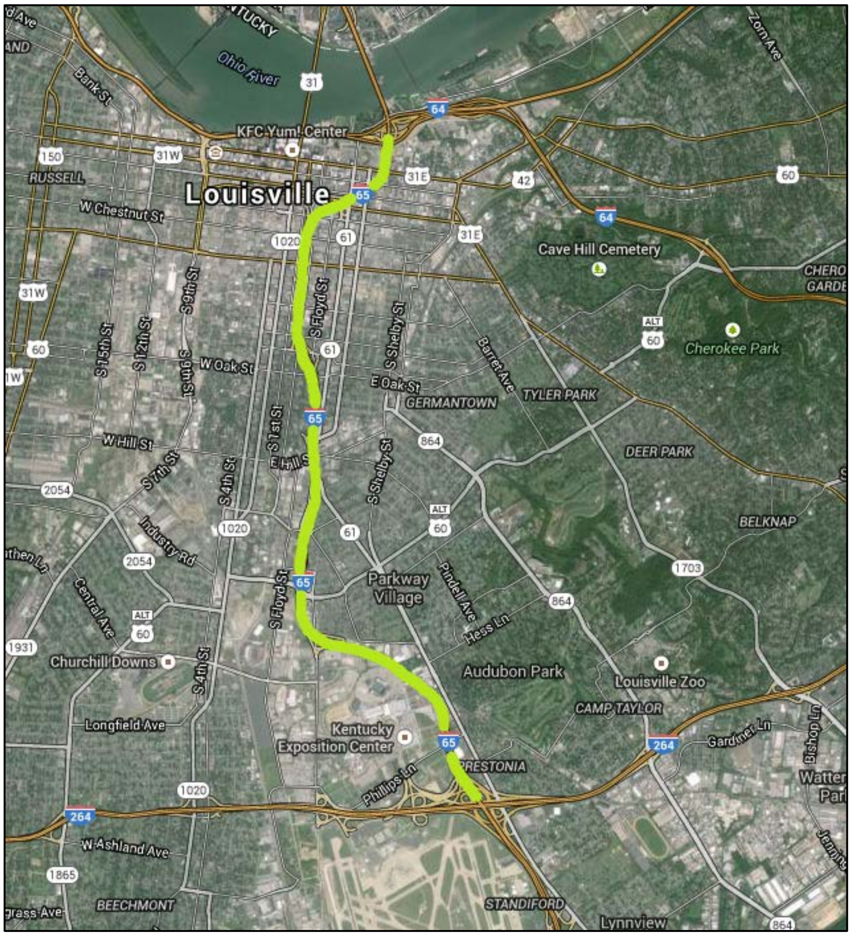


Figure 5: Study Corridor of I-65

3.3.2 TRIMARC Data

TRIMARC is a regional ITS data source designed to improve the performance of the freeway system in metropolitan Louisville, which extends into Southern Indiana. The TRIMARC data environment contains time, speed, volume, and lane occupancy data for each day. This information was recorded at 15-minute intervals along each detector section. Data were originally recorded in 30-second slots. The TRIMARC server aggregated them every 15 minutes. There are 15 TRIMARC sensors located on I-65 N and 11 sensors on I-65 S. The average spacing between two sensors is approximately 0.4 mile. Figure 6 maps the sensor locations on I-65 N.

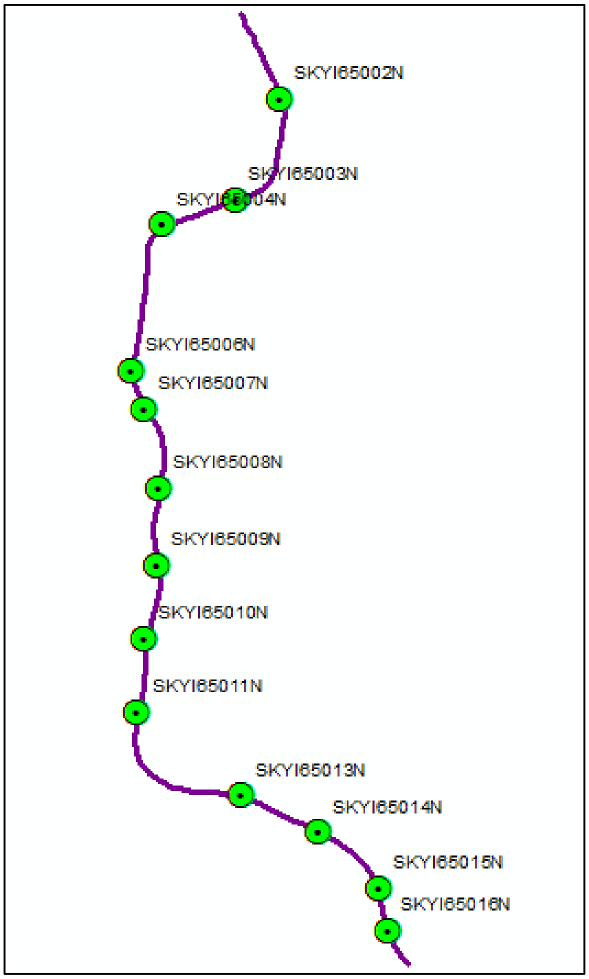


Figure 6: Location of TRIMARC Sensors on I-65N

The TRIMARC data environment contains incident data: the location, direction, time, incident type, duration, and information on blocked lanes. To interpret the effect of crashes, accident crash-affected time slices (15 minutes) are identified based on the duration of an incident.

3.3.3 NAVTEQ Data

NAVTEQ speed information data are found in the smart driver network, which aggregates traffic data from probe vehicles. These data were used only for I-65 N, from milepoint 131.2 to milepoint 136.6. NAVTEQ data are aggregated by weekday for specific months.

3.4 Methodology

3.4.1 Visualization of Traffic Data

To analyze the impact of crashes on traffic, the team built a visualization tool in Microsoft Excel. For TRIMARC data, the model relied on vehicle speed and the location of sensors. The horizontal axis represented the time of the day and the vertical axis showed the distance, or the location of the sensors/length of the segment. A space-time velocity map (also known as a heat map) recorded average speeds on those segments at different times of the day. Using this heat map, the user can input the date and find the traffic speed at different times of day. The heat map also contained a link between the incident data and the TRIMARC data, so that inputting an incident identifier (Incident ID) generated incident information and corresponding traffic speed for that day. Ultimately, this heat map increased visual understanding of the crash impacts.

3.4.2 Influencing Factor Analysis

Travel rates (seconds/mile) were treated as a measure of effectiveness to understand how crashes affect the travel rate. This analysis can guide agencies toward improvements in the operation of road networks. Travel time reliability monitoring systems allow agencies to enhance system reliability. For example, when an agency experiences unreliable travel times because of indents, the agency may increase its spending on incident management systems and safety improvements. Conducting the Influencing Factor Analysis required the following steps:

- Select the region or facilities of interest and study period
- Compile travel rate data for each facility
- Identify what types of nonrecurring events (peak/off-peak accident, different types of accident etc.) are present in the data
- Develop cumulative distribution functions (CDFs) of travel rate (TR) for each combination of nonrecurring events

3.4.3 Travel Time Reliability Measurements

Significant research on travel time reliability has resulted in several proposed reliability measures. There is an ongoing debate over which reliability measure is the most effective. Lomax et al. (2003) reviewed existing travel time reliability measures, and categorized them into four main categories: statistical range measures, buffer measures, tardy-trip indicators and probabilistic measures. In this study, the buffer and planning time indices proposed by Lomax et al. (2003) were used to measure travel time reliability.

Buffer Index is the amount of extra time that a driver should add to their average travel time to ensure an on-time arrival. The Buffer Index (BI) is calculated by finding the difference between the 95th percentile travel rate and the average travel rate and dividing that number by the average travel rate:

$$\mathbf{BI = (TR_{95th} - \mu)/\mu}$$

Where TR_{95th} is the 95th percentile travel rate and μ is the average travel rate. Units for both measures are seconds per mile.

Planning Time Index (PTI) compares the total percentage of time to a driver's free-flow travel time (ensuring a 95th percentile on-time arrival). PTI is calculated by dividing the 95th percentile travel rate by the free flow travel rate:

$$\mathbf{PTI = TR_{95th}/TR_{free\ flow}}$$

Where TR_{95th} is the 95th percentile travel rate and $TR_{free\ flow}$ is the free flow travel rate. Units are both in seconds per mile.

The **Travel Time Index (TTI)** was also included in this study because it is a key criterion for mobility analysis. It represents the average time a driver would take to complete a trip during an incident, compared to free-flow conditions. It is calculated as the ratio of average travel rate across the entire year to travel rate at free-flow condition.

$$\mathbf{TTI = \mu \,/\, TR_{free\ flow}}$$

Where μ is the average travel rate and $TR_{free\ flow}$ is the free flow travel rate, with both in seconds per mile.

Travel Time Variability (TTV) is the difference between the 90th and 10th percentile travel rate in different traffic condition:

$$\mathbf{TTV = TR_{90th} - TR_{10th}}$$

Where TR_{90th} is the 90th percentile travel rate and TR_{10th} is the 10th percentile travel rate.

3.5 Data Analysis and Results

3.5.1 Space-Time Velocity Map/ Heat Map

A space-time velocity map visualizes traffic data during a defined space-time window; it is also referred to as a heat (or color) map. Heat maps clarify actual conditions rather than focusing on numerical values. Figure 7 includes three panels:

- 1st panel shows the incident information — when and where incident happened, how long it lasted, the type of incident, condition, lane blocked information, and other information.
- 2nd panel shows the TRIMARC's heat map
- 3rd panel shows the travel time information at different time of the day

A narrative description of Figure 7 would read as follows: *On February 20, 2012, at 1:08 pm an accident occurred at milepoint 134.8. The crash blocked one lane of the freeway, and the accident zone was cleared at 2:19 pm. Traffic was interrupted for 71 minutes.*

Placing the incident identifier of the crash in our Excel framework, we created a space-time velocity map (heat map) of that day for TRIMARC data. Readers can easily spot the incident that took place at the sensor (MP 134.6); this accident affected times at the sensor both temporally and spatially. There are five horizontal red boxes (each red box equals 15 minutes), meaning the crash's impact lasted approximately 75 minutes where it occurred. It also affected the four immediate upstream sensors. In this region, a significant decline in vehicle speeds occurred. When the crash cleared, the traffic resumed normal flow. Based on the travel time information, it is evident the impact of the crash starts and ends at just about the same time as the crash's start and end time. After clearing the accident, travel time returned to normal condition.

24206	Incident ID	
2/20/2012	Start Date	#
Monday	Week day	M
	Start Time	1:08 PM
	End Time	2:19 PM
	Total (min)	71
	HWY	I-65
	MP	134.8
	Est. Clear	1 Hours
	Incident Type	Accident
	Lanes Blocked	1 Lane(s) Blocked
	Notes	The left lane is blocked. May be viewed on TRIMARC camera 3.
	Condition	Dry Pavement, Sunny, Possible Injury, Rear-End Collision, Vehicle Damage, Car, Van, SUV

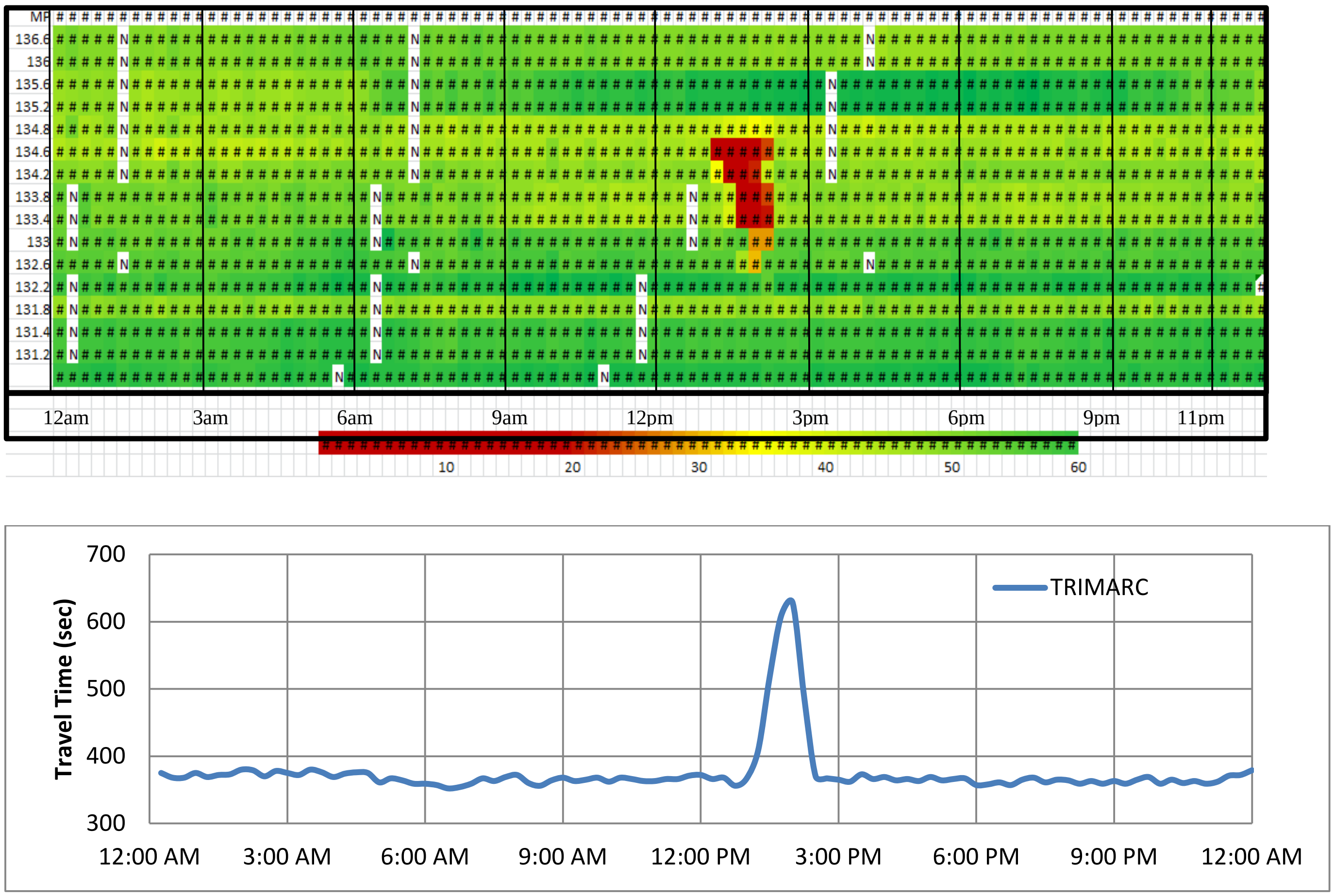


Figure 7: Visualization Tool (Heat Map)

Later, we developed a trajectory diagram (Figure 8) based on the TRIMARC data to illustrate how the corridor was impacted during the crash-affected period. It reveals how the shockwave propagated along the corridor during the crash period.

