DEVELOPING A DATA-DRIVEN SAFETY ASSESSMENT FRAMEWORK FOR RITI COMMUNITIES IN WASHINGTON STATE

FINAL PROJECT REPORT

by

Yinhai Wang, Wei Sun, Hao Yang, Christopher Gottsacker, Sam Ricord, and Shuyi Yin
University of Washington

for

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, AK 99775-5910

INE/CSET 19.25

In cooperation with U.S. Department of Transportation,
Research and Innovative Technology Administration (RITA)

rural · isolated · tribal · indigenous

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE

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

4. Title and Subtitle	5. Report Date
Developing a Data-Driven Safety Assessment Framework for RITI Communities in Washington State	Sep 30, 2019
	6. Performing Organization Code
7. Author(s) and Affiliations	**8. Performing Organization Report No.**
Yinhai Wang, Wei Sun, Hao Yang, Christopher Gottsacker, Sam Ricord, and Shuyi Yin University of Washington	
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910	**11. Contract or Grant No.**
12. Sponsoring Organization Name and Address	**13. Type of Report and Period Covered**
United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	Final report, July 2018 – Sep 2019
	14. Sponsoring Agency Code

15. Supplementary Notes

Report uploaded to:

16. Abstract

In the history of this country, rural, isolated, indigenous, and tribal (RITI) communities were commonly overlooked with regards to social infrastructure and support. This issue is evident in the development of the transportation networks of these areas and the distinct lack of road safety in these types of communities. RITI communities carry a significantly disproportionate amount of traffic collisions and fatalities compared to urban areas. In order to improve the traffic safety conditions of the RITI communities in Washington State, it is necessary to build a traffic safety management system. A baseline data platform was developed by integrating the collected safety related data for the RITI communities in Washington State in the Year 1 Center for Safety Equity in Transportation (CSET) project. Besides the baseline data, the traffic safety management also requires the safety assessment framework, which is the corner stone of the traffic safety management system. Therefore, this project aims to develop a data-driven safety assessment framework to enable an effective roadway safety management system and improve the traffic safety conditions for RITI communities. The framework is based on an effective and efficient database management system for traffic and crash-related data of the RITI communities. In addition, in order to assist transportation agencies in practices such as the identification of high-risk roadway segments, the developed database management system has powerful visualization functions. Besides the database management and visualization platform, this project also develops roadway safety performance indices and traffic safety assessment methods in the safety assessment framework. This project also provides guidance on how to utilize these safety performance indices and results of safety assessment methods for visualization and analysis.

17. Key Words	18. Distribution Statement
Database management system; Safety assessment framework; Safety performance indices; Safety assessment methods; RITI communities	

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	45	N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Washington State has faced many challenges to meet the transportation safety needs of Rural, Isolated, Tribal, or Indigenous (RITI) communities. As a result, it is critical to build a safety management system that can manage and analyze multi-source traffic safety data and provide assistance to decision-making that improves the safety conditions of RITI communities in Washington State. The safety management system requires the baseline data platform and the safety assessment framework. In the Year 1 Center for Safety Equity in Transportation (CSET) project, a baseline data platform was developed by integrating the collected safety related data for the RITI communities in Washington State. In this case, the research team of this project aimed to develop a data-driven safety assessment framework, which will be the corner stone of the roadway safety management system for RITI communities.

The research team developed the safety assessment framework with roadway safety performance indices and safety assessment methods supported by a database management and visualization system. The data management system is able to store, access, share, visualize, and analyze the multi-source data collected from the RITI communities, such as traffic data, crash records, roadway characteristics data, weather data, etc. In order to better assist transportation agencies on roadway safety management, the database management system also has a powerful visualization function. Along with the database management system, the research team developed quantitative roadway safety performance indices for rural and tribal areas. Besides the development of safety performance indices, the research team also provided guidance on how to apply the safety performance indices in practice, such as hotspot identification for high-risk roadway segments, via visualization on the data management platform. In addition, the research team also implemented safety assessment methods on the dataset obtained for roadway segments in rural areas. Statistical methods including Poisson regression, negative binomial regression and Empirical Bayes were implemented for crash modeling analysis. The users should take this framework as a technical guidance on how to perform traffic safety assessment on RITI communities and modify the elements of the framework as needed considering the specific characteristics of the situation at hand.

CHAPTER 1. INTRODUCTION

1.1. Project Background

In the history of this country, rural, isolated, indigenous, and tribal (RITI) communities were commonly overlooked with regards to social infrastructure and support. This issue is evident in the development of the transportation networks of these areas and the distinct lack of road safety in these types of communities. RITI communities carry a significantly disproportionate amount of traffic collisions and fatalities compared to urban areas. The Federal Highway Administration (FHWA) in a study from 2012 showed that 54% of roadway fatalities occurred on rural roads, while these areas only contain 19% of the total US population (Federal Highway Administration, 2012). As shown in Figure 1, the fatality rate in rural areas has been almost 2.5 times higher than in urban places from 2007 to 2016. This equates to a fatality rate that is almost 2.5 times higher than in urban areas (National Highway Traffic Safety Administration, 2018). There are many issues prevalent in addressing these safety issues in RITI communities. In general, RITI communities lack the necessary resources and are unable to solve these traffic safety issues, leading to problems in addressing these safety concerns. This has created a feedback loop that has caused RITI communities to continually fall behind the safety standards of urban areas.

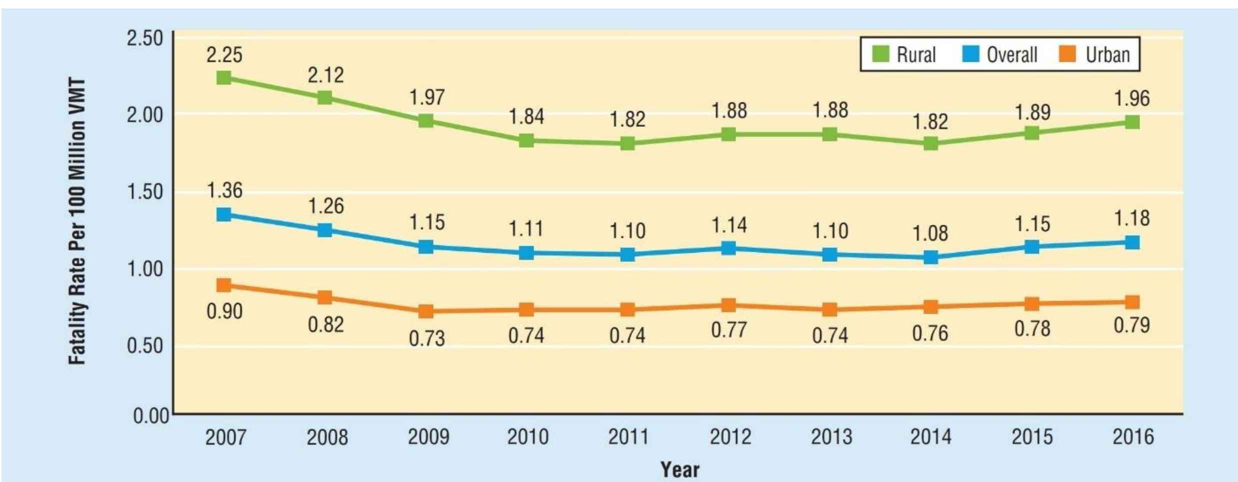

Figure 1 Fatality rates per 100 million vehicle miles traveled, by year and location, 2005 – 2014 (National Highway Traffic Safety Administration, 2018). Source: FARS 2007-2015 final file, 2016 Annual Report File

In Washington State, there are 29 federally recognized tribes. Tribal communities have in the history of this country been treated unjustly and been marginalized by our society. This is clearly evident in the traffic safety of these areas, as Native Americans are almost 3 times more likely to be killed in a rural collision than other ethnicities (Washington Traffic Safety Commission, 2013). This discrepancy in fatalities reflects the historical lack of support of these communities in our society. The traffic safety for the rural areas and tribes in our state is not just an issue of life safety, but also of social justice. These communities should not bear such a hugely disproportionate amount of traffic fatalities. In order for RITI communities to develop in a positive and sustainable way, an effort must be made to rectify some of the longstanding injustice put on these communities.

1.2. Problem Statement

Washington State has faced many challenges regarding the traffic safety of RITI communities. Many studies suggest that roadway segments in rural areas have higher fatality rates than urban areas (Federal Highway Administration, 2012; National Highway Traffic Safety Administration, 2018). It is clear that the traffic safety problem in the rural and tribal areas has become an important social issue in Washington State. In order to improve the roadway safety conditions, many transportation agencies have implemented a certain type of traffic management system to assist traffic safety analysis (Tufte et al, 2010; Xiao et al, 2015; Wang et al., 2016; Cui et al., 2019).

In order to improve the traffic safety conditions of the RITI communities in Washington State, it is necessary to build a traffic safety management system. A baseline data platform was developed by integrating the collected safety related data for the RITI communities in Washington State in the Year 1 Center for Safety Equity in Transportation (CSET) project. Besides the baseline data, traffic safety management also requires the safety assessment framework, which is the corner stone of the traffic safety management system. The traffic assessment framework should be able to classify crash-related data for RITI communities and manage and visualize the traffic and crash-related data in an effective and efficient database system. It is critical to develop safety performance indices that could assist the identification of high-risk roadway segments. In addition, the development and implementation of traffic safety assessment methods such as statistical modeling of crash frequency and severity are also of great importance to assist transportation practitioners in their efforts to improve the traffic safety of RITI communities. Therefore, this project developed a data-driven safety assessment framework to enable an effective roadway safety management system and improve the traffic safety conditions for RITI communities.

1.3. Research Objective

The objective of this research was to develop a data-driven safety assessment framework for RITI communities. The framework is based on an effective and efficient database management system for traffic and crash-related data of the RITI communities. In addition, in order to assist transportation agencies in practices such as the identification of high-risk roadway segments, the database management system has powerful visualization functions. Besides the database management and visualization platform, this project developed roadway safety performance indices and implement traffic safety assessment methods in the safety assessment framework.

CHAPTER 2. LITERATURE REVIEW

The safety assessment framework proposed by the project included a traffic data management and visualization platform and traffic safety assessment techniques, such as safety performance indices and safety assessment methods. The related literature is classified into three categories, including the database management system, applications of database management system in Intelligent Transportation System (ITS), and traffic safety assessment, to introduce the current state-of-the-art and practice as follows.