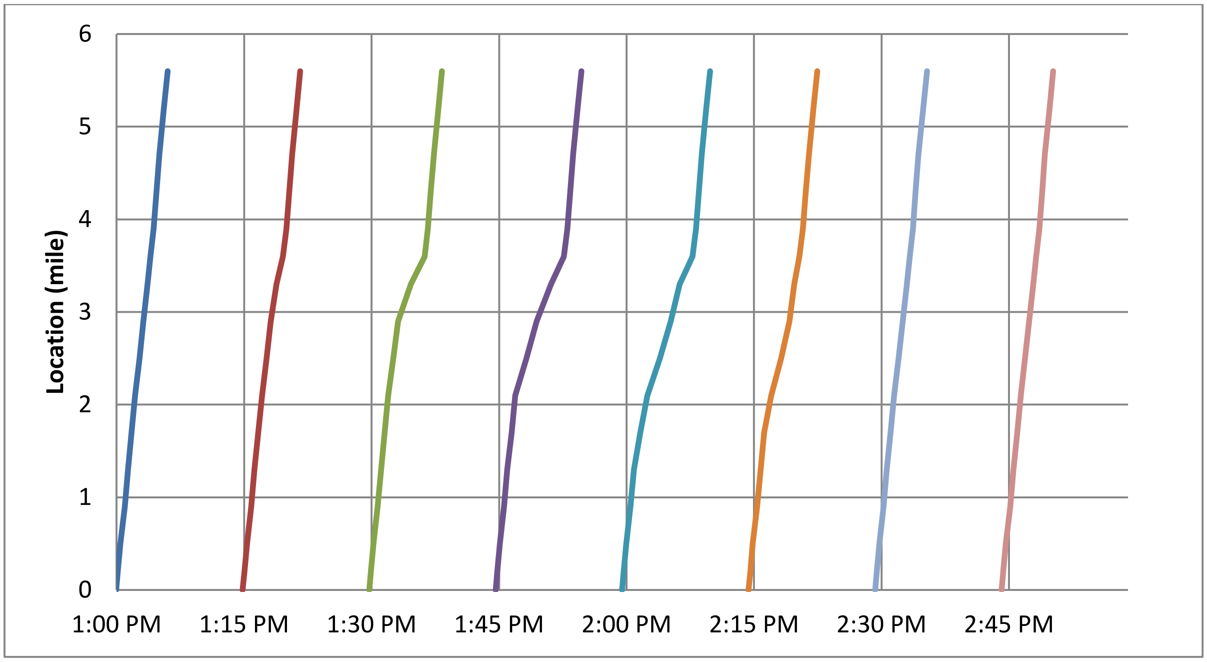


Figure 8: Trajectory Diagram

We also created a heat map framework for NAVTEQ data to visualize the crash's impact. Given that NAVTEQ data are aggregated weekly for a specific month, we aggregated our TRIMARC data in NAVTEQ format and built another space-time velocity map for TRIMARC. This map displays every day's traffic speed and its weekly aggregation level. We then connected that heat map with incident data to analyze every accident just by putting in incident identifier. For example; Figure 9, 10, 11, and 12 are traffic speeds for the individual day (Friday) in October. Figure 13 aggregates all Fridays in October based on the TRIMARC data.

Figure 14 is the heat map for NAVTEQ data, which aggregates October Fridays. Analyzing traffic for each Friday in October showed that crashes happened on October 26 (Figure 12). Both heat maps (TRIMARC data in Figure 13 and NAVTEQ data in Figure 14) reveal the impact of those crashes on weekly aggregated speed at the incident sites. Heat maps enhance our visual understanding of the impact of crashes. Visual intuition is preferable to examining numerical information.

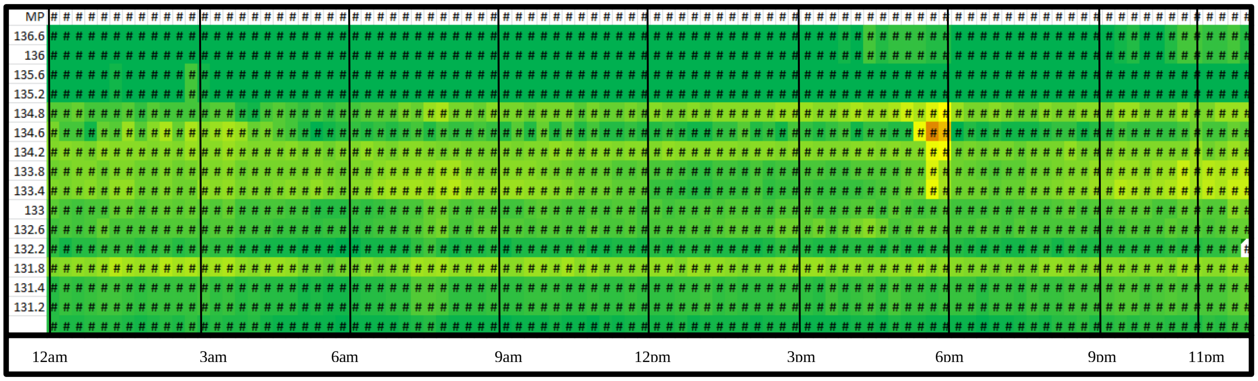


Figure 9: October 5, 2012; Friday- No crashes

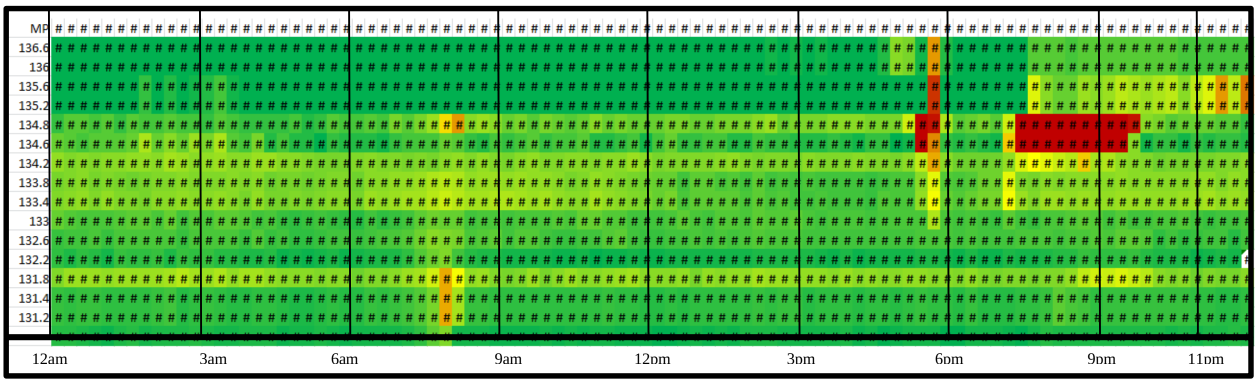


Figure 10: October 12, 2012; Friday- No crashes

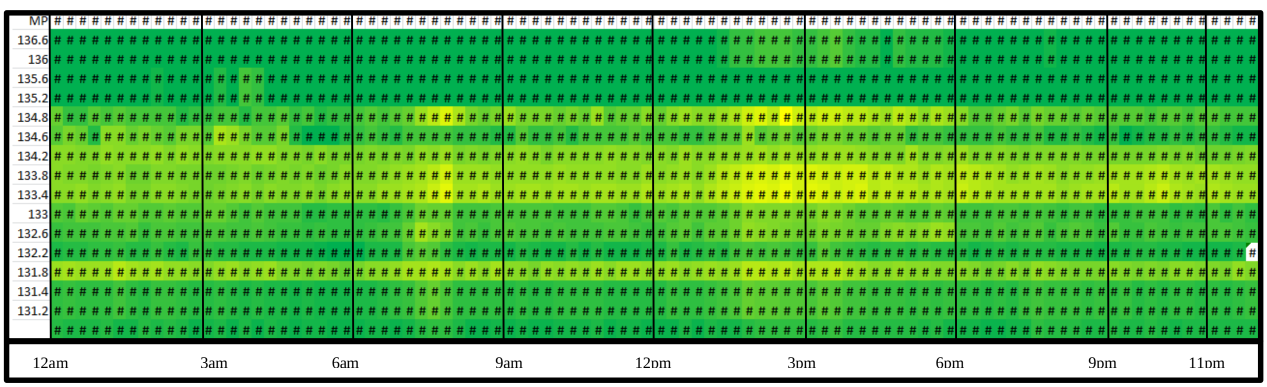


Figure 11: October 19, 2012; Friday- No crashes

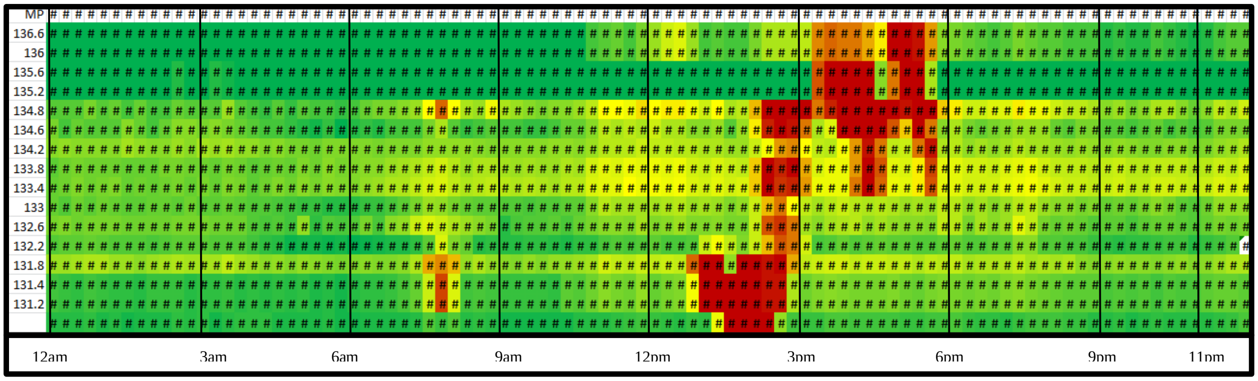


Figure 12: October 26, 2012; Friday- Crash

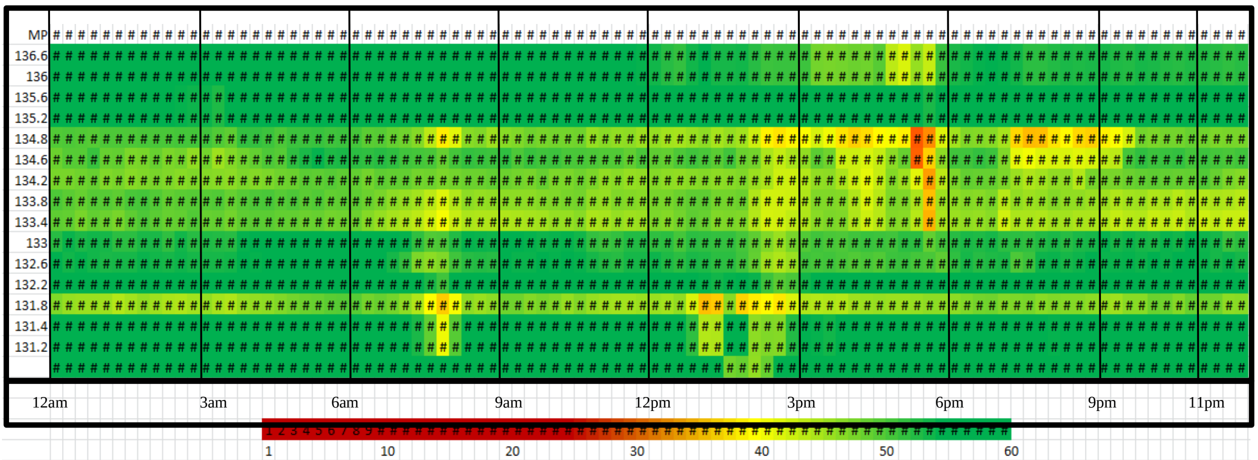


Figure 13: October_Friday (TRIMARC)

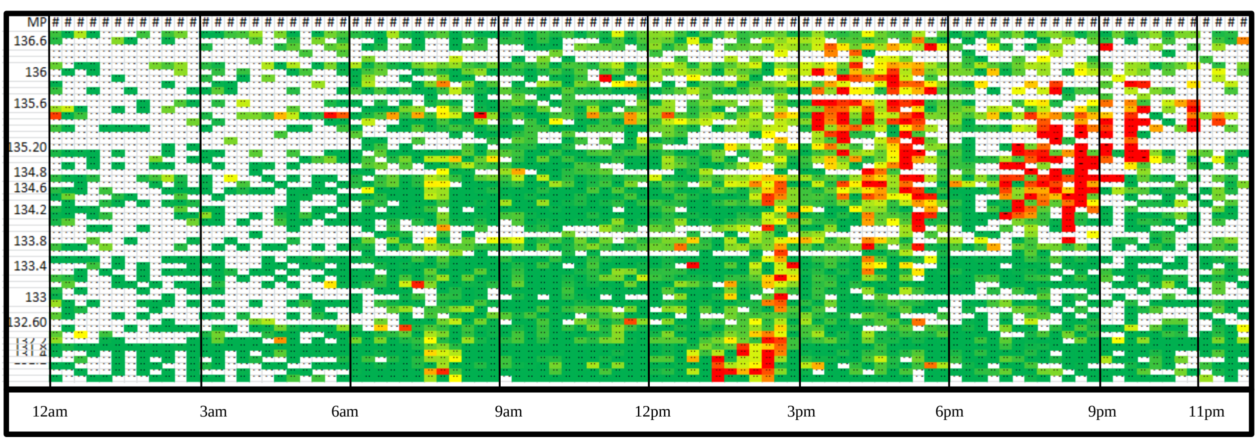


Figure 14: October_Friday (NAVTEQ)

3.5.2 Influencing Factor Analysis:

Next, we analyzed the impact of crashes on I-65 N and S. Travel rate is considered as a measure of effectiveness to understand how a crash influences traffic flow.