2.1. Database Management System

A Database Management System (DBMS) is a software that defines, manipulates, stores, retrieves and analyzes data across databases. It consists of a set of programs that allows for database creation, data manipulation (e.g. creation, insertion, deletion) and information retrieval at the requests by its users. Users define features of the DBMS before its use, such as data format, field name, storing structure and validation rules. The DBMS allows multi-user access, removes redundancy and allows for easy recording of complex transactions. There are four types of organization in DBMS: hierarchical, network, relational, and object-oriented, where data are stored in tree-like structure modeling one-to-many relationships, in graphs modeling many-to-many complex relationships, in logically independent tables sharing common fields, and in objects with data and methods as attributes. DBMS has been widely applied in all sectors. Common DBMS products include MySQL, Microsoft Access, Oracle, PostgreSQL, and Microsoft SQL Server.

2.2. Applications of Database Management System in ITS

DBMS has also been increasingly applied in transportation data management, especially with the advancing of Intelligent Transportation System (ITS) and its data collection and analysis demands. A lot of efforts have been focusing on the design and development of online systems to extract, analyze and visualize real-time transportation information.

2.2.1. Freeway Performance Measurement System (PeMS)

Sponsored by the California Department of Transportation (Caltrans), Chen et al from the University of California, Berkeley developed the Freeway Performance Measurement System (PeMS) in 1998. PeMS collects real-time traffic data from sensors installed on freeways and measures performance in all parts of the system. Its audience includes traffic engineers, operators and planners. After collecting data from sensors, PeMS processes them into useful information that can be visualized in plots and summarized in reports, both of which can be generated on PeMS website. Information, including speed, vehicle-hours of delay, vehicle-miles traveled and travel-time analysis is available. To these archives of traffic information, PeMS fits empirical models and is thus able to predict travel time from archived data and to detect data errors. PeMS allows policy makers to evaluate their decisions, planners to monitor congestion and respond with measures, engineers to focus on specific locations, and researchers to model large-scaled traffic behavior with statistically adequate amount of samples (Chen et al, 2001; Chen et al, 2003).

2.2.2. Portland Oregon Regional Transportation Archive Listing (PORTAL)

Portland Oregon Regional Transportation Archive Listing (PORTAL) was developed by Portland State University, originally as a research prototype that stores only freeway loop detector data. Having met its original goals, the system has been progressing towards a second-generation archive with more sources of data and more functionalities in analyzing, modeling and visualizing. PORTAL 2.0 includes weather data, incident data, freight data and transit data. It also provides a custom API integrating R programming language so that graphical data can be displayed with customized plotting in R and people from a wider group can contribute to the project (Xiao et al, 2015; Tufte et al, 2010).

2.2.3. Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net)

The Smart Transportation Applications and Research Laboratory (STAR Lab) at the University of Washington has developed several transportation data storage and management approaches and tools. In order to help Washington Department of Transportation (WSDOT) manage and analyze the huge amount of transportation data, including loop detector data, vehicle trajectory data, etc., the research team of STAR Lab developed a transportation data management, analysis, visualization platform, named Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net) (Wang et al., 2016). The multi-source data used by the DRIVE Net system is stored and managed in the MS SQL server and PostgreSQL databases. The DRIVE Net system is based on the transportation data fusion framework to integrate multiple transportation related spatiotemporal datasets, including loop detector data, probe vehicle data, weather data, incident data, etc. to conduct comprehensive transportation analysis. In this framework, those datasets are integrated by mapping to their spatial information onto a shared geospatial information roadway layer for referencing. In this way, the datasets stored in relational databases can be efficiently queried, summarized, and analyzed.

2.3. Traffic Safety Assessment

2.3.1. Safety Assessment Method

Previous studies on traffic safety assessment have laid solid ground on roadway crash modeling analysis. Several statistical models were developed by Vogt and Bared studying the accidents on two-lane roadway segments in rural areas (Vogt and Bared, 1998). A study by Karlaftis and Golias found that average annual daily traffic to be the most critical contributing factor of crash rates for both two-lane and multilane roadway segments in rural areas, other critical influencing factors included geometric design and pavement condition (Karlaftis and Golias, 2002). Harwood et al. developed an algorithm to predict the safety performance of a rural two-lane highway. The prediction algorithm took into account various factors, such as lane width, shoulder type and width, grades, curvature, passing lanes, and sight distance (Harwood et al., 2000). Abbas investigated the traffic and crash data for five main rural roads in Egypt over 1990-1999 and provided a traffic safety assessment of rural roads by developing several statistical models for crash frequency prediction (Abbas, 2004). Schalkwyk and Washington studied multivariate approaches to realize cost effective safety improvements on two-lane rural state roads in Washington State (Schalkwyk, 2008). The studies mentioned above provided guidance and foundation for developing roadway safety assessment methods for RITI communities.

2.3.2. Rural and Tribal Related Crashes

2.3.2.1 Rurality Classification

There are many definitions to classify rural and urban places. Some definitions are based on jurisdictional boundaries, e.g., the county boundary. One such classification was defined by the Washington State Office of Financial Managements, which categorized a list of counties based on population density and county size (Washington State Office of Financial Management, 2019). Other definitions created by the Washington State Department of Health include the Rural-Urban Continuity Codes (RUCCs) and the Rural-Urban Commuting Area Codes (RUCAs). While the RUCAs categorized urban and rural status as the zip code level, the RUCCs included nine levels of rurality at the county level. The RUCCs was later applied in this project for traffic safety modeling to represent levels of rurality.

2.3.2.2 Rural crashes

The literature suggests a higher fatality rate on roadway segments at rural areas than urban areas (National Highway Traffic Safety Administration, 2018). Driving behaviors, such as driving under influence and reckless driving are contributing to the higher fatality rate in rural areas (Wu, et al., 2013). In addition, the difference of roadway design characteristics between roadway segments in rural and urban areas, such as lane width, shoulder width, grade, number of curves and degree of curvature, etc. could also have impacts on the fatal crash rate. Other than roadway characteristics and driving behaviors, one likely reason is the access to medical resources which relates to the spatial characteristics. In this case, it is important to include spatial correlation when analyzing crash frequency and severity for rural areas (Aguero-Valverde & Jovanis, 2008). Due to jurisdiction issues or crash reporting methods, underreporting has been a problem for the fatal crashes reported by tribes (Ragland, Bigham, Oum, Chen, & Felschundneff, 2014). Outreach activities that lead to a long lasting relationship with the tribal communities could help to obtain the complete crash data and perform better traffic safety analysis for the tribal areas.

CHAPTER 3. OUTREACH EFFORTS

In order to most effectively implement positive improvements for Tribal and RITI communities, the research team must first connect with these communities to create effective channels of communication to find optimal solutions. To connect with Tribes, the research team attended several different conferences for Native American issues to connect with leaders from those communities. This connection, when coupled with face to face meetings with different Tribal leaders proved an effective way for connections to be made between the UW and different Tribes.

The first conference the research team attended was the Tribal Leadership Summit, held at the UW on May 11th of 2018. This conference did not have a specific focus on transportation safety, but it proved an effective platform to connect with a wide range of tribal leaders. At the conference, the research team presented a poster on this research, and we were able to begin many conversations with several different tribes. Additionally, the connections made created the foundation to receive invitations to other conferences, including the Affiliated Tribes of Northwest Indians Conference.

The Affiliated Tribes of Northwest Indians Conference was held from May 21st to the 24th of 2018 in Yakima Nation. For this conference, we contributed funding to become a Silver Sponsor. This increased our visibility to other tribes as well as showing commitment to solving some of the issues faced by tribal communities. This conference, being larger than the previous conference, did have a specific breakout section dedicated to traffic safety that we attended, in which the Tribal Transportation Planning Organization (TTPO) held its quarterly meeting. In this breakout session, the research team connected with 4 tribal transportation leaders, and one expressed an interest in creating a collaboration with the UW immediately. This tribe, The Confederated Tribes of Colville, began working with the UW to create a data sharing agreement for their collision data for which they had received funding from the Washington Traffic Safety Commission to support their efforts. From these connections, further meetings were set up to continue the communication between these tribes and the UW.

The third conference the research team attended was The Bureau of Indian Affairs Tribal Transportation Symposium, which was held in Spokane Washington in February 2019. At this conference, the TTPO conducted another meeting with several different related presentations. This increased the visibility of the research team in the Tribal Traffic Safety Community. Around this time, as the data sharing agreement with The Confederated Tribes of Colville was being written, the tribe visited the laboratory of the research team to learn more about the expertise of the laboratory and how that can be applied to their traffic safety needs. Eventually, the data sharing agreement was completed and signed with collision data shared with the research team. This represented the first concrete connection with a tribe that resulted in a data transfer.

After attending the conferences, the research team was contacted by another tribe in the state, the Tulalip Tribe. They were also interested in connecting with us to share data and see if there were any other services the team could provide for them. As with the Confederated Tribes of Colville, after the initial contacts and meetings, the research team invited members of the tribe to the laboratory to show the potential services the team can provide for the tribes. The research team settled on several courses of action. Firstly, like the Confederated Tribes of Colville, a research agreement will be signed to share collision data. In addition to this, they also expressed interest in working to upgrade their collision data collection infrastructure. This could include an app-based collision recording system along with potential database upgrades. This is a service the team intends to extend to the Confederated Tribes of Colville as well if they are interested.

While the research team has been successful in the goal of outreach to various tribes in the region, it is important to continue the effort to create and maintain connections with tribes in the region. It is important to maintain credibility with the tribes, and to be able to connect with more tribes as they reach out, which will allow the research team to further enhance the project in the future.

CHAPTER 4. SAFETY DATA MANAGEMENT SYSTEM

4.1. Data Collection

The main work before development of the methodology consisted of data collection, processing and building a project database. The overarching objective of this task was to compile the data necessary to assess the relationship among the safety of rural, isolated, tribal, and indigenous (RITI) communities, crash potential, roadway performance, and weather/road surface conditions that would ultimately help guide development of the RITI communities in Washington State safety performance. To do this, the project team collected the following data:

- State Patrol incident data from the Computer Aided Dispatch (CAD) system, including traffic incidents, objects in the roadway, and traffic stops.
- Roadway geometry, in the form of GIS files describing road line geometry, lane count, and roadway class.
- Weather data for National Weather Service stations in Washington State, from the UW Department of Atmospheric Sciences.
- The information on counties in Washington State.
- The information of tribes in Washington State.