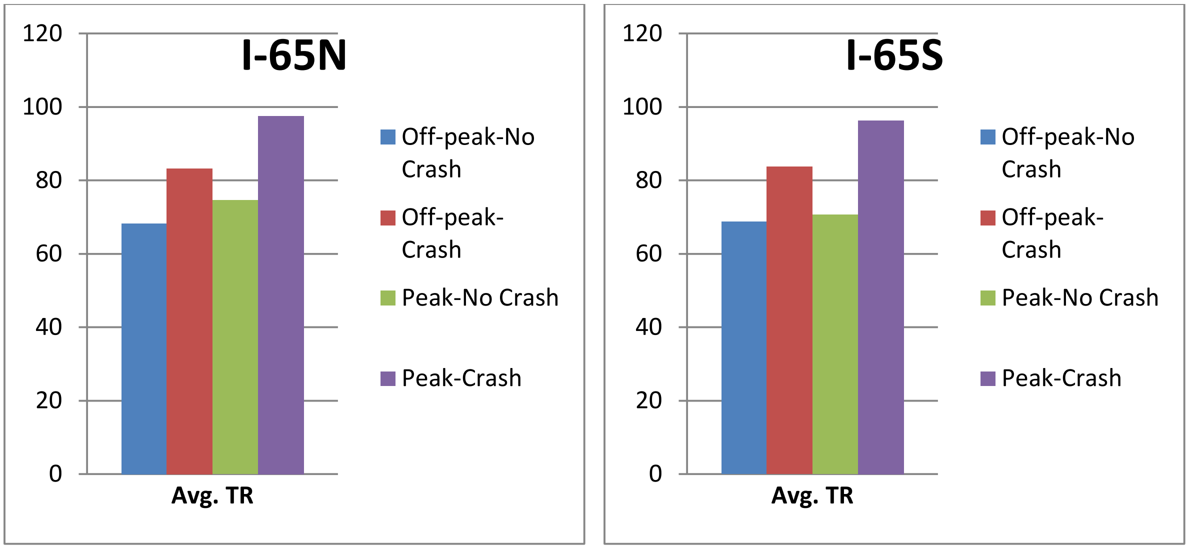


Figure 15: Average Travel Rate (sec/mile)

First, we calculated the 2012 daily travel rate (sec/mile) (based on TRIMARC speed data of 15-minute intervals). Travel rates were then separated into two groups: peak period and off-peak period travel rate. The morning and afternoon peak periods occurred between 6 AM and 9 AM and 4 PM and 7 PM, respectively. Our analysis was based on the crashes in the study segment. Peak travel rates were grouped into two classes: one containing the 15- minute travel rates that were influenced by crashes (termed “peak, crash”) and the other travel rates during periods that were unaffected by crashes (“peak, no crash”). Time slices impacted by crashes were selected based on incident duration. Off-peak travel rates were also separated into two groups: off-peak, no crash, and off-peak crash. As Figure 15 indicates, during a crash, the travel rate increases significantly over to the no-crash condition. Crashes during peak periods increase travel rates (which means a higher delay) for both directions on the interstate.

Figures 16 and 17 depict a CDF of travel rate; they tell a better story about route performance. Consider a travel rate of 80 sec/mile. According I-65N (Figure 16), more than 95 percent of vehicles could travel at 80 sec/mile during an off-peak hour no crash situation, while 83 percent of vehicles travel at this rate during peak hours when there are no crashes. However, if a crash occurred during an off-peak period, only 65 percent of vehicles could travel at 80 sec/mile. If a crash happened during the peak period, the situation worsened. Only 43 percent of vehicles could achieve that travel rate. Similar trends were observed for I-65 S (Figure 17). Table 6 summarizes this information for a travel rate of 80 sec/mile.

Table 6: Comparison Chart

Route	Off-Peak, No Crash	Peak, No Crash.	Off-Peak, Crash	Peak, Crash
I-65N_TRIMARC	96%	83%	65%	43%
I-65S_TRIMARC	97%	90%	69%	44%

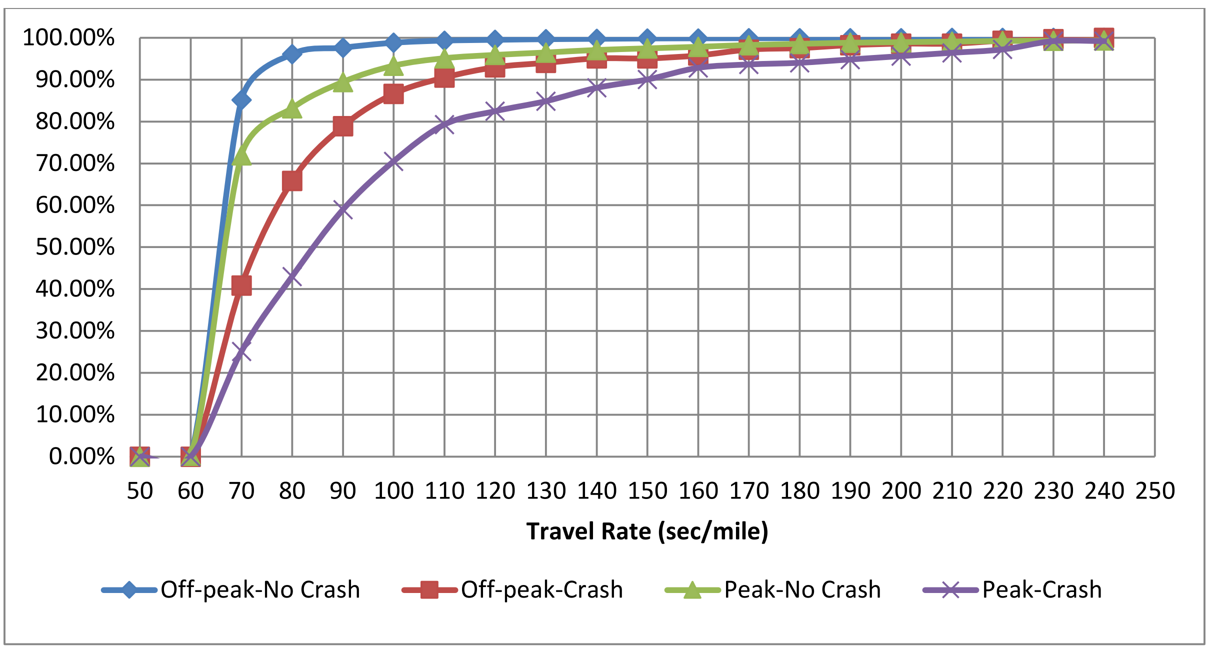


Figure 16: CDF of Travel Rates for I-65N

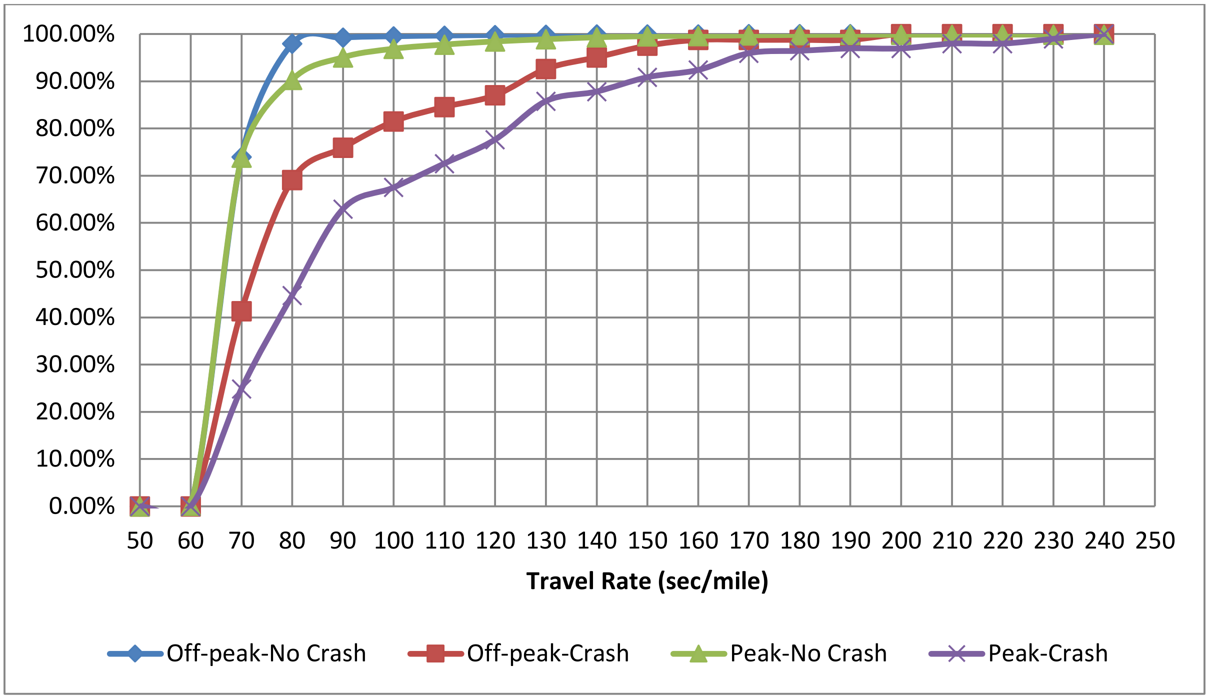


Figure 17: CDF of Travel Rates for I-65S

Next, we analyzed travel rate information for each route under four scenarios to determine the impacts of each on route travel time. These four conditions were:

1. No crashes: normal condition with no crashes
2. Crash with one lane blocked: crashes with a single lane blocked scenario with and without the shoulder blocked
3. Crash with multiple lanes blocked: crashes with multiple lanes blocked scenario with and without the shoulder blocked
4. Crash with only the shoulder and/or zero lanes blocked : crashes that block only the shoulder and/or zero lanes.

We drew a CDF of travel rate for the four scenarios to analyze the effect of different incident types. Taking a travel rate of 80 sec/mile as a baseline, along I-65N_TRIMARC (Figure 18), about 92 percent of vehicles could travel at 80 sec/mile during when there were no crashes. However, when a crash occurred, just 58 percent of vehicles maintained this travel rate when a single lane was blocked. Only 54 percent of vehicles moved at this rate under scenarios when just the shoulder was blocked. When multiple lanes were blocked due to a crash, the situation worsened: only 44 percent of vehicles traveled at 80 sec/mile travel rate. Similar trends were evident on I-65 S (Figure 19). Table 7 compares information for a travel rate of 80 sec/mile.

Table 7: Comparison Chart

Route	No Crash	Single-Lane Crash	Shoulder Crash	Multi-Lane Crash
I-65N_TRIMARC	92%	58%	54%	44%
I-65S_TRIMARC	96%	58%	52%	52%

However, there was an exception on I-65 S (Figure 19). In this data set, all three crash types had a comparable impact on travel rate. Although somewhat counterintuitive, incidents that blocked a single lane or just the shoulder produced slowdowns commensurate with crashes that closed multiple lanes.

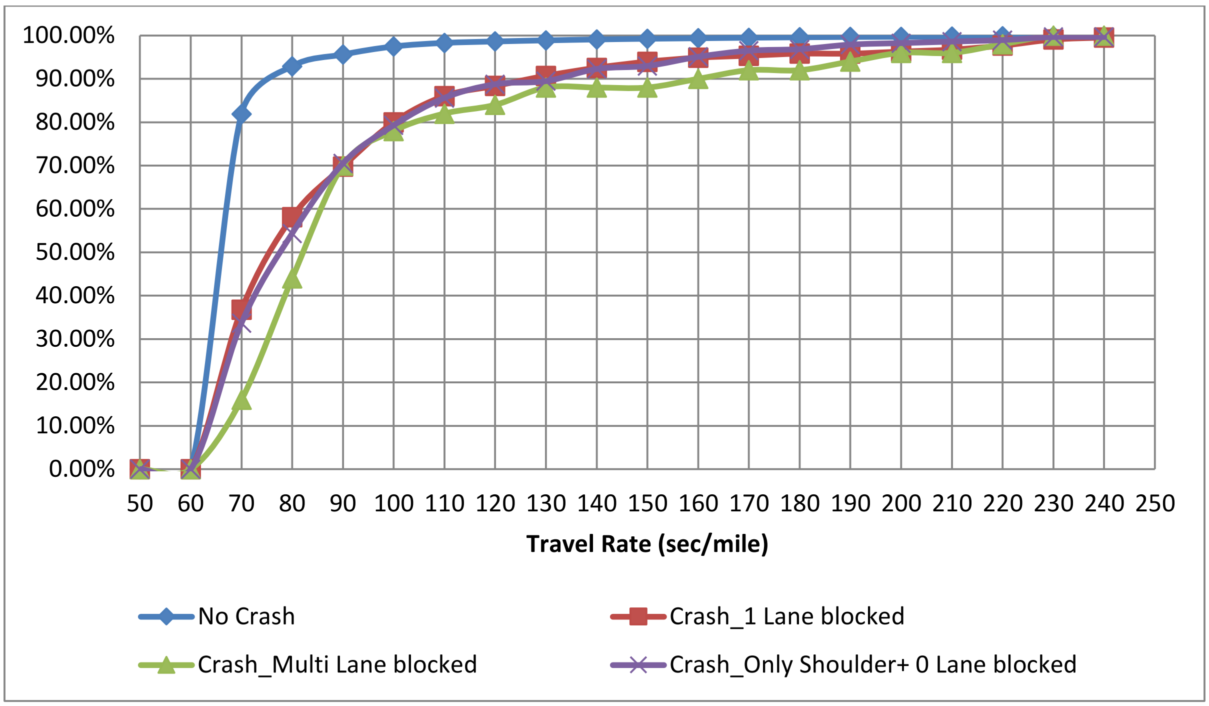


Figure 18: CDF of Travel Rates for I-65N

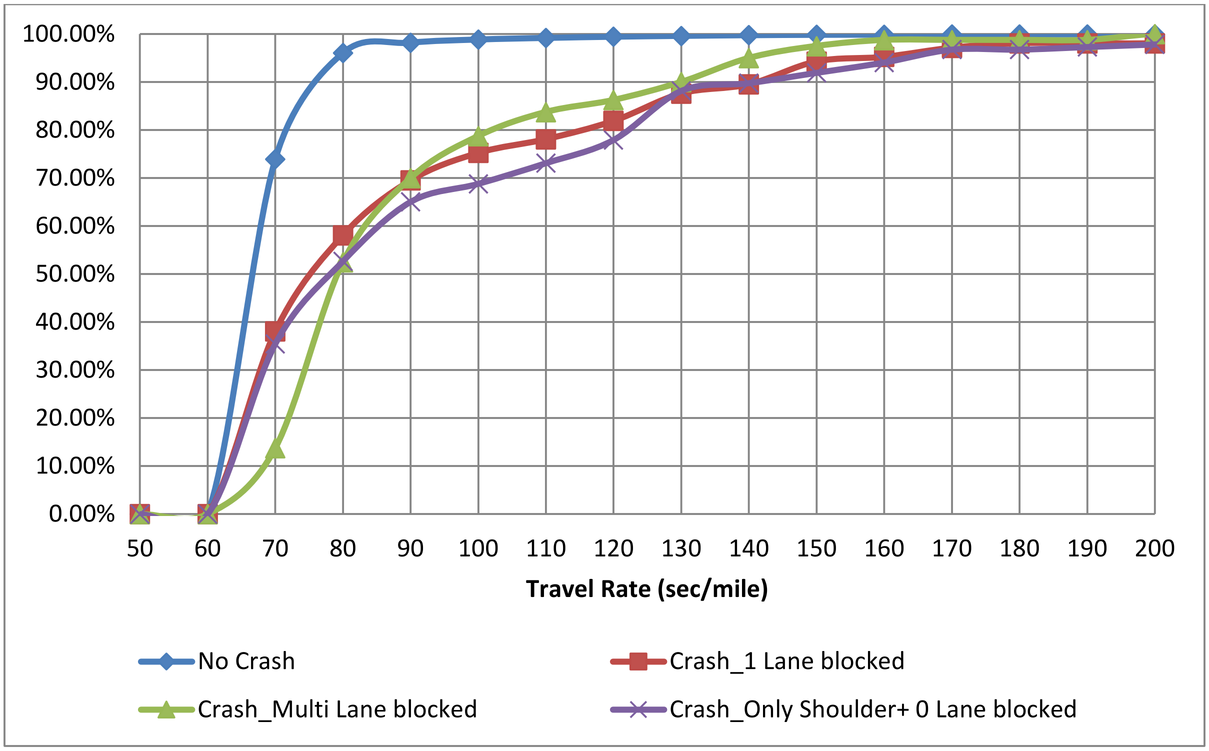


Figure 19: CDF of Travel Rates for I-65S

3.5.3 Reliability Measurements

Travel Time Reliability measurements were calculated, including Travel Time Variability (TTV) (Figure 20), Buffer Index (BI) (Figure 21), Planning Time Index (PTI) (Figure 22), and Travel Time Index (TTI) (Figure 23). Each index was calculated for the four scenarios noted above: no crash, crash blocking one lane, crash blocking multiple lanes, and crash that blocks only the shoulder and/or no lanes. These reliability measurements (Figure 24, Figure 25, Figure 26, Figure 27) were calculated for the crashes at different times of day (e.g.: off-peak, no crash; off-peak, crash; peak, no crash peak, crash). The following sections describe our findings.

Impact of a Crash with One Lane Blocked

Compared to the no crash scenario, crashes blocking one-lane crashes have a larger impact on TTV, Buffer Index, PTI, and TTI. According to TRIMARC, for I-65 N and S, TTV was six to seven times when a crash blocked one lane compared free flowing conditions. The buffer index increased by a multiple of three, PTI by two, and TTI by about 40 percent compared to free flowing conditions. It took about 40 percent more time to navigate these segments when there was a crash blocking one lane, compared to normal travel conditions. Therefore, drivers should add about 80 percent more buffer time to ensure on-time arrival.

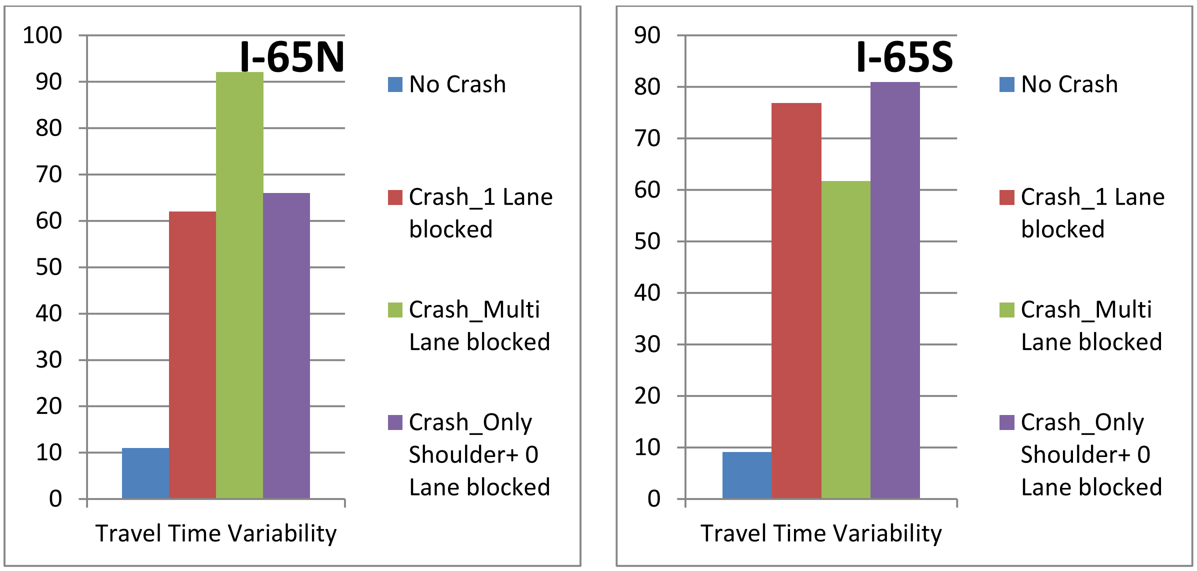


Figure 20: Travel Time Variability (sec./mile) for Different Types of Crash

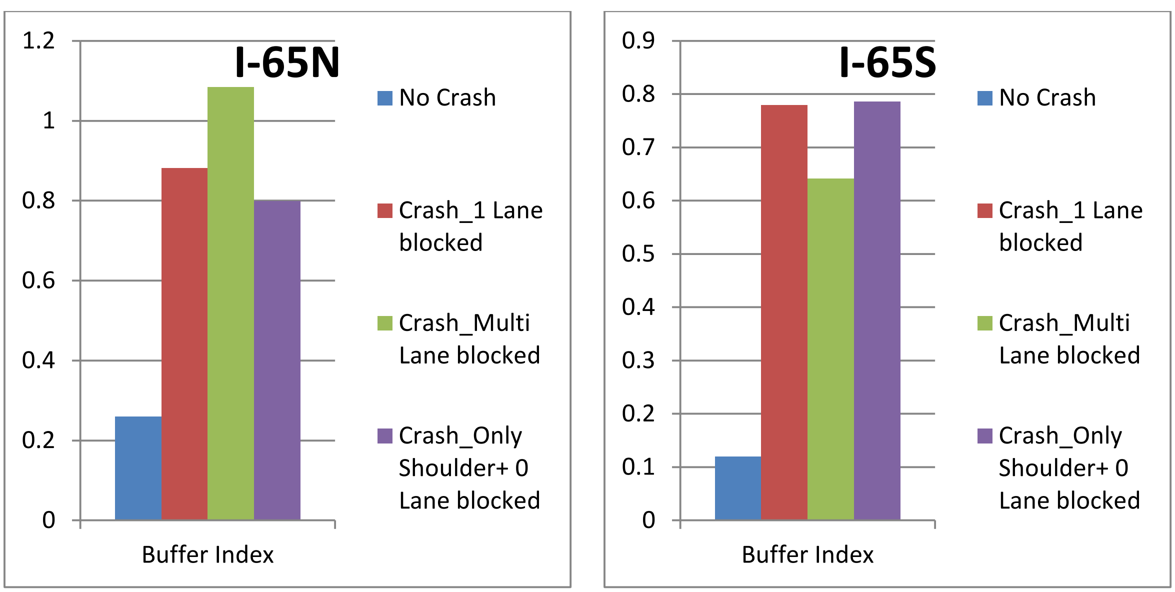


Figure 21: Buffer Index for Different Types of Crashes

During normal conditions, TTV was approximately 10 sec/mile, but when crashes blocked one lane, it increased to 60-75 sec/mile.

Impact of a Crash with only Shoulder or Zero Lanes Blocked

Crashes that only block the shoulder had a similar effect as crashes that blocked a single lane. They also had a larger impact on reliability measures compared to free flowing conditions.