4.1.1. Washington State Patrol Incident Data

A single Microsoft Excel file containing the CAD incident data and metadata was provided by Richard J. Warren, CAD Systems Supervisor with the Washington State Patrol (WSP). In total, this file contained 52,901 records dating from December 31, 2009 to September 30, 2017. The structure of the table is shown in Table 1.

Table 1 CAD incident data format

Field	Description	Field Type
District	District Identifier Code	integer code
Area	Area Identifier Code	character
DateTime	Time and Date of Incident	date time stamp
Source	Source of Record	Integer

Field	Description	Field Type
Incident	Incident Identifier	character
Type	Incident Type Code	character
Unit	Reporting Unit Identifier	character
Disposition	Disposition Code	character
Location	Location Description	character
Latitude	Latitude of Incident Location	numeric
Longitude	Longitude of Incident Location	numeric

The primary fields of interest for this project included location and time, type, and disposition. The type field is a text code referencing an incident type, which includes crash and injury types, crimes, traffic stops, and a variety of other incident types that are were associated with some WSP action. The disposition field is a text code indicating the conclusion and/or actions of the responding officer(s). For example, disposition codes include Driving Under the Influence (DUI) arrest, impounded vehicle, and various traffic citation types. This field was included primarily to provide some indication of whether an incident involved a commercial vehicle. Scripts were developed to load the CAD data into a PostGIS database, and several queries were developed to provide some initial insight into the overarching patterns and trends. Table 2 shows the incident count for the top eight incident types by occurrence. Together, these types constituted over 91 percent of all records. In total, 27 incident types were contained in the incident records database.

Table 2 Top eight incident types

Incident Type	Count
DISABLED VEHICLE	21998
TRAFFIC HAZARD BLOCKING	5191
COLLISION PROPERTY DAMAGE	4455
ABANDONED	4282
TRAFFIC HAZARD	3903
DISABLED VEHICLE / TOW ENROUTE	3646
PEDESTRIAN	3489

Incident Type	Count
DISABLED VEHICLE BLOCKING	1587

Our initial investigation suggested that the incident locations were geocoded with a text description of the actual locations, and therefore, the spatial resolution of the locations was quite low. That is, in many cases dozens and even hundreds of incidents were (indicated to be) located at the exact same point in space, and there are substantial stretches of roadway with zero incidents. A map of incident locations from January 2014 through August 2017 in the RITI communities is shown in Figure 2.

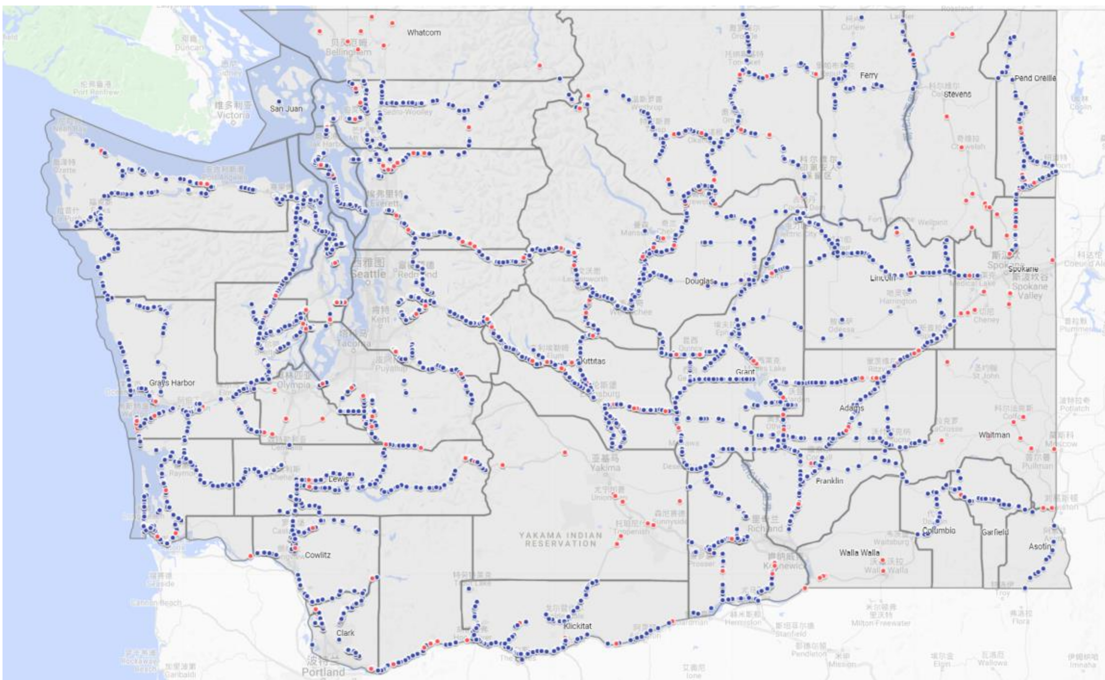

Figure 2 Map of incidents from 2014 – 2017 in the RITI communities based on WSP

4.1.2. Roadway Geometry

Because of the low spatial resolution of the incident data (i.e., that in many cases, dozens or even hundreds of incidents were noted to have taken place at the exact same point in space), the research team decided that roadway line geometry would provide sufficient detail for characterizing incident patterns. For this purpose, state route and interstate line geometries were obtained from the Washington State Department of Transportation at 1:24,000 scale. This data set is updated annually and is used primarily for linear referencing of features along the state highway network and for general purpose geospatial mapping (Blake, 2016). Scripts were developed to convert the ESRI shapefile into an appropriate format and load the file into a PostGIS database. The fields present in this data set are described in Table 3.

Table 3 WSDOT route data format

Field	Description	Field Type
barm	Begin Accumulated Route Mile	numeric
direction	Travel Direction (increasing/decreasing/both milepost, ramp)	character
display	labels for shields	character
earm	End accumulated route mile	numeric
lrs_date	date for linear referencing operations	date string
objectid	feature id	integer
region	geographic/administrative area of WSDOT responsibility	character
RelRouteQual	Related Route Qualifier, description of feature location	character
RelRouteType	Related Route Type, coded route type	character
RouteID	Unique route identifier	character
RT_TypeA	Coded Route Type (text code)	character
Rt_TypeB	Coded Route Type (integer code)	integer
StateRouteNumber	Number assigned to state route	character
STLength	Length of feature	numeric
geom	Feature Geometry	geometry

4.1.3. Weather Data

Weather data for the Olympia, Washington, area were obtained through the UW Department of Atmospheric Sciences website. Data were obtained for the station located at the airport in Olympia, Washington, which was the nearest National Weather Service station providing data for the entire study period (2010 – 2017). No stations providing road surface condition observations for the locations of interest were found. A script was developed to perform the necessary conversions and load the weather data into a PostGIS database. A description of the weather data format is provided in Table 4. A number of additional summary fields were included in this data set, but these are not included for brevity.

Table 4 Weather data format

Field	Description	Field Type
Date	Date and time of Record (GMT)	character
date_julian	Date in Julian format	numeric
pressure	atmospheric pressure in millibars	numeric
air_temp	air temperature in degrees F	numeric
dew_temp	Dew Temperature in degrees F	numeric
wind_dir	Wind direction in degrees from north	integer
wind_spd	Wind Speed in Nautical Miles / Hour	numeric
cloud_cov	Cloud Cover in 1/8ths of sky	integer
cloud_hei	Cloud Height in 100's Feet	numeric
vis_miles	Visibility in Miles	numeric
solar_ird	Solar iradience in Watts/meter^2	numeric
rel_humid	Relative Humididty in percent	numeric
rain	Rain in inches	numeric
sum_rain	Cumulative rain fall in inches	numeric

4.1.4. *County and City Data*

Figure 3 Map of all counties in Washington States

In this research, the counties and cities data are necessary for the safety analysis. According to statistics from the 2010 and 2018 US Census, there are a total of 6.24 million 4,443 people in the state, making it the 13th most populous state. Washington is also the 20th largest land area, covering an area of 66,455.52 square miles (172,119.0 square kilometers). The whole state is divided into 39 counties, a total of 281 cities and towns. In this project, the research team built a database containing the cities and towns information including name, level of the area, population information, etc. Details of the information are in Table 5.

Table 5 City and town data format

Field	Description	Field
name	Name of the city or town	character
level	The level of the city	numeric
county	The county where the city located in	character
pop_2018	The population based on 2018 estimation	numeric

Field	Description	Field
pop_2010	The population based on 2010 estimation	numeric
land_area	The land area of the city/town (acers)	numeric

4.1.5. Tribes Data

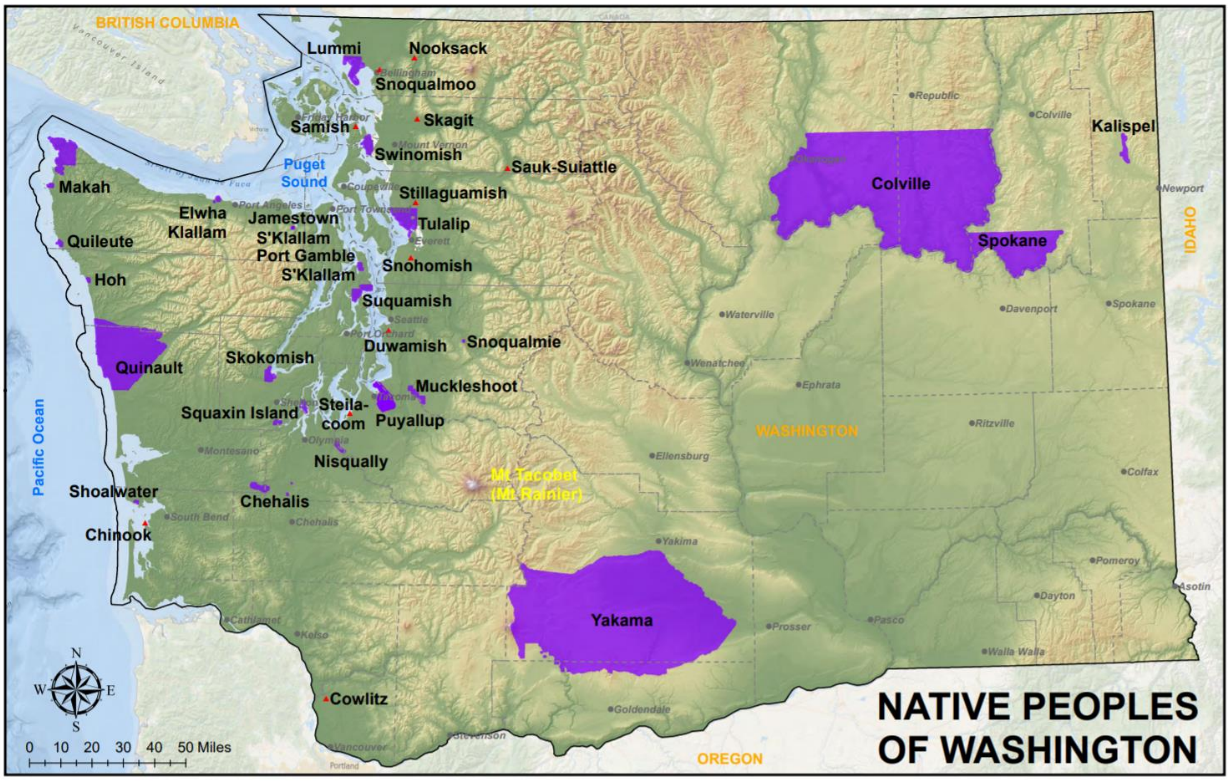

Figure 4 Map of all tribes in Washington States (Source: Native Peoples of Washington, from Puyallup Tribe GIS Department, April 2014)