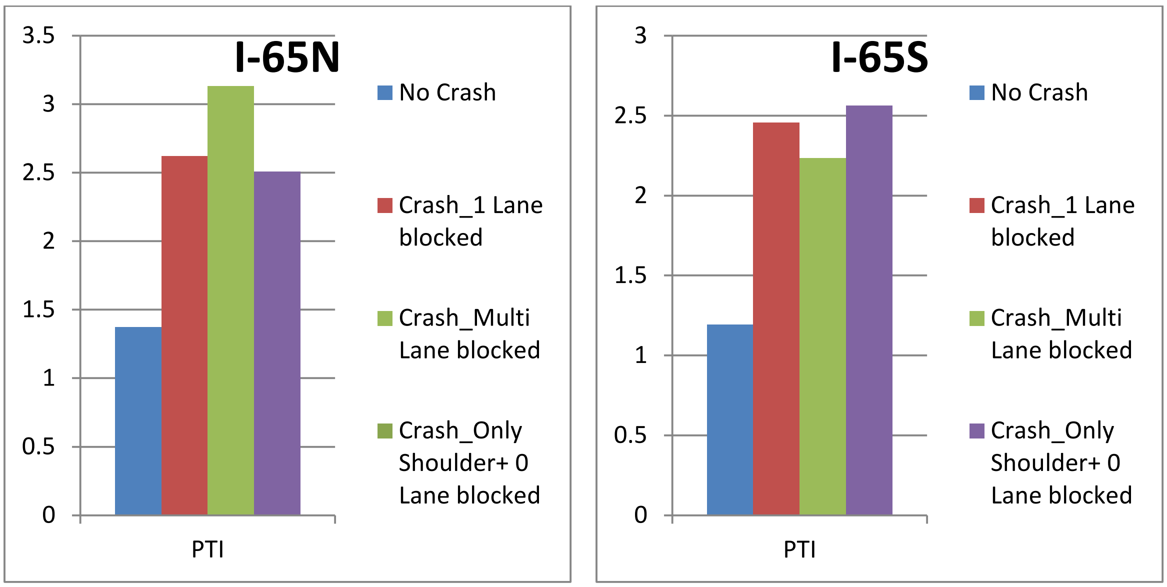


Figure 22: Planning Time Index for Different Types of Crashes

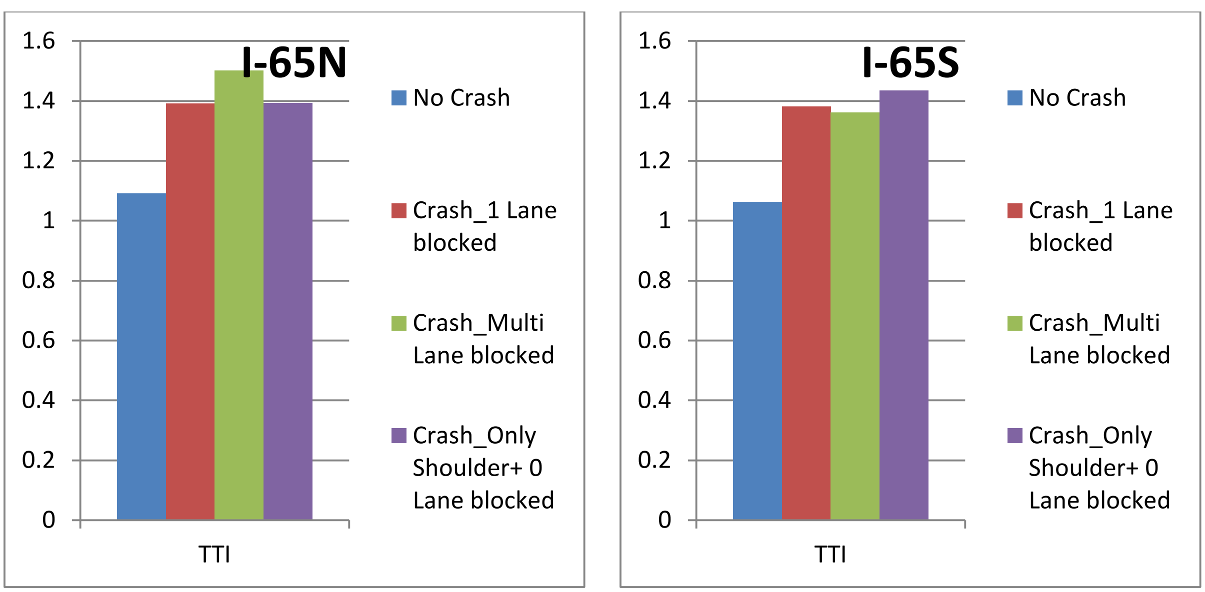


Figure 23: Travel Time Index for Different Types of Crashes

All criteria were similar when crashes blocked one lane. Like crashes that only blocked the shoulder, TTI increased by about 40 percent. Under these conditions, drivers will spend about 40 percent more time traveling than free flow traffic and should add about 80 percent more time (BI) to ensure on-time arrival. During normal conditions, TTV was 10 sec/mile, but when a crash blocked a single lane, this grew to 65 to 80 sec/mile.

Impact of a Crash with Multiple Lanes Blocked

Crashes that blocked multiple lanes had the largest influences on TTV, Buffer Index, PTI, and TTI. On I-65N, TTV increased by a factor of nine when a crash shuttered multiple lanes. The buffer index increased 4 times, PTI grew about 2.5 times, and TTI increased approximately 40-50 percent. As such, it took 40-50 percent more time to navigate this segment when multiple lanes were impassable, compared to normal conditions. During normal conditions TTV was 10 sec/mile but when crashes blocked multiple lanes, it climbed to more than 90 sec/mile. However, there was an outlier in the I-65 S data set. When multiple lanes were blocked, there were lower values across all reliability measures compared to when a crash blocked one lane, just the shoulder, and/or zero lanes.

Impact of Crashes at Different Times of the Day

Reliability analysis measures the impacts of crashes based on the time of day they occur. When crashes were grouped according to peak and off-peak periods and compared to free flow traffic conditions for the corresponding period, crashes yielded a higher index on all reliability measures.

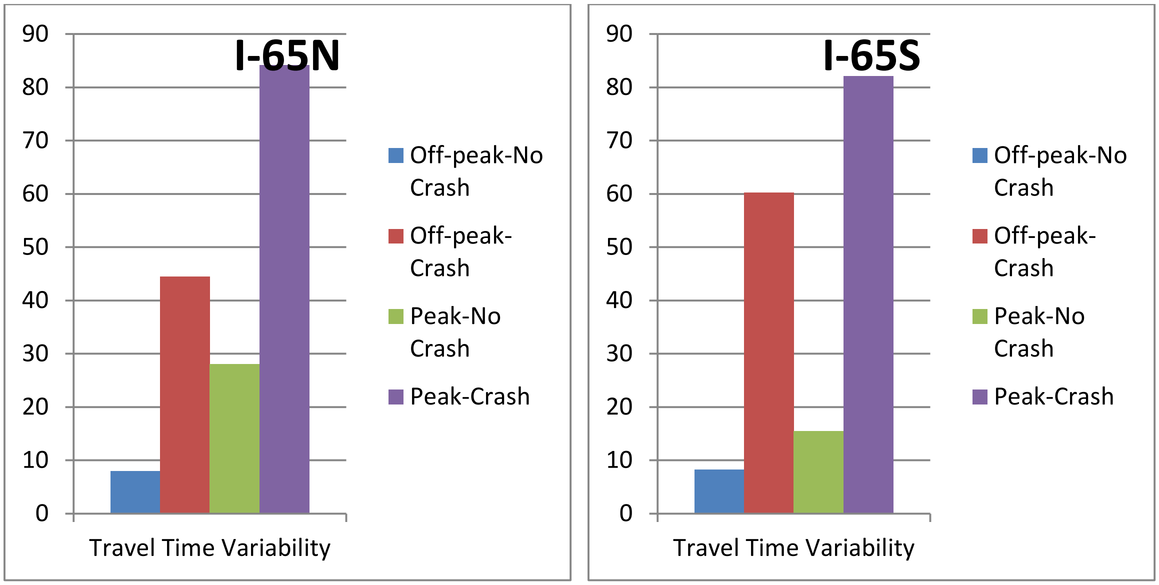


Figure 24: Travel Time Variability (sec/mile) for Crashes at Different Times of the Day

From TTV (Figure 24) measures, we found that during off-peak hours when there were no crashes TTV was about 8 sec/mile. When a crash took place during off-peak hours, TTV increased to 45-60 sec/mile.

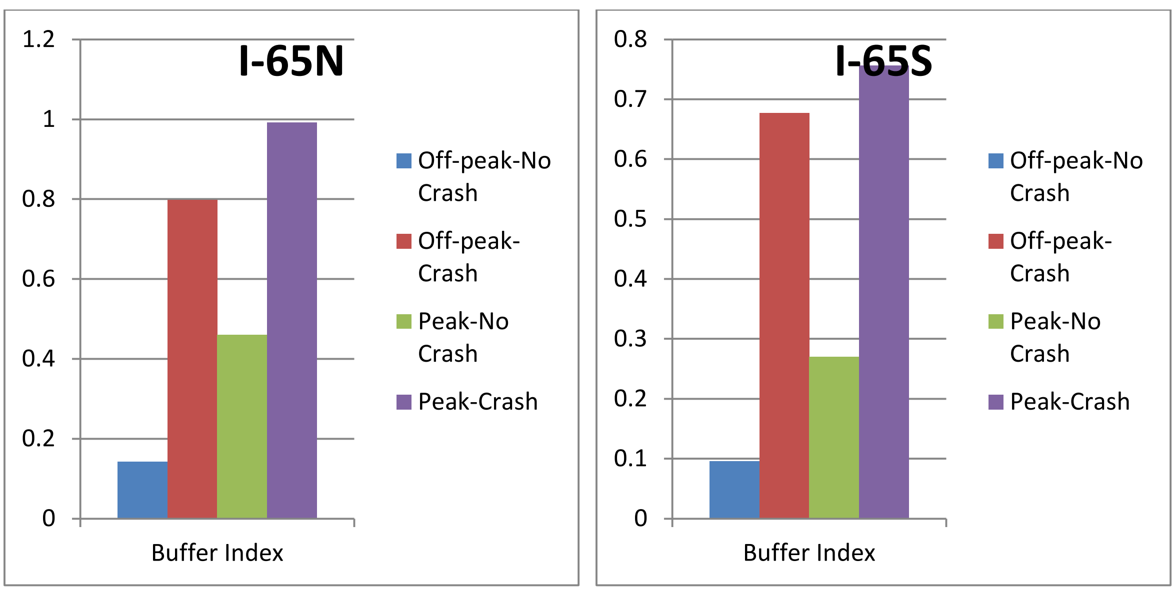


Figure 25: Buffer Index for Crashes at Different Times of the Day

When crashes occur during peak traffic times, this measure became even higher — over 80 sec/mile. Using the Buffer Index (Figure 25), we estimated that a driver would have to spend 80 percent more time than average to ensure on-time arrival when crashes took place during off-peak hours; for peak hours, this increased to 100 percent. Also the PTI (Figure 26) increased between 2.25 to 3 during crash condition. Under these conditions, a driver would require 2.25 to 3 times more travel time than necessary during free flowing traffic to ensure a 95th percentile on-time arrival. When crashes occurred during off-peak times, TTI was about 1.3 and during peak crashes it rose to 1.5 (Figure 27).

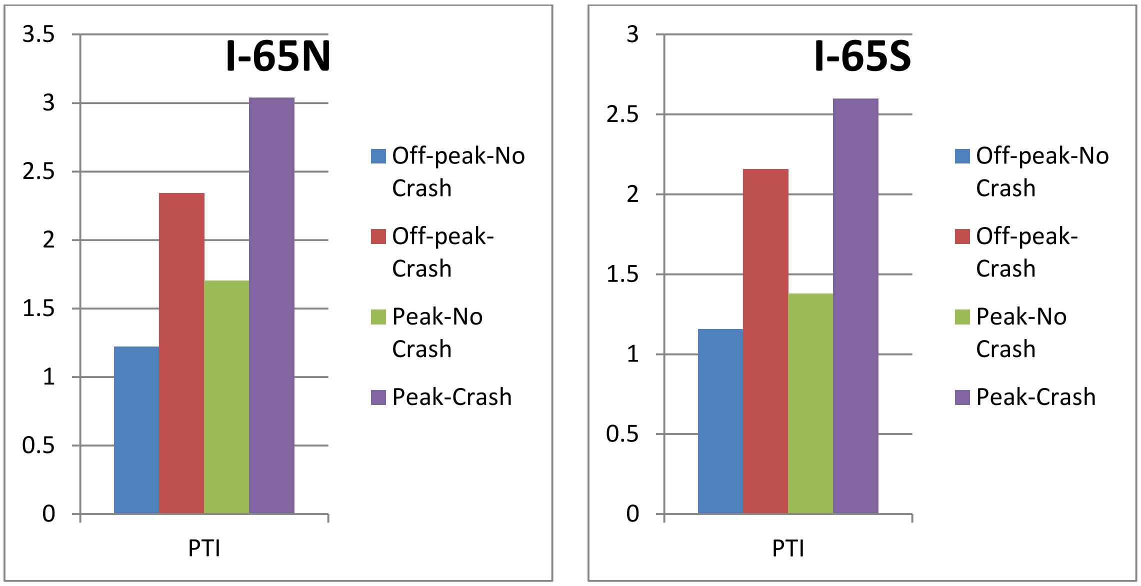


Figure 26: Planning Time Index for Crashes at Different Times of the Day

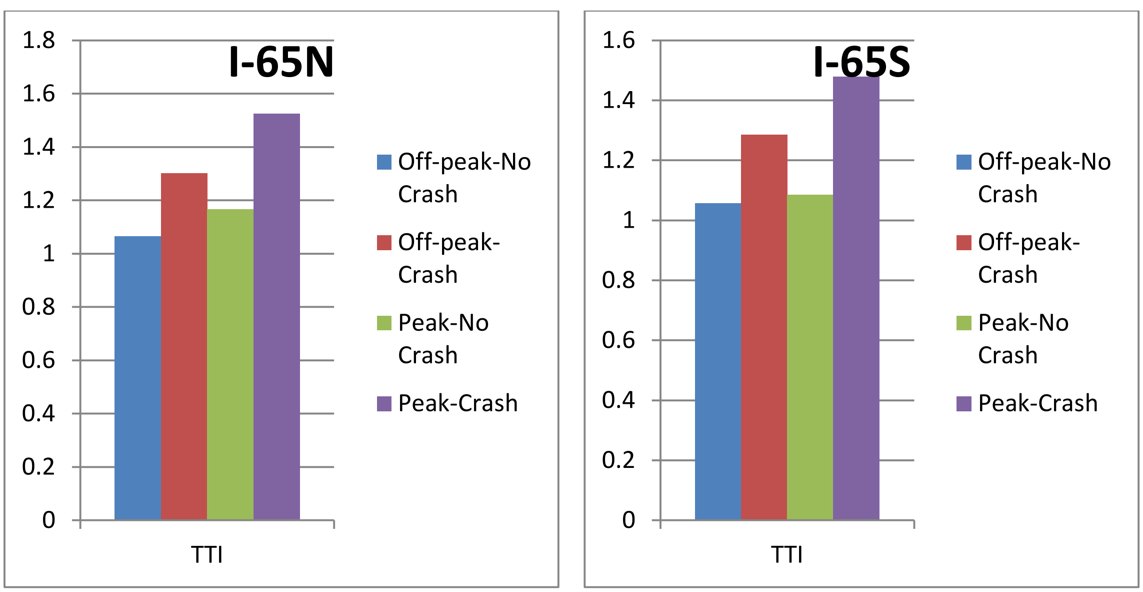
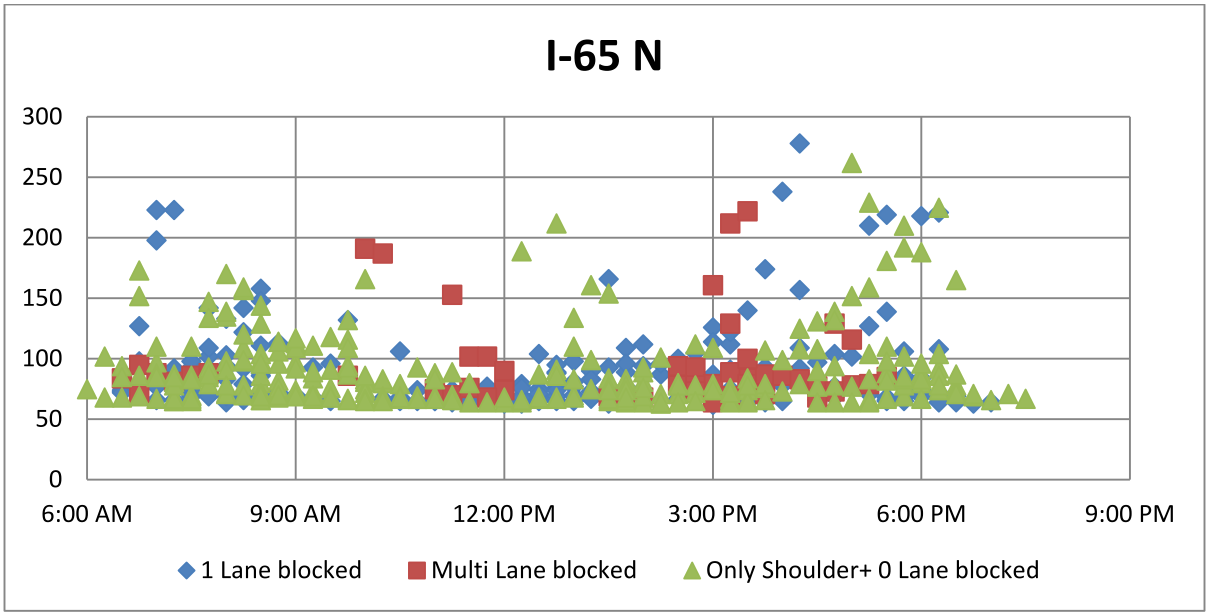


Figure 27: Travel Time Index for Crashes at Different Times of the Day

3.5.4 Variability of Travel Rates Due to Crashes at Different Times of the Day

To determine the dependency of travel rate on the time of day that a crash took place, we filtered the 15-minute intervals during which crashes occurred from the data sets and plotted travel rate (sec/mile) versus time of the day. Crashes that affected travel rates were divided into three groups (crashes blocking one lane, crashes blocking multiple lanes, and crashes



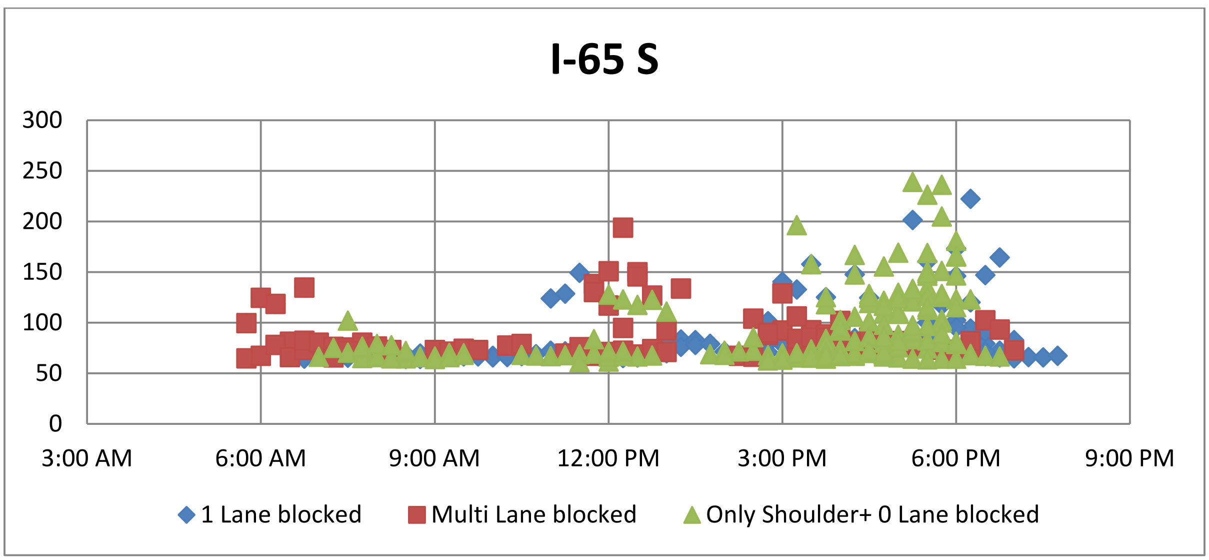
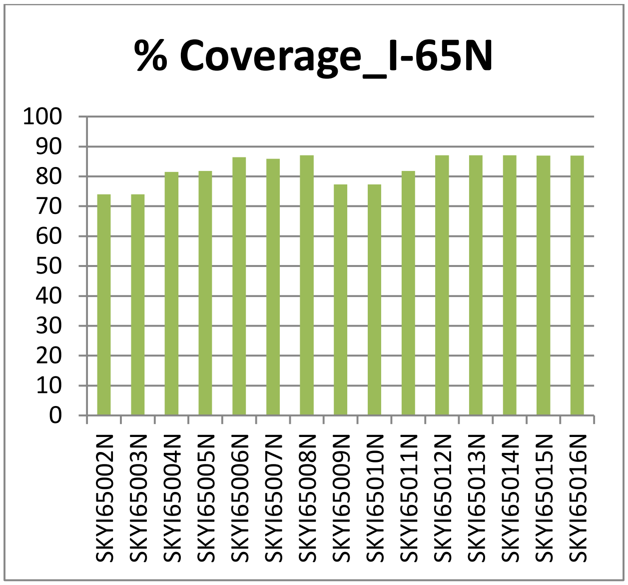


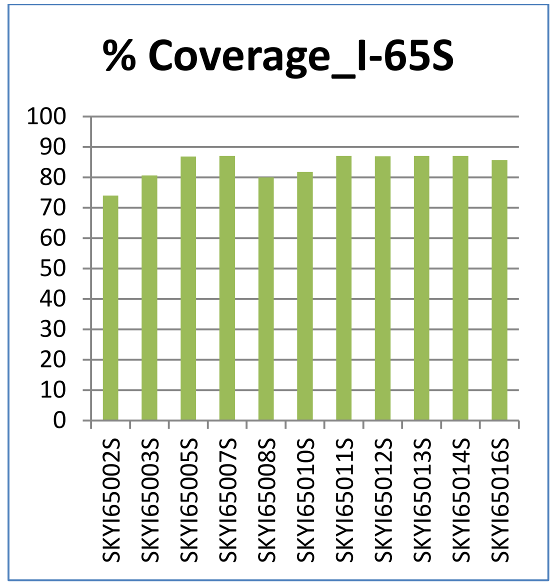
Figure 28: Variability of Travel Rate Due to Crashes

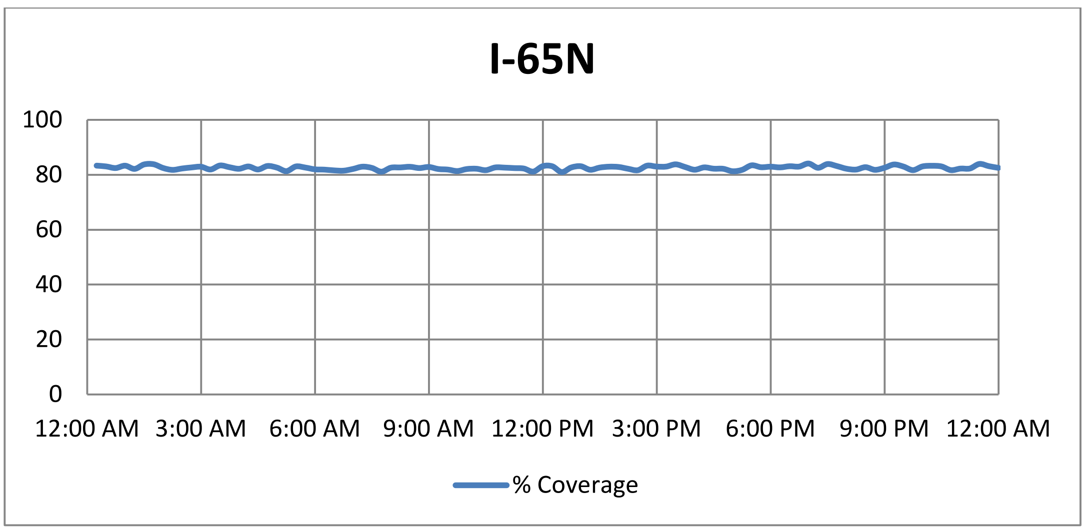
blocking only the shoulder and/or zero lanes). As Figure 28 indicates, if crashes happened during the late-afternoon hours, it caused more variability and higher travel rates. Multi-lane crashes also caused more variability and higher travel rates during any time of the day. Crashes that obstructed a single lane or the shoulder accidents showed a similar pattern, except for a few cases where crashes that only blocked the shoulder had greater variability and higher travel rates than single-lane crashes. The travel rate versus time of the day plot (Figure 28) provides clear insights into how travel rates vary due to time of the day and different crash types.

3.5.5 Temporal coverage:

We examined the temporal coverage of different TRIMARC sensors on I-65 N and S. Two steps were used to perform this analysis — first, we derived temporal coverage for individual sensors, and second, we calculated temporal coverage for different times of the day. For the TRIMARC dataset shown in Figure 29, temporal coverage for individual sensor was uniform. It varied from 75 to 87 percent. For the majority of sensors, coverage exceeded 80 percent. When we computed the temporal coverage at different times of the day, we found that coverage was greater than 80 percent over the entire day. TRIMARC has better coverage for both individual sensors and at different times of the day, this phenomenon observed on both directions of I-65.







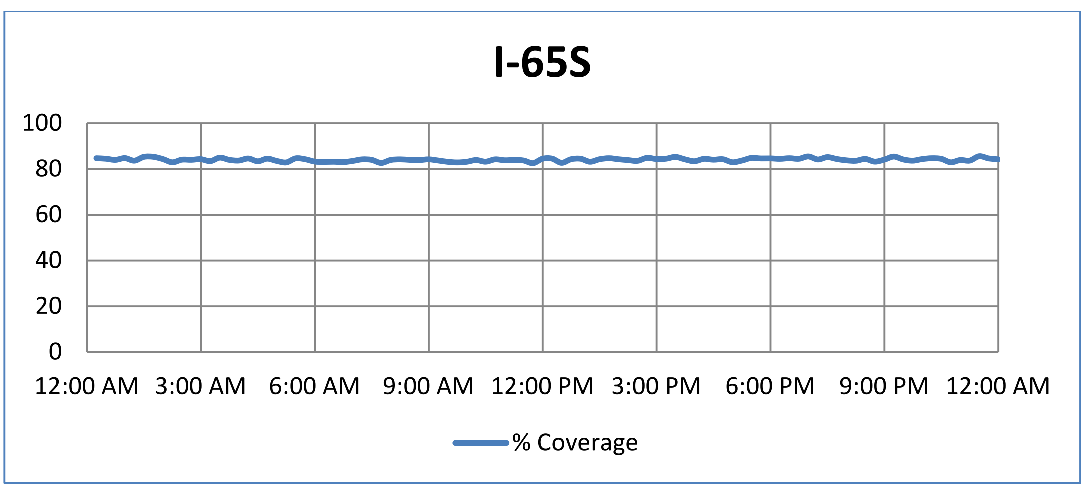


Figure 29: Temporal Coverage of TRIMARC Dataset

3.6 Conclusions

Our application of Big Data used several analyses to study the impacts of traffic crashes on travel time reliability/variability, variation in crash impact during peak/off-peak period, and different crash types. Visualization tools (i.e., heat maps) illustrated the effects of crashes on traffic flow. These tools promise to enhance our visual understanding of the impact of crashes. Our analysis yielded the following conclusions:

- Crashes significantly impact traffic movements, as suggested by the heat maps. They clearly visualize how crashes affect traffic temporally and spatially.
- Crashes negatively impact travel time reliability. TTV, TTI, PTI and buffer index values all increase to some extent during crash condition.
- The CDF of travel rates tells a better story about the impacts of crashes — it shows to what extent crashes influenced travel rates during peak and off-peak periods.
- Among the different crash scenarios, crashes that blocked multiple lanes induce the most significant negative impacts on travel rate. Compared to normal conditions, when 83 to 96 percent of vehicles could travel at a specified travel rate, only 9 to 42 percent of vehicles could travel at that rate during multi-lane crashes. TTV, TTI, PTI, and the Buffer Index all showed highest values during multi-lane accidents.
- Crashes that blocked an individual lane or just the shoulder produced similar impacts on travel rates for both data sets and in both directions.
- Crashes that occurred in the afternoon (3PM–6PM) produced greater variability and higher travel rates under all crash scenarios. Multi-lane crashes led to higher variability and larger travel rates during any time of the day.

Operation efficiency and traffic safety are considered as the most important elements among highway system performance measurement. Traffic congestion serves as a proxy for efficiency, and crash analysis can be used to evaluate highway safety. With the advances in Big Data, improving operations and safety in real-time is now possible. However, to fully realize the power of this data, we need develop more uses for these data. This chapter has illustrated how large data sets can be analyzed with innovative methods. Doing so will improve our understanding of crashes, their impacts, and ultimately their distribution in spatial temporal domain.

REFERENCES

Ahmed, M. M., Abdel-Aty, M., & Yu, R. (2012). Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather, and Traffic Data. *Transportation Research Record: Journal of the Transportation Research Board, (2280),* 51–59

Asai, Nobuhiro; & Akiyama, Kazuhito. (2013). Distributed Intelligence to Transform "Uncertain Big Data" Into "Valuable Data", *20th ITS World Congress,* Tokyo, Japan.