Tribal data are one of the key databases in this project. In this project, the research team summarized a tribal information database including all tribes located in Washington State. The database included the official name of the tribes, population numbers, land areas and the location of the reservations. The detailed information on tribal data format is shown in Table 6.

Table 6 Tribal data format

Field	Description	Field
name	Name of the tribal official name	character

Field	Description	Field
pop	The population of the tribe	numeric
area	The land area of the tribe (acers)	numeric
loc	The location of the tribe reservation area	character

4.2. Data Storage, Management and Visualization

4.2.1. Data Storage

In this research, the five databases were managed and stored in the SQL server located at the University of Washington. The existing Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net) server at the Smart Transportation Applications and Research Laboratory (STAR Lab) at the UW was used to host the datasets. Researchers, scientists, engineers, students, and transportation agencies are all potential target users of DRIVE Net. As part of establishing baseline crash data infrastructure, the proposed data platform can enable effective traffic safety program management at all levels in RITI communities, and it can further help with design and implementation of appropriate countermeasures to mitigate rural crash risks and reduce associated crash severities.

4.2.2. Data Management and Quality Control

The data quality control and management process in this project is well-designed. Due to the variety of traffic accidents and random occurrences, it is difficult to manage them in a basic standard. Some of the missing and inconsistent data records can be fixed and used based on the data-imputation algorithms. Data mining and analysis need to find more valuable information based on effective data. In order to obtain effective data records, it is necessary to clean the abnormal data, including redundant data processing, missing data processing, and noise data.

4.2.2.1 Redundant data processing

The duplicate data are called redundant data. In general, the attribute value of the data should be unique, but data may be integrated due to data acquisition, transmission, storage, etc., or data integration due to data similarity. When the resulting data are directed to the same entity, multiple similar records appear. Increasing the independence of data and reducing data redundancy are prerequisites for processing redundant data. The processing of redundant data is generally handled by deletion. The redundant data in the highway toll data can be generally divided into two cases, that is, the attribute values are all redundant data and part of the data is redundant data. Generally, redundant data are processed in a manner that all attribute values are duplicated, and the entire data are deleted. For some cases where redundant attributes are used, correlation analysis and monitoring should be performed to determine whether the original data can be restored.

4.2.2.2 Missing data processing

The lack of data values is a common problem in data analysis. The reasons for the lack of data include that the information is temporarily unavailable, the information is missing, some objects or some attributes of some objects are not available, and the system has high real-time requirements. Before processing the missing data, we must first analyze the mechanism and form of the missing data. According to the distribution of the missing, we divide it into completely random deletions, random deletions and completely non-random deletions. Different treatment methods should be determined according to different situations. When it is judged that the missing value is very small, the missing value can be deleted directly.

4.2.2.3 Noise data processing

Noise in the data is also referred to as "meaningless data" or "damaged data" due to random error or errors. An important part of processing noise is to find and delete these data that are significantly different from other data. In statistics, methods such as binning, clustering, and regression are often used. Noise is generated for a variety of reasons, such as data device failures, data input and output errors, and storage media corruption. The data for processing noise are generally deleted and re-screened according to the principle of uniqueness.

4.2.3. DRIVE Net System and Safety data visualization function

4.2.3.1 Brief introduction of DRIVE Net system

Traffic safety and efficiency are critical for our daily life and are thus of common interest. Due to the complexity and dynamic features of transportation systems, our understanding of the characteristics and nature of a transportation system is shallow and stays more or less in the intuitive stage. To further investigate and learn these key issues, more and better data together with a supporting analytics platform are indispensable. Our research product, named the Digital Roadway Interactive Visualization and Evaluation Network (DRIVE Net), was designed to address this important need by integrating various traffic related data for modeling, analysis, and decision-making on top of a digital map.

Government agencies may use the system to quantify important performance and safety measures, companies may see DRIVE Net as a great data source and analytics tool for their relevant business activities (e.g., calibration of simulation models, roadway design optimization, etc.), and individual road users may select their travel modes, routes, and departure times based on DRIVE Net. Figure 5 shows the STAR Lab DRIVE Net platform.

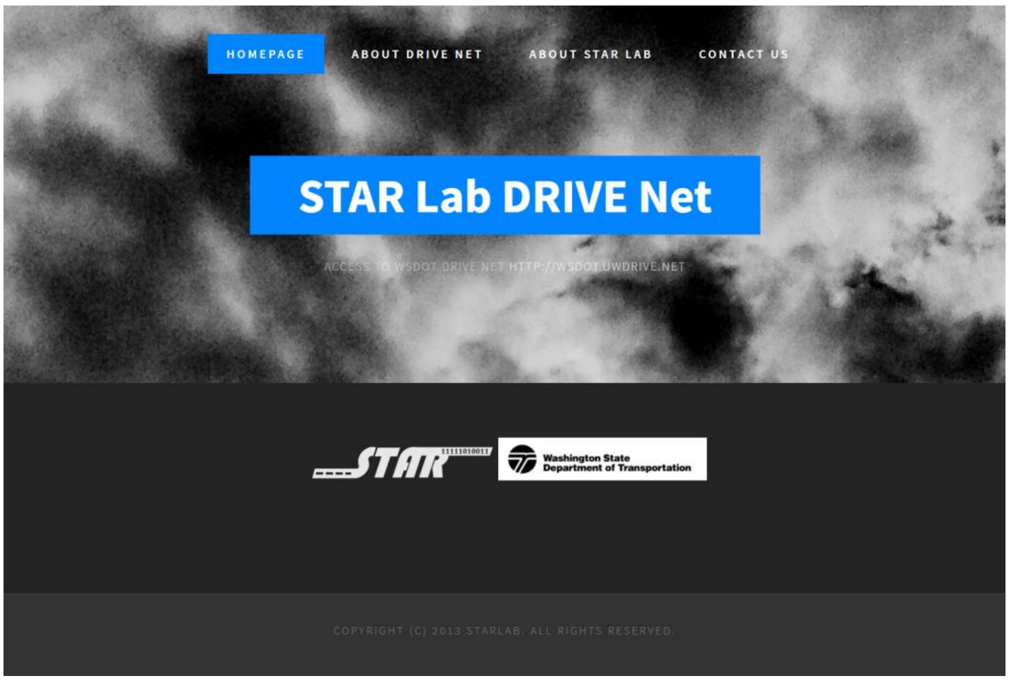

Figure 5 DRIVE Net login page

DRIVE Net serves as a platform to enable data fusion, calculations, and decision support to enable Smart Transportation services in the Smart City context based on big data (such as that provided for governments via sensor networks and various public data sources). The major breakthrough and innovation of DRIVE Net lies in achieving new methods for both how to frame transportation problems and how to provide an innovative and technical means of solving them.

In the past, transportation research mainly relied on mathematical modeling and development of such models was based upon small samples and sporadic data. With increasing amounts of data being collected from intelligent transportation system sensors, data-driven research (as seen in DRIVE Net) is expected to gain increasing prominence in the realm of transportation research and decision-making. At the same time, as internet technologies flourish as they have in recent years, combining concepts from web-based platforms and data-driven methods to solve transportation problems will become increasingly common. Currently, unlike other web-based transportation platforms, DRIVE Net provides a unique implementation of these aforementioned concepts. Further, most existing online systems can only handle one type of transportation data such as that collected from freeway inductive loop detectors. In contrast, our system provides several data layers to enable the integration of multiple data sources (e.g., traffic sensor, incident, and weather data). With the computation done by a powerful computational layer kept separate from the client side of the platform, complex analyses and supporting visualizations, which are based upon user input/option specification, can be carried out and developed in an intuitive and accurate way.

The fusion of large amounts of data, efficient calculations for a host of complicated analysis tasks, and visualizations and reports to support provision of quality transportation services and other decision-making are the major breakthroughs in the overall technical framework of DRIVE Net. In order to achieve the integration of different data sources, the design of DRIVE Net obeys three basic principles: ensuring all necessary inner connections of the perceptual layer, ensuring the efficiency and accuracy of the computational layer, and optimizing the decision layer. Recently, the development and maturity of internet of things technologies, big data, and cloud computing etc. provides the technical support necessary to solve these kind of data integration and analysis problems. DRIVE Net integrates the application of traditional traffic theory, machine learning methods, and optimization and control theory in transportation to exploit the strengths of each methodology. In this way, DRIVE Net is able to produce algorithms for big data processing and provide solutions to traffic problems that are aimed at the user, be they a traveler or transportation agency employee/decision maker. Figure 6 and 7 show the road safety analysis and travel time reliability analysis functions of DRIVE Net.

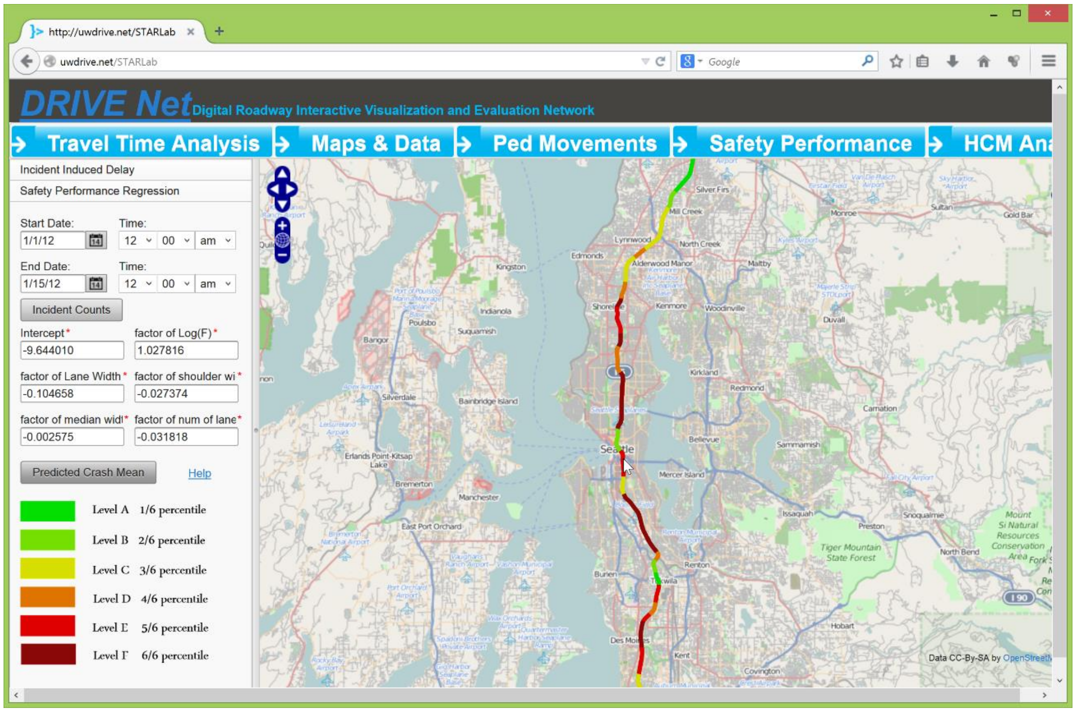

Figure 6 DRIVE Net road safety analysis

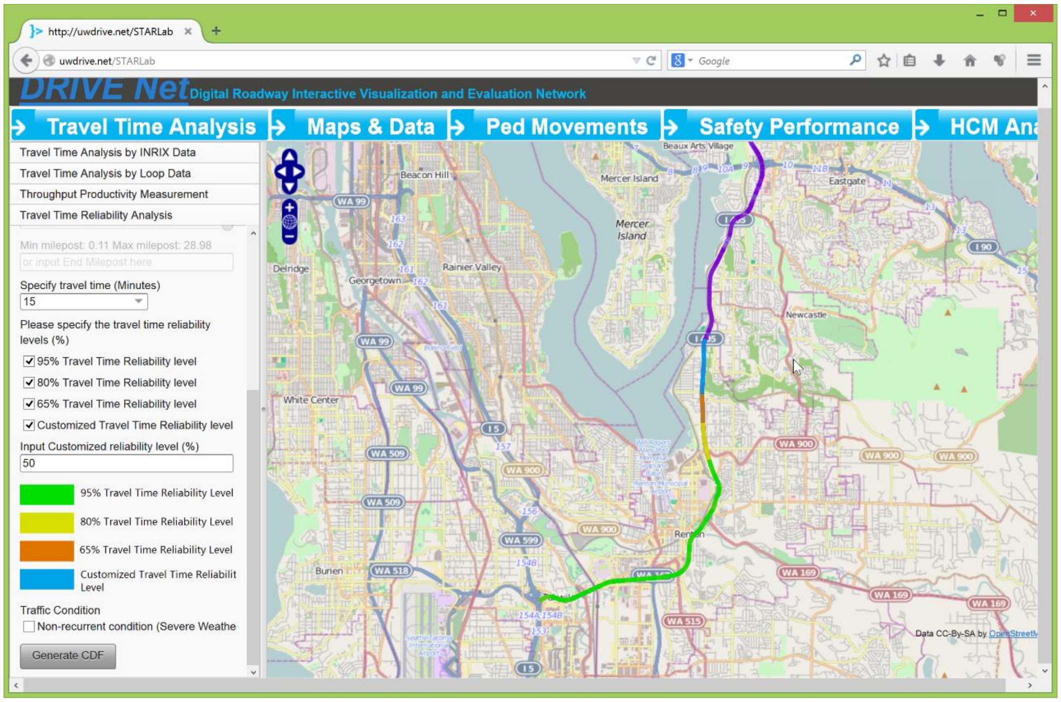

Figure 7 DRIVE Net travel time reliability analysis

4.2.3.2 Safety data visualization function

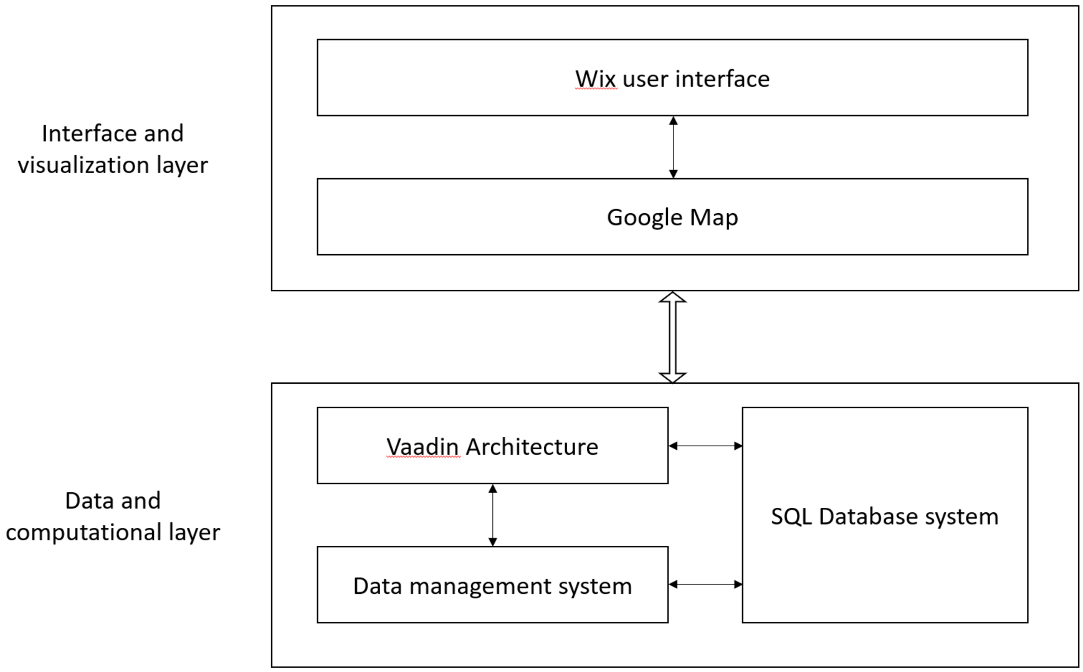

Figure 8 Safety data visualization platform architecture

As shown in Figure 8, the data visualization platform is based on two parts: the interface and visualization layer and the data and computational layer. In the first part, the Wix interface framework, the Google Map system is used to interact with users and visualize the results. Wix is a cloud based on-line open source interface framework based on HTML language. Wix provided a free and user-friendly integrated online interface. Also, with the help of Google Map, the on-line map system can be visited at any corner of the world with Internet without any other settings or packages.

For the second part, the Vaadin architecture is an open source platform for web application development based on Java. The Vaadin platform includes a set of Web components, a Java Web framework, and a set of tools and application launchers. The Vaadin component can be used to extend this demo or create a completely new interface using the Google Maps API or by simply verifying the functionality of the Vaadin application. Also, the data management system based on SQL database are used in this project. Right now, this system uses the crash data for the Washington State Country Routes extracted from WSDOT for data visualization demonstrations. Future features will ensure the robustness of processing large amounts of data. The overall login interface of the on-line safety data visualization platform is shown in Figure 9.