Bakshi, Kapil. (2012). A Primer on Big Data in State and Local Government. *Government Technology*. Retrieved from http://www.govtech.com/e-government/Primer-Big-Data-State-Local-Government.html.

Booz, Allen; Hamilton (2014). Safety pilot model deployment – one day sample data environment data handbook, *USDOT*. Retrieved from https://www.its-rde.net/data/showds?dataEnvironmentNumber=10014

Brunson, S., Kyle, E., Phamdo, N., & Preziotti, G. (2002). *Alert algorithm development program--NHTSA rear-end collision alert algorithm*. National Highway Traffic Safety Administration, DOT-HS-809.

Buckley, Stephen, & Lightman, Deborah. (2015). Ready or Not, Big Data is Coming to a City (Transportation Agency) Near You, *Transportation Research Board Annual Conference.*

Cull, Bill. (2013). 3 ways big data is transforming government. *FCW*. Retrieved from http://fcw.com/articles/2013/09/25/big-data-transform-government.aspx.

(2012). Cloud computing and big data analytics come to local government in growing market. *GovPlace*. Retrieved from http://www.govplace.com/2012/09/cloud-computing-and-big-data-analytics-come-to-local-government-in-growing-market/.

Desouza, Kevin. (2014-a, February 18). Big Data is a Big Deal for Local Government. *International City/County Management Association*. Retrieved from http://icma.org/en/icma/knowledge_network/blogs/blogpost/2162/Big_Data_Is_a_Big_Deal_for_Local_Government.

Desouza, Kevin. (2014-b). Realizing the Promise of Big Data: Implementing Big Data Projects. *IBM Center for the Business of Government.* Retrieved from http://www.businessofgovernment.org/report/realizing-promise-big-data.

Drees, Laura, & Castro, Daniel. (2014, August 18). *State Open Data Policies and Portals.* Center for Data Innovation.

Edmiston, Kelly. (2003). State and Local E-Government: Prospects and Challenges. *American Review of Public Administration, 33(1),* 20-45.

(2015). Four Federal, State and Local Big Data Applications. *Ingram Micro Advisor*. Retrieved from http://www.ingrammicroadvisor.com/big-data/four-federal-state-and-local-big-data-applications.

Giuliano, Genevieve; Rhoads, Mohja, & Chakrabarti, Sandip. (2014). New data, new applications: using transportation system data for regional monitoring. *Transport Research Arena 2014*, Paris.

Gou, Jizhan; Gunda, Veera; Zhang, Li; Huang, Zhitong; & Li, Xiaoling. (2015). Traffic Engineers are Preparing for the Coming Big Data Era: Prototype Development of a Comprehensive Intersection Volume Analysis Tool. *Transportation Research Board Annual Conference.*

Govardhan, M., Parashar, A., & Pendyala, R. (2000). Using GPS Technologies to Collect Multiday Travel Data. *Transportation Research Board Annual Meeting.*

Hojati, A. T., Ferreira, L., Charles, P., & Raduan bin Kabit, M. (2012). Analysing freeway traffic-incident duration using an Australian data set. *Road and Transport Research, 21(2),* 19-31.

Horner, R. (1999). Meeting demands: Citizens, empowered by technology-driven access to customized services, are now demanding similar service from state and local government. *Minnesota Cities, 84(4),* 5-6.

Jones, P., & Clarke, M. (1988). The significance and measurement of variability in travel behaviour. *Transportation,* 15.

Layne, Karen, & Lee, Jungwoo. (2001). Developing fully functional E-government: A four stage model. *Government Information Quarterly*, 18, 122-136.

Liu, Xintao; Yan, Wai Yeung; Chow, Joseph. (2015). Time-geographic relationships between vector fields of activity patterns and transport systems. *Journal of Transport Geography, 45*, 22-33.

Lomax, T., Schrank, D., Turner, S., & Margiotta, R. (2003, May). Selecting travel reliability measures. *Texas Transportation Institute monograph.*

Lv, Yisheng; Duan, Yanjie; Kang, Wenwen; Li, Zhengxi; & Wang, Fei-Yue. (2015). Traffic flow prediction with big data: a deep learning approach. *Intelligent Transportation Systems, IEEE Transactions, 16*(2), 865-873.

Miaou, Shaw-Pin & Harry, Lum (1993). Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accident Analysis and Prevention*, *25*.

Okamoto, H. & Koshi, M. (1989). A Method to Cope with Random Errors of Observed Accidents Rates in Regression Analysis. *Accident Analysis Prevention*, *21*.

Parikh, Gordon, & Hourdos, John. (2014). *Implementation of High Accuracy Radar Detectors for Traffic Safety Countermeasure Evaluation*. Minnesota Traffic Observatory, CTS 14-05. University of Minnesota.

Pas, E.I., (1986). Multiday samples, parameter estimation precision, and data collection costs for least squares regression trip-generation models. *Environment and Planning A*, *18*, 73-87.

Shueh, Jason. (2015). Big Data Could Bring Governments Big Benefits. *GovTech.* Retrieved from
http://www.govtech.com/data/Big-Data-Could-Bring-Governments-Big-Benefits.html.

StateTech. (2013). 8 Benefits of Big Data for State and Local Governments. Retrieved from http://www.statetechmagazine.com/article/2013/05/8-benefits-big-data-state-and-local-governments.

Tarko, Andrew P. (2012). Use of crash surrogates and exceedance of crash statistics to estimate road safety. *Accident Analysis and Prevention, 45*, 230-240.

Tene, Omar, & Polonetsky, Jules. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. *Stanford Law Review Online, 64*, 63-69.

Tu, H., Li, H., Van Lint, H., & van Zuylen, H. (2012). Modeling travel time reliability of freeways using risk assessment techniques. *Transportation Research Part A: Policy and Practice*, *46*(10), 1528-1540.

Vasudevan, Meenakshy; Negron, Daniel; Feltz, Matthew; Mallette, Jennifer; & Wunderlich, Karl. (2015). Predicting Congestion States from Basic Safety Messages using Big Data Graph Analysis. *Transportation Research Board Annual Conference.*

Wigan, Marcus, & Clarke, Roger. (2013). Big Data's Unintended Consequences. *IEEE Computer Society, 6*, 46-53.

West, D. M. (2000). *Assessing e-government: The Internet, democracy, and service delivery by state and federal governments*. Providence, RI: Taubman Center for Public Policy, Brown University.

Wiliszowski, C., Fell, J., Smither, D., Vegega, M., Auld-Owens, A., & Namuswe, N. (2010). *Feasibility of Collecting Traffic Safety Data from Law Enforcement Agencies.* National Highway Traffic Safety Administration, DOT HS 811 447.

Wright, B., Zou, Y., & Wang, Y. (2015). The Impact of Traffic Incidents on the Reliability of Freeway Travel Times. In *Transportation Research Board 94th Annual Meeting* (No. 15-5920).

Wu, Xinkai, & Lu, Henry. (2014). Using high resolution event-based data for traffic modeling and control: An overview. *Transportation Research Part C, 42*, 28-43.

Xie, Fei; Gladhill, Kristie; Dixon, Karen; & Monsere, Christopher. (2011). Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record*, *2241*, 19-28.

Yu, R., Shi, Q., & Abdel-Aty, M. (2013). Feasibility of Incorporating Reliability Analysis in Traffic Safety Investigation. *Transportation Research Record: Journal of the Transportation Research Board (2386),* 35–41.

Yu, R., Abdel-Aty, M. A., Ahmed, M. M., & Wang, X. (2013). Crash-Type Propensity Analysis with Bayesian Models Using Microscopic Traffic and Weather Data. *Transportation Research Board 92nd Annual Meeting.*

Zhong, Cheng; Sisipiku, Virginia; Ksaibati, Khaled; Zhong, Tao. (2011). Crash Prediction on Rural Roads. *International Conference on Road Safety and Simulation.*

APPENDIX A

Data.gov Results

A&I - Crash Statistics: Trends in Motor Vehicle Crashes (1975 - 2000)

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures.(http://catalog.data.gov/dataset/ai-crash-statistics-trends-in-motor-vehicle-crashes-1975-2000)

A&I - Crash Statistics: National Summary of Large Trucks and Buses Involved in Crashes, 2006 - 2010

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-national-summary-of-large-trucks-and-buses-involved-in-crash)

A&I - Crash Statistics: State Profiles

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-state-profiles)

A&I - Crash Statistics: Large Truck and Bus Crash Facts 2009: Early Release - Vehicles: Data Tables 47-58

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-large-truck-and-bus-crash-facts-2009-early-release-vehicles-)

A&I - Program Effectiveness: Compliance Review Effectiveness Model

The objective of the FMCSA's Program Effectiveness research is to measure the effectiveness of the FMCSA Safety Programs. The Compliance Review Effectiveness Model and the Intervention Model provide estimates of the beneficial impact of these programs on reducing crashes resulting in lives saved and injuries avoided. The Resource Allocation model utilizes the results of these two models to analyze the allocation of state resources. (http://catalog.data.gov/dataset/ai-program-effectiveness-compliance-review-effectiveness-model)

A&I - Safety Programs: Data Mining Tool

This area of the website provides information on three of the safety programs established by FMCSA to support this mission. The three programs covered by this area include reviews, roadside inspections of commercial vehicles and drivers, and traffic enforcement stops of CMVs operating in an unsafe manner. Each program is implemented in conjunction with the states and devoted to improving motor carrier safety by reducing the number and severity of crashes involving large trucks and buses. (http://catalog.data.gov/dataset/ai-safety-programs-data-mining-tool)

A&I - Crash Statistics: Trends in Fatal Crash Data by State, 1996 - 2008 (complete report)

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-trends-in-fatal-crash-data-by-state-1996-2008-complete-report)

A&I - Crash Statistics: Large Truck and Bus Crash Facts 2009: Early Release - People: Data Tables 59-70

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-large-truck-and-bus-crash-facts-2009-early-release-people-da)

A&I - NAFTA Safety Stats: NAFTA Safety Stats

NAFTA Safety Stats presents information and statistics on the U.S. operations of all U.S. registered interstate and intrastate motor carriers broken out by national domicile of the carrier. The information and statistics are presented based on the carrier's domicile within each of the three North American Free Trade Agreement (NAFTA) nations: the United States, Canada, and Mexico. The reports present motor carrier safety statistics for 2005-2009. (http://catalog.data.gov/dataset/ai-nafta-safety-stats-nafta-safety-stats)

A&I - Crash Statistics: Large Truck and Bus Crash Facts 2009: Early Release - Trends: Data Tables 1-34 and Graphs 1-8

Crash Statistics are summarized crash statistics for large trucks and buses involved in fatal and non-fatal Crashes that occurred in the United States. These statistics are derived from two sources: the Fatality Analysis Reporting System (FARS) and the Motor Carrier Management Information System (MCMIS). Crash Statistics contain information that can be used to identify safety problems in specific geographical areas or to compare state statistics to the national crash figures. (http://catalog.data.gov/dataset/ai-crash-statistics-large-truck-and-bus-crash-facts-2009-early-release-trends-da)

Carrier Safety Measurement System (CSMS, or SMS) - Raw Data: HAZMAT Motor Carriers

The Federal Motor Carrier Safety Administration's (FMCSA) Safety Management System (SMS) is an automated data system used by FMCSA to monitor motor carrier on-road safety performance. FMCSA analyzes safety performance by grouping carrier data in the SMS into seven Behavioral Analysis and Safety Improvement Categories (BASICs) which are, in turn, used to identify potential safety problems with individual carriers and determine when an enforcement intervention might be appropriate. (http://catalog.data.gov/dataset/carrier-safety-measurement-system-csms-or-sms-raw-data-hazmat-motor-carriers-614e0)

Carrier Safety Measurement System (CSMS, or SMS) - Raw Data: Intrastate Non-HAZMAT

The Federal Motor Carrier Safety Administration's (FMCSA) Safety Management System (SMS) is an automated data system used by FMCSA to monitor motor carrier on-road safety performance. FMCSA analyzes safety performance by grouping carrier data in the SMS into seven Behavioral Analysis and Safety Improvement Categories (BASICs) which are, in turn, used to identify potential safety problems with individual carriers and determine when an enforcement intervention might be appropriate. (http://catalog.data.gov/dataset/carrier-safety-measurement-system-csms-or-sms-raw-data-intrastate-non-hazmat)

Carrier Safety Measurement System (CSMS, or SMS) - Raw Data: Intrastate Non-HAZMAT

The Federal Motor Carrier Safety Administration's (FMCSA) Safety Management System (SMS) is an automated data system used by FMCSA to monitor motor carrier on-road safety performance. FMCSA analyzes safety performance by grouping carrier data in the SMS into seven Behavioral Analysis and Safety Improvement Categories (BASICs) which are, in turn, used to identify potential safety problems with individual carriers and determine when an enforcement intervention might be appropriate. (http://catalog.data.gov/dataset/carrier-safety-measurement-system-csms-or-sms-raw-data-intrastate-non-hazmat-f0526)

Fatality Analysis Reporting System (FARS): FTP Raw Data

The program collects data for analysis of traffic safety crashes to identify problems, and evaluate countermeasures leading to reducing injuries and property damage resulting from motor vehicle crashes. The FARS dataset contains descriptions, in standard format, of each fatal crash reported. To qualify for inclusion, a crash must involve a motor vehicle traveling a traffic-way customarily open to the public and resulting in the death of a person (occupant of a vehicle or a non-motorist) within 30 days of the crash. Each crash has more than 100 coded data elements that characterize the crash, the vehicles, and the people involved. The specific data elements may be changed slightly each year to conform to the changing user needs, vehicle characteristics and highway safety emphasis areas. The type of information that FARS, a major application, processes is therefore motor vehicle crash data. (http://catalog.data.gov/dataset/fatality-analysis-reporting-system-fars-ftp-raw-data)