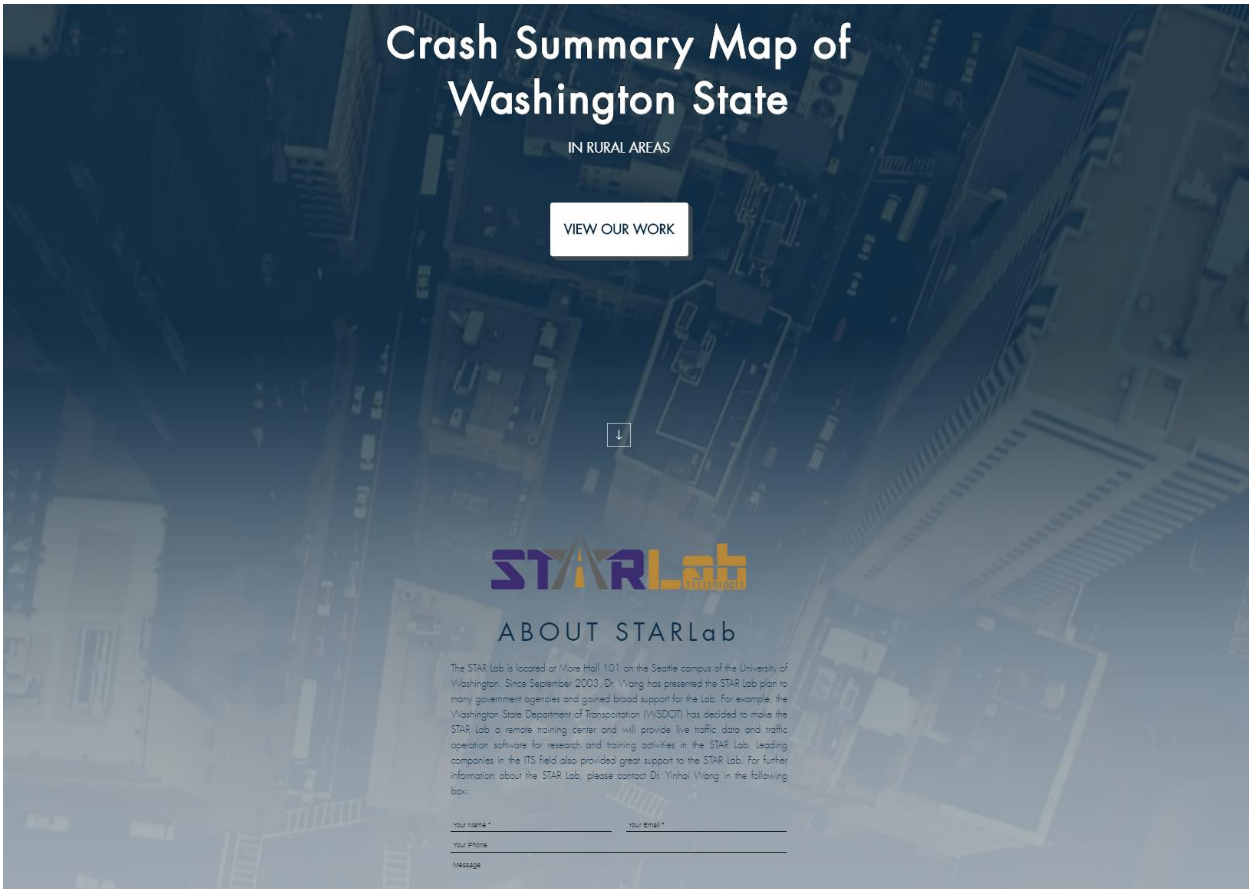

Figure 9 User Interface of the on-line safety module for RITI communities

As shown in Figure 10, each point represents a collision/collision record on a rural route in Washington State. A red dot indicates a fatal collision and a blue dot indicates an injured collision. Users can select a specific year or crash severity from the options in the left window to display those crash locations in the Google Map-based interface. In order to identify the location of each crash on the map, the attribute "(state plane x, state plane y)" of each crash record in the data set is converted to longitude and latitude values, which can be used to locate the position of each latitude and longitude collision. This visualization function may help to present the spatial-temporal distribution of different types of collisions (i.e., death and injury), which lays the foundation for developing hotspot identification analysis functions in this database system.

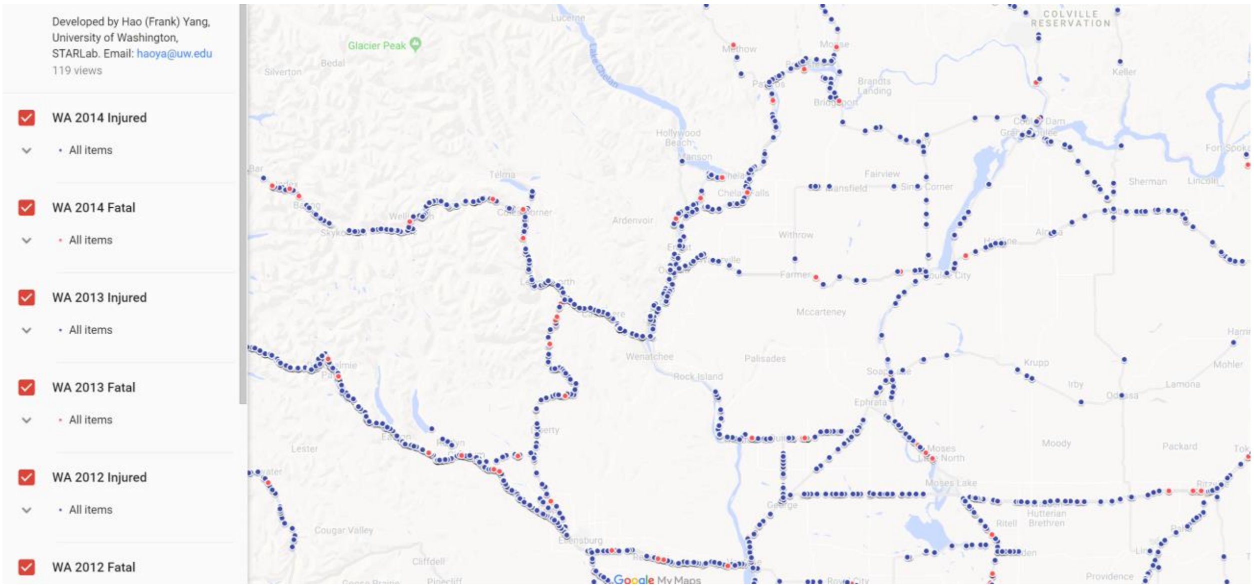

Figure 10 Year and crash severity options

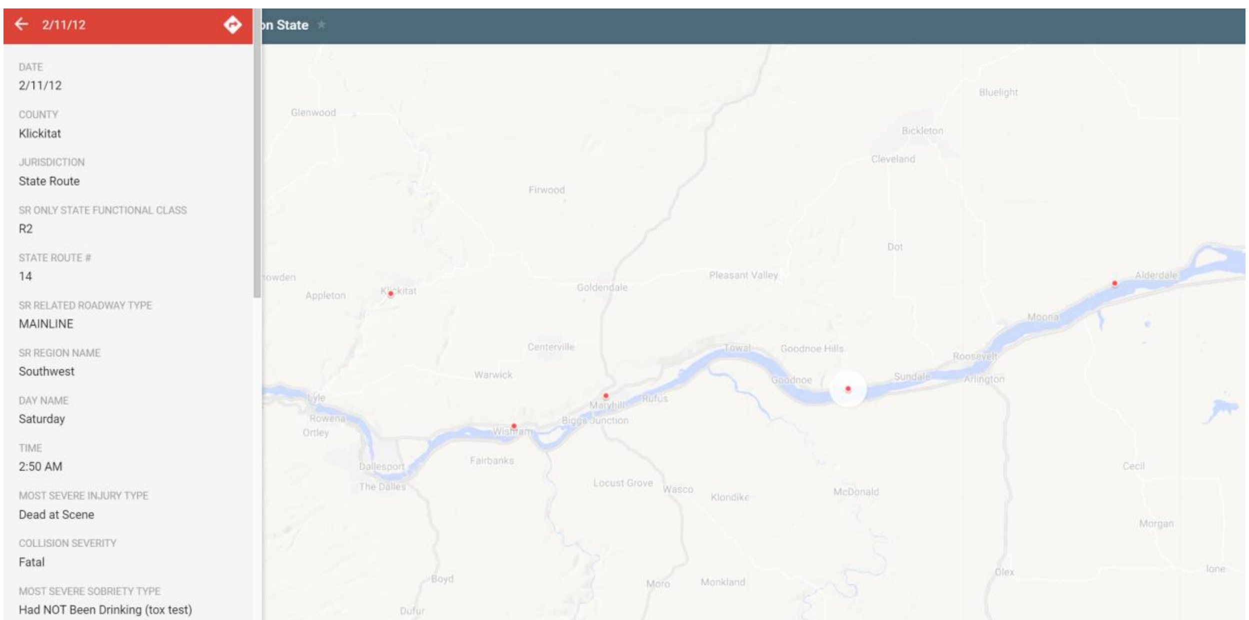

Figure 11 Detailed information for a specific crash record

In addition, as shown in Figure 11, the user can also interact with the collision by selecting a specific point to view its detailed information, such as collision report date, county, route type, route direction, date, time, and the like. This interaction can also show important collision characteristics, such as whether damage is involved, weather conditions, and the disaster relief situations listed in the police report. The user can also obtain the detailed information for a specific crash record. Among other factors, these factors are important for analysis in rural areas because of the high incidence of crashes affected by behavior. However, viewing only the basic information of each crash may not be sufficient to correctly identify the correlation between different crashes, so the analysis function was developed to analyze a set of presentations. Also, it's important to note that all the information that can be viewed with each crash is still linked to the SQL Server database, and the analysis of the data that can be sorted is still not possible, but it is not possible to visualize more complex analyses and results. As such, the demo is still well suited to the purpose of the project, and the user interface is certainly similar to what is expected.

CHAPTER 5. ROADWAY SAFETY ASSESSMENT

This section describes the roadway safety performance indices developed in this project. In addition, various safety assessment methods were implemented in this project for better crash modeling analysis for the RITI communities. Besides the development of safety performance indices and the implementation of safety assessment methods, this project also provided examples of how to visualize the results of the safety performance indices and assessment methods, which will be useful to transportation practitioners for practices such as hotspot identification of high-risk roadway segments.

5.1. Roadway Safety Performance Index

This project developed a new safety performance index (SPI) to assist transportation agencies and practitioners to identify high-risk roadway segments (Wang et al., 2019). In addition, a potential safety improvement index (PSII) was also developed to help identify roadway segments with potential safety improvements (Wang et al., 2019). These two indices could be applied to the database management and visualization platform to color-code roadway segments based on safety performance and highlight areas with potential safety improvements.

5.1.1. Safety Performance Index (SPI)

The SPI was developed using the risk weight and generalized nonlinear model-based mixed multinomial logit approach (Zeng et al., 2017) combining with the Empirical Bayes (EB) approach. Besides crash frequency, the identification of high-risk roadway segments should also take into account crash severities. In the database system for RITI communities, the crash data were categorized into three types of severities, i.e., property damage only (PDO), injury, and fatal. A risk weight factor associating the crash cost with crash severity probability was applied. In order to develop a combined crash density and severity score (CCDSS), the equivalent property damage only (EPDO) method (Washington et al., 2014) is implemented for crash severity categories. The developed SPI is defined as follows (Wang et al. 2019):

$$SPI_i = \lambda_i ECCDSS_i + (1 - \lambda_i) OCCDSS_i \quad (1)$$

Where $ECCDSS_i$ and $OCCDSS_i$ are the expected CCDSS and observed CCDSS for roadway segment i, respectively. λ_i is the weighting factor for roadway segment i.

5.1.2. Potential Safety Improvement Index (PSII)

In addition to the SPI, the potential safety improvement index (PSII) was also developed, which was defined as the difference between SPI and the ECCDSS, as follows (Wang et al., 2019):

$$PSII_i = SPI_i - ECCDSS_i = (1 - \lambda_i)(OCCDSS_i - ECCDSS_i), i = 1,2, \dots, n, \qquad (2)$$

While the SPI could be applied to the data visualization platform to reflect the safety performance of the roadway segments, the PSII could be used to show the areas with potential safety improvement. When PSII value is positive, i.e., the observed CCDSS is greater than the expected CCDSS, it means that the roadway segment has potential safety improvement. On the other hand, when the PSII value is negative, i.e., the observed CCDSS is smaller than the expected CCDSS, it means that there is no significant potential for safety improvement.

5.2. Safety Assessment Methods

This project investigated three modeling techniques for traffic safety analysis, i.e., the Poisson regression, the negative binomial regression, and the Empirical Bayes model. The crash frequency and severity prediction analysis in this project used the dataset obtained from the Highway System Information System (HSIS) for state routes in Washington from 2007 to 2017. Besides crash records, the HSIS dataset also included the characteristics, curvature and grade data of roadway segments. Detailed results and interpretation of the following traffic crash analysis were documented in Christopher Gottsacker's Master's thesis at University of Washington (Gottsacker, 2019).

5.2.1. Poisson Regression

The Poisson regression method was implemented for crash frequency and severity prediction. The probability mass function is shown as follows:

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad (3)$$

The assumption of Poisson distribution is equal mean and variance, which was violated by the HSIS dataset. In this case, the Poisson regression method was not used to estimate crash frequency. However, the Poisson model still indicated the core factors influencing traffic safety.

5.2.2. Negative Binomial Regression

Since the negative binomial regression method is able to handle overdispersion of the data, the negative binomial regression was also implemented for crash frequency and severity prediction. However, the negative binomial model is less suitable when sample size is low. The negative binomial model is shown as follows:

$$\log(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (4)$$

The HSIS dataset contains a default attribute defining whether the roadway segment is located at rural or urban area. However, the attribute has empty values in many records of the dataset. Therefore, this project first implemented the negative binomial model using this built-in attribute of the HSIS dataset to define rurality. After that, the negative binomial regression was implemented with the crash data and characteristics from the HSIS dataset, while including a predictor representing degree of rurality considering the RUCCs codes.

6.2.2.1 HSIS Built-In Rural Definition

As shown in Table 7, the crash frequency of a roadway segment was found to be significantly influenced by factors such as speed limit, average grade, curve count, maximum degree of curvature, the log of segment length, and the log of annual average daily traffic. With the negative parameter estimate for the speed limit, the result of the model indicated that the roadway segments with higher speed limit have lower crash frequency, which seemed to be against common sense. Since roadway characteristics were also considered in the model, one reason could be the segments with higher speed limits were likely to have better geometric conditions, such as better curvature, grade, and visibility conditions. The results indicated that the greater the average grade, the lower the crash frequency. This was also not very intuitive and might be caused by large portion of crashes happening on zero or lower grade segments. As expected, the curve count, maximum degree of curvature, the log of segment length, and the log of annual average daily traffic all had increasing impact on crash frequencies.

Table 7 Negative Binomial results for rural segments (Gottsacker, 2019)

term	estimate	std.error	statistic	p.value
(Intercept)	-4.2160	0.0731	57.7049	0.0000
spd_limt	-0.0059	0.0009	-6.5992	0.0000
avg_grad	-0.0129	0.0060	-2.1646	0.0304
curv_count	0.0576	0.0062	9.3314	0.0000

term	estimate	std.error	statistic	p.value
max_deg_curv	0.0054	0.0017	3.1762	0.0015
log_length	0.4966	0.0076	65.4050	0.0000
log_aadt	0.7339	0.0094	77.8293	0.0000
AIC	66006			

6.2.2.2 Considering RUCCs Codes

When including the RUCCs codes as the degree of rurality, the results of the negative binomial regression model are shown in Table 8. The lower AIC than the previous model with HSIS rurality attribute indicated that the model was a better fit of the data, which indicated that adding the degree of rurality improved the performance of the crash frequency prediction model. The results implied that the lane width and the number of lanes were also significant contributing factors to the crash frequency. Different from the previous model, the speed limit was not found to have significant impact on crash frequency. Lane width was implied to have an increasing impact on crash frequency, which was against common sense, but could be cause by the large amount of one lane width roadway segments in the dataset. Intuitively, the results indicated that the number of segment lanes had a decreasing impact on crash frequency. As expected, the curve count, maximum degree of curvature, the log of segment length, and the log of annual average daily traffic all had increasing impact on crash frequency, which was also consistent with the previous model. The model also implied that an increase in rurality resulted in a decrease in crash frequency. More rural roadway segments having lower average annual daily traffic could cause this. Note that the model only considered crash frequency not severity.

Table 8 Negative Binomial results including RUCCs (Gottsacker, 2019)

term	estimate	std.error	statistic	p.value
(Intercept)	-4.3987	0.0965	45.5671	0.0000
rucc	-0.0887	0.0051	17.4436	0.0000
lanewid	0.0097	0.0037	2.6464	0.0081

term	estimate	std.error	statistic	p.value
no_lanes	-0.1891	0.0116	16.2333	0.0000
avg_grad	-0.0188	0.0059	-3.1913	0.0014
curv_count	0.0619	0.0060	10.2493	0.0000
max_deg_curv	0.0055	0.0017	3.2930	0.0010
log_length	0.4997	0.0076	65.7805	0.0000
log_aadt	0.7979	0.0120	66.2400	0.0000
AIC	65450			

5.2.3. Empirical Bayes

The Empirical Bayes model was implemented with the significant contributing factors obtained from the results of the negative binomial model. The index of expected safety could be obtained through the following equation.

$$\pi_i = w_i * SPF_i + (1 - w_i)K_i \quad (5)$$

Where w_i is a weighting factor between 0 and 1, the *SPF* is the result from the negative binomial model, and K_i is the observed crash count for segment *i*.

The crash reduction potential (CRP) was estimated based on the safety performance function (SPF) using the negative binomial method. The CRP is calculated as follows:

$$CRP = (1 - w_i)(K_i - SPF_i) \quad (6)$$

The estimation of CRP could be used in visualization to help identify high-risk roadway segments along with their potential for safety improvement. As shown in Figure 12, the CRP of the roadway segments in the HSIS dataset were estimated and plotted, the red points indicated there were 1,191 roadway segments with CRP greater than 3, and blue points indicated there were 2,569 roadway segments with CRP greater than 1. For users to implement this safety assessment framework, other statistics of the CRP such as standard deviation and percentile values could also be estimated as needed.

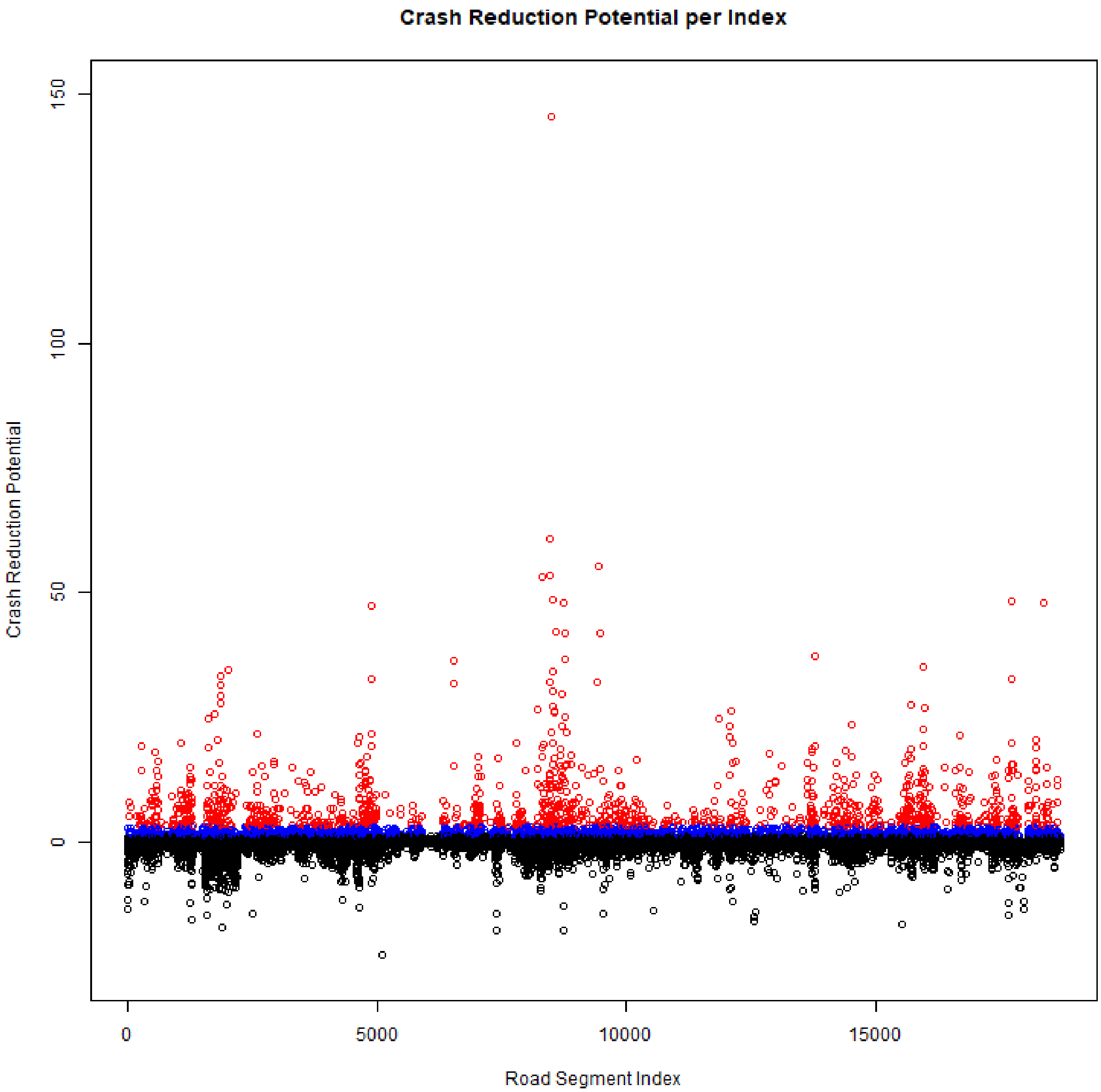

Figure 12 Crash reduction potential per road segment index (Gottsacker, 2019)

CHAPTER 6. SOLUTIONS FOR SAFETY IMPROVEMENTS

To generate positive safety outcomes for the RITI communities and Tribes in our region, we chose to focus our efforts on working with tribal collision data and enhancing the procedures and infrastructure built around this data. The outcomes of our project can generally be split into 4 major categories of collision data: collection, sharing, visualization, and analysis. Each of these categories provides ways to not only improve the traffic safety for the tribes in Washington, but also as catalysts to help begin repairing the issues caused by the social inequities faced by the tribes in Washington and the rest of the country. Additionally, each of these ideas individually represent major improvements in traffic safety, but when combined together their positive impacts build upon the other improvements creating a multiplying effect.