Fatality Analysis Reporting System (FARS): Online Query Tool

The program collects data for analysis of traffic safety crashes to identify problems, and evaluate countermeasures leading to reducing injuries and property damage resulting from motor vehicle crashes. The FARS dataset contains descriptions, in standard format, of each fatal crash reported. To qualify for inclusion, a crash must involve a motor vehicle traveling a traffic-way customarily open to the public and resulting in the death of a person (occupant of a vehicle or a non-motorist) within 30 days of the crash. Each crash has more than 100 coded data elements that characterize the crash, the vehicles, and the people involved. The specific data elements may be changed slightly each year to conform to the changing user needs, vehicle characteristics and highway safety emphasis areas. The type of information that FARS, a major application, processes is therefore motor vehicle crash data. (http://catalog.data.gov/dataset/fatality-analysis-reporting-system-fars-online-query-tool)

Highway Rail Accidents: Accident/Incident Overview by State/Region

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-accident-incident-overview-by-state-region)

Highway Rail Accidents: Hwy/Rail Table By Railroad

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-hwy-rail-table-by-railroad)

Highway Rail Accidents: Hwy/Rail Incidents Summary Tables

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-hwy-rail-incidents-summary-tables)

Highway Rail Accidents: Highway-Rail Crossing Accident

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-highway-rail-crossing-accident)

Highway Rail Accidents: Frequency of Crossing Collisions

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-frequency-of-crossing-collisions)

Highway Rail Accidents: Master Web Service

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-master-web-service)

Highway Rail Accidents: Accident/Incident Overview

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-accident-incident-overview)

Highway Rail Accidents: Overview Charts By Railroad

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-overview-charts-by-railroad)

Highway Rail Accidents: Overview Charts By State

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-overview-charts-by-state)

Highway Rail Accidents: Highway-Rail Crossings

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-highway-rail-crossings)

Highway Rail Accidents: Hwy/Rail Detail Report

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years 1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-hwy-rail-detail-report)

Highway Rail Accidents: Ten Year Accident/Incident Overview by Railroad/Region/State/County

This file contains reported cases of impacts between on-track equipment and any user of a public or private highway-rail intersection. National files from 1975 through the current year are available for download. In addition, individual files by State are available for the years

1991 through the current year. (http://catalog.data.gov/dataset/highway-rail-accidents-ten-year-accident-incident-overview-by-railroad-region-state-county)

Hazmat 10 Year Incident Summary Reports: Data Mining Tool

Series of Incident data and summary statistics reports produced which provide statistical information on incidents by type, year, geographical location, and others. The data provided is that from the Hazardous Materials Incident Report Form 5800.1 (http://catalog.data.gov/dataset/hazmat-10-year-incident-summary-reports-data-mining-tool)

Hazmat Yearly Incident Summary Reports: Data Mining Tool

Series of Incident data and summary statistics reports produced which provide statistical information on incidents by type, year, geographical location, and others. The data provided is that from the Hazardous Materials Incident Report Form 5800.1 (http://catalog.data.gov/dataset/hazmat-yearly-incident-summary-reports-data-mining-tool)

Highway Statistics: Data Browser

The Highway Statistics is a national transportation publication providing annual information covering highway travel, travel condition, infrastructure performance, highway financing, highway fuel usage, registered vehicles, and licensed drivers. (http://catalog.data.gov/dataset/highway-statistics-data-browser)

Large Truck Crash Causation Study (LTCCS): File 1 (TXT)

The Large Truck Crash Causation Study (LTCCS) is based on a three-year data collection project conducted by the Federal Motor Carrier Safety Administration (FMCSA) and the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation (DOT). LTCCS is the first-ever national study to attempt to determine the critical events and associated factors that contribute to serious large truck crashes allowing DOT and others to implement effective countermeasures to reduce the occurrence and severity of these crashes. (http://catalog.data.gov/dataset/large-truck-crash-causation-study-ltccs-file-1-txt)

Large Truck Crash Causation Study (LTCCS): File 2 (Excel)

The Large Truck* Crash Causation Study (LTCCS) is based on a three-year data collection project conducted by the Federal Motor Carrier Safety Administration (FMCSA) and the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of

Transportation (DOT). LTCCS is the first-ever national study to attempt to determine the critical events and associated factors that contribute to serious large truck crashes allowing DOT and others to implement effective countermeasures to reduce the occurrence and severity of these crashes. (http://catalog.data.gov/dataset/large-truck-crash-causation-study-ltccs-file-2-excel)

Motor Carrier Compliance Reviews and Safety Audits: Data Mining Tool

Contains data on compliance reviews and new entrant safety audits performed by FMCSA and State grantees.(http://catalog.data.gov/dataset/motor-carrier-compliance-reviews-and-safety-audits-data-mining-tool)

NHTSA's Office of Defects Investigation (ODI) - Early Warning Reporting: EWR

Manufacturers of motor vehicles, motor vehicle equipment, child safety systems, and tires are required to submit Early Warning Reporting (EWR) information and documentation to NHTSA in order to comply with the Transportation Recall, Enhancement, Accountability and Documentation (TREAD) act. Public or non-confidential manufacturer EWR data is accessible from the web site.Use the EWR Data Search pages to search for manufacturer EWR data associated with Production (for Light Vehicles only), Property Damage, and Death and Injury records.

NHTSA's Office of Defects Investigation (ODI) - Recalls: NHTSA API

Manufacturers who determine that a product or piece of original equipment either has a safety defect or is not in compliance with Federal safety standards are required to notify the National Highway Traffic Safety Administration (NHTSA) within 5 business days. NHTSA requires that manufacturers file a Defect and Noncompliance report as well as quarterly recall status reports, in compliance with Federal Regulation 49 (the National Traffic and Motor Safety Act) Part 573, which identifies the requirements for safety recalls. This information is stored in the NHTSA database. Use this data to search for recall information related to:- Specific NHTSA campaigns - Product types (http://catalog.data.gov/dataset/nhtsas-office-of-defects-investigation-odi-recalls-nhtsa-api-1e65f)

New Car Assessment Program (NCAP) - 5 Star Safety Ratings: NHTSA FTP

NCAP rates vehicles to determine crash worthiness and rollover safety. The safety ratings are gathered during controlled crash and rollover tests conducted at NHTSA research facilities. Vehicles with a rating of five stars indicate the highest safety rating, whereas a one star indicates the lowest rating. (http://catalog.data.gov/dataset/new-car-assessment-program-ncap-5-star-safety-ratings-nhtsa-ftp)

New Car Assessment Program (NCAP) - 5 Star Safety Ratings: NHTSA API

NCAP rates vehicles to determine crash worthiness and rollover safety. The safety ratings are gathered during controlled crash and rollover tests conducted at NHTSA research facilities. Vehicles with a rating of five stars indicate the highest safety rating, whereas a one star indicates the lowest rating. (http://catalog.data.gov/dataset/new-car-assessment-program-ncap-5-star-safety-ratings-nhtsa-api-7cb17)

New Car Assessment Program (NCAP) - 5 Star Safety Ratings: NHTSA API

NCAP rates vehicles to determine crash worthiness and rollover safety. The safety ratings are gathered during controlled crash and rollover tests conducted at NHTSA research facilities. Vehicles with a rating of five stars indicate the highest safety rating, whereas a one star indicates the lowest rating. (http://catalog.data.gov/dataset/new-car-assessment-program-ncap-5-star-safety-ratings-nhtsa-api-8d8e3)

New Car Assessment Program (NCAP) - 5 Star Safety Ratings: NHTSA OGD

NCAP rates vehicles to determine crash worthiness and rollover safety. The safety ratings are gathered during controlled crash and rollover tests conducted at NHTSA research facilities. Vehicles with a rating of five stars indicate the highest safety rating, whereas a one star indicates the lowest rating. (http://catalog.data.gov/dataset/new-car-assessment-program-ncap-5-star-safety-ratings-nhtsa-ogd)

NTD Safety & Security Summary Data Set: Time Series

Summary ("count") data submitted to the Safety & Security Module of the NTD. Reflects counts of incidents, fatalities, injuries, fires, collisions, etc. (http://catalog.data.gov/dataset/ntd-safety-security-summary-data-set-time-series)

NTD Safety & Security Summary Data Set: Major Only Time Series

Summary ("count") data submitted to the Safety & Security Module of the NTD. Reflects counts of incidents, fatalities, injuries, fires, collisions, etc.(http://catalog.data.gov/dataset/ntd-safety-security-summary-data-set-major-only-time-series)

Preliminary Accident/Incident Data: Daily Data File

Provides preliminary accident/incident data.(http://catalog.data.gov/dataset/preliminary-accidentincident-data-daily-data-file)

APPENDIX B

	1. FDOT Orlando	2. Leesburg VA	3. Minnesota DOT	4. San Diego
Speed	▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪Elevation (m), ▪Heading (deg) , ▪ Count in milliseconds ▪ Documented in spreadsheet			▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪Heading (deg), (collected by GPS) ▪ Documented in spreadsheet
Flow & Lane Occupancy				▪5-minute, hourly, and daily aggregations on Freeway ▪ Documented in spreadsheet
Weather			▪Real time weather (Temp, Wind Speed, Visibility , Precipitation etc)	▪Real time weather(Temp, Wind Speed, Visibility, Precipitation etc)
Incident				▪ Incident description by Highway patrol in Freeway, ▪ Location ▪ ALK Grid ID ▪ Documented in spreadsheet
GPS				▪ALK GPS data, ▪ Collected at 3-second intervals from users of CoPilot, ▪link description ▪ Documented in spreadsheet
Other	▪Basic Safety MsgbyVehicle Awareness Devices (VADs)	▪Basic Safety MsgbyVehicle Awareness Devices (VADs)	▪Vehicle to Infrastructure(V2I)	

	5. Proof of Concept	6. (NCAR) 2009	7. (NCAR) 2010	8. Safety Pilot Model
Speed	▪ Individual OBE_ID ▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪ Documented in spreadsheet	▪ Individual OBE_ID ▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪ Documented in spreadsheet	▪ Individual OBE_ID ▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪ Documented in spreadsheet	▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪Elevation (m), ▪Heading (deg) , ▪contains trip-level summaries ▪ Distance from nearby vehicle & obstacle ▪ Documented in spreadsheet
Flow & Lane Occupancy				
Weather	▪ Real time weather(Temp, Wind Speed&Direction Precipitation etc)	▪ Real time weather(Temp, Wind Speed&Direction Precipitation etc)	▪Real time weather(Temp, Wind Speed&Direction Precipitation etc)	▪Real time weather(Temp, Wind Speed&Direction Precipitation etc)
Incident				
GPS				
RSE Snapshot (Road Side Equipement)	▪ Taken every 5-20s depending on veh. status ▪Longitude(deg), ▪Latitude (deg), ▪ Speed ▪ Acceleration ▪ Bearing ▪ Gradient ▪ Origin- destin. etc ▪ Documented in spreadsheet	▪ Taken every 5-20s depending on veh. ▪Longitude(deg), ▪Latitude (deg), ▪ Speed ▪ Acceleration ▪ Bearing ▪ Gradient ▪ Origin- destin. etc ▪ Documented in spreadsheet ▪ Focused on Data duing rainy or snowy weather	▪ Taken every 5-20s depending on veh. ▪Longitude(deg), ▪Latitude (deg), ▪ Speed ▪ Acceleration ▪ Bearing ▪ Gradient ▪ Origin- destin. etc ▪ Documented in spreadsheet ▪ Focused on comparing weather data bet. Veh. Sensors & Station	
Other	▪ Test bed discription	▪ Test bed discription	▪ Test bed discription	▪Basic Safety MsgbyVehicle Awareness Devices (VADs) ▪ Potential Connected vehicle research data source

	9. Pasadena	10. Portland	11. Seattle
Speed	▪Time stamp, ▪Longitude(deg), ▪Latitude (deg), ▪Elevation (m), ▪Heading (deg) ▪Link Capacity ▪ Database (SQL format),	▪ Freeway loop data ▪ 1hr, 5min, 15min & 20s interval ▪ volume ▪Documented in spreadsheet	▪ Seatle Sensys data which contains: speed , volume ▪ Documented in spreadsheet
Flow & Lane Occupancy	▪ link and turn volumes in arterial System ▪current volumes for 5 minute intervals ▪forecasted volumes for 30 minute intervals ▪ Database (SQL format), Plain Text and VISUM Files	▪ Arterial volume and occpancy data by loop detector ▪ Travel time data from Bluetooth radar ▪ Documented in spreadsheet	▪ Freeway volume and occpancy data by loop detector ▪ Arterial Travel time data every 5 minutes from Automatic Licence plate reader ▪ Documented in spreadsheet
Weather	▪ Real time weather ▪ Database (SQL files) and xml schema files	▪Real time weather(Temp, Wind Speed&Direction , Humidity etc) ▪ Documented in spreadsheet	▪Real time weather(Temp, Wind Speed&Direction, Humidity etc) ▪ Documented in spreadsheet
Incident	▪ Incident description by Highway patrol / county sherrif, ▪ Location ▪Severity ▪ stored as an xml file, which is convertible in spreadsheet	▪ Freeway Incident description from Oregon DOT ATMS. ▪ several entities for each incident ▪ Location ▪Severity ▪ Documented in spreadsheet	▪ Incident description byHighway service patrol ▪ Location ▪Severity ▪ Documented in spreadsheet
GPS			
Other	▪Turn Capacity and Delays	▪ Transit Data (bus, metro, light rail) ▪ schedule , stop event, passengers count ▪ Documented in spreadsheet	▪ Transit Data ▪ Actual Bus arrival & scheduled time data. ▪ Documented in spreadsheet