Currently, the tribal police with whom we are working have an extremely outdated method for reporting collisions and keeping records of them. Physical paper and carbon copy forms are still the predominant method by which tribal police agencies record their collision data. This creates many problems for these agencies to use collision data in a useful way because it exponentially increases the amount of work required to synthesize and analyze the data. To combat this, we are developing an app-based collision reporting system that can be deployed on either a tablet or smartphone system. This will be coupled with a server system to securely store and maintain the collision data and information. Implementing such a system will have significant benefits for collision data collection. Primarily, this digitalizes all of the collision data collection, which when stored on a secure server makes it much more accessible and easier to use. Additionally, much more information can be gathered about a collision than with paper only reporting. For example, the devices can automatically save a coordinate location to pinpoint the exact location of the collision, which will constitute a major increase in location accuracy over the paper recording method that requires roadway mileposts to determine collision location.

Another aspect of collision data where we hope to have significant improvement with the tribes is in their data sharing practices. Currently, collision data sharing is extremely rare in tribal communities for a variety of reasons. One such reason is that most Tribes still use paper collision reports, which makes it incredibly difficult to share any information. As detailed above, having paper copies makes it exponentially harder to synthesize data, and as an extension of that transmit said data. It requires either scanning each report and sending them that way or manually copying the valuable information for later use. If collision reports were digitized, it would make data sharing as simple as a click of a button to send all the desired data through the internet, showcasing an example of how these improvements can compound on top of one another. There is however a larger issue at play with the data sharing habits of tribes: they have a history of having an internal focus and being wary of sharing their information with outsiders. Given the historical treatment of tribes in the United States, this behavior is understandable. A major goal of our work is to begin breaking down the barriers that exist, which is possible as universities can act as a third party and garner trust from tribes, which can eventually lead to more openness and cooperation with tribes.

A third aspect of collision data that can be improved for tribes is in the data visualization. Again, this improvement will be made even more powerful when coupled with the previously described actions. Having a digitized record with more data, such as coordinate location points, makes visualization of that data significantly easier. For example, the coordinates allow the creation GIS type maps that show the location of all collisions in relation to each other, which allows sorting and other data to be extrapolated via location. Additionally, if this data is shared, it will be easier for outside groups to help create these visualizations either for the tribes or with permission of the tribes. The main platform we have developed to achieve this goal is called Safety Net. It is a web-based GIS visualization system that shows the location of different severities of collisions. Eventually, it is our goal to have this system set up such that tribes can securely input their collision data directly into the Safety Net to have immediate visualization tools at their disposal.

The final major improvement to tribal collision data practices we aim to implement revolves around collision data analysis. All of the improvements mentioned above will greatly enhance the ability for this to be done more regularly, efficiently, accurately, and effectively for tribes. Improving the data collection of collisions makes it significantly easier to store and manipulate that data. Increasing the shareability of the data makes analysis by outside parties to benefit the tribe significantly easier. Better visualization techniques make many types of analysis that were previously impossible easily and readily available. Eventually, it is our goal to provide some analysis service to the tribes of Washington. By making all of these improvements, it makes it possible for us to conduct analyses such as collision hotspot analysis as an example. This type of analysis will be hugely beneficial to tribes because it can show them in a more comprehensive way than previously used methods which areas of the roadway are most prone to collisions and where funding would be best utilized to reduce severe collisions.

While these improvements all represent significant steps forward for rural areas and tribes in the state, what is more important is the positive benefits they bring the tribal communities. By having an efficient data collection, storage, sharing, visualization, and analysis, it allows the leaders of the communities to begin addressing the traffic safety issues that are prevalent. The new analysis techniques can show which areas of the transportation network are in most dire need of attention so that the most dangerous parts of the network can be addressed first. This will have a major positive impact on the community by generally increasing the safety of the roadway that will result in fewer injuries and fatal collisions. Of just as much importance, the results of these improvements can also be used to showcase the issued faced in transportation safety to a wider range of the populace and policy makers. In short, this analysis could open new channels of funding to address the safety of these communities. If decision makers are shown concrete evidence that these communities need more funding to address their safety needs, they are much more likely to get some of that funding. While these improvements by no means address all of the social issues faced by tribal and RITI communities, they are a great step towards addressing the social injustices faced by these communities and have the real possibility of creating positive outcomes for these communities in regard to traffic safety.

CHAPTER 7. CONCLUSION

This project aimed to develop a data-driven safety assessment framework for the RITI communities in Washington State. The safety assessment framework is based on an effective and efficient database system able to pull from multiple sources to store, categorize and analyze safety data for RITI communities. Primarily working with tribes from Washington State, we aimed to create a platform for these communities to effectively collect, store, share, visualize, and analyze traffic safety data for their use, particularly collision data. Besides the data management system, the project aimed to develop safety performance indices and safety assessment methods to support decision-making that improves the traffic safety conditions of RITI communities.

To better understand current practices in the subject, similar systems were investigated in a comprehensive literature review. The literature review focused on the current state-of-the-art and practices of database management system, applications of database management system in ITS, and traffic safety assessment.

The research team integrated the multi-source traffic and crash-related data from RITI communities into a web-based database visualization platform. This platform takes collision data with coordinate information and displays it in a GIS style format. Initially, data from the WSDOT (Washington State Department of Transportation) of collisions on rural roads was inputted into the database. Then, after a research agreement was signed between the UW and The Confederated Tribes of Colville, the tribal collision data was shared and input into the database. Currently, we are in discussions with the Tulalip Tribe attempting to sign another research agreement with that tribe to include their data in the database. The database system is able to effectively and efficiently store, process, analyze, share and visualize the data. Besides the database management platform, the research team also developed quantitative roadway safety performance indices for rural and tribal areas. In addition, the research team also provided guidance on how to apply the safety performance indices in practice, such as hotspot identification for high-risk roadway segments, via visualization on the data management platform. Other than the safety performance indices, the research team also implemented safety assessment methods on the dataset obtained for roadway segments in rural areas. Statistical methods including Poisson regression, negative binomial regression and Empirical Bayes were implemented for crash modeling analysis. The users of the framework should take it as technical guidance on how to perform traffic safety assessment on RITI communities and modify its elements as needed considering the specific characteristics of the situation at hand.

Additionally, the research team is currently in the process of developing an app-based collision reporting system that can be implemented on a tablet or smartphone to enhance Tulalip's ability to digitally record collision information, including collision location. The goal is for this information to be inputted into the database such that the tribes can have immediate visualization ability on Safety Net. This will greatly enhance the tribe's ability to gather, store and analyze collision safety data. This in turn will greatly enhance the ability of the tribes to use that data to make positive safety improvements to their transportation network and capture more funding than they otherwise would be able to.

By creating these positive outcomes, we have taken a positive step in the direction of increasing the equity of the transportation network for these traditionally underserved communities. While these improvements certainly do not solve all of the social justice issues faced by tribal communities in our country today, these improvements represent movement in the right direction so we as a society can begin to close the longstanding gap in social service to tribal and RITI communities.

CHAPTER 8. REFERENCES

Abbas, K.A., 2004. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. *Accident Analysis & Prevention*, *36*(2), pp.149-163.

Aguero-Valverde, J., & Jovanis, P. P. 2008. Analysis of Crash Frequency with Spatial Models. *Transportation Research Record: Journal of the Transportation Research Board, 2061*(1), 55-63.

Blake, T. A. 2016. State Routes of Washington State 1:24,000. Unrouted Lines Metadata. Olympia, WA: Washington State Department of Transportation. Retrieved from http://www.wsdot.wa.gov/mapsdata/geodatacatalog/maps/NOSCALE/DOT_TDO/LRS/24kLRS.htm

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., & Jia, Z. 2001. Freeway performance measurement system: mining loop detector data. *Transportation Research Record*, *1748*(1), 96-102.

Chen, C., Kwon, J., Rice, J., Skabardonis, A., & Varaiya, P. 2003. Detecting errors and imputing missing data for single-loop surveillance systems. *Transportation Research Record*, *1855*(1), 160-167.

Cui, Z., Henrickson, K., Pu, Z., Guo, G., & Wang, Y. 2019. A New Multi-Source Traffic Data Integration Framework for Traffic Analysis and Performance Measurement. *Transportation Research Board (TRB) Annual Meeting* (No. 19-02736).

Federal Highway Administration, 2012. *Traffic Safety Facts "Rural and Urban Comparison"*. U.S. Department of Transportation.

Gottsacker, C. 2019. *A Pilot Data System and Analytical Framework for Tribal and Rural Community Traffic Safety Equity Assessments*. Master's thesis. University of Washington. Available at: https://digital.lib.washington.edu/researchworks/handle/1773/44116?show=full (Accessed: September 30, 2019).

Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E. and Vogt, A., 2000. *Prediction of the expected safety performance of rural two-lane highways* (No. FHWA-RD-99-207,, MRI 4584-09, Technical Report). United States. Federal Highway Administration.

Karlaftis, M.G. and Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis & Prevention*, *34*(3), pp.357-365.

National Highway Traffic Safety Administration, 2018. Traffic Safety Facts: 2016 Data - Rural/Urban Comparison of Traffic Fatalities. United States Department of Transportation.

Native Peoples of Washington, from Puyallup Tribe GIS Department, April 2014, Source: http://www.puyallup-tribe.com/gis/downloads/NativePeoplesOfWashington_Web.pdf

Ragland, D., Bigham, J., Oum, S., Chen, K., & Felschundneff, G. 2014. *Traffic Injury on Tribal Lands in California.* UC Berkeley: Safe Transportation Research and Education Center.

Van Schalkwyk, I., 2008. *Cost effective safety improvements on two-lane rural state roads in Washington State* (No. WA-RD 695.1). Washington State Department of Transportation.

Vogt, A. and Bared, J., 1998. Accident models for two-lane rural segments and intersections. *Transportation Research Record*, *1635* (1), pp. 18-29.

Wang, Y., Zhang, W., Henrickson, K., Ke, R., & Cui, Z. 2016. Digital roadway interactive visualization and evaluation network applications to WSDOT operational data usage (No. WA-RD 854.1). *Washington (State). Dept. of Transportation*.

Wang, Y., Zeng, Z., Gottsacker, C. and Yang, H.F., 2019. *Develop a Regional Multi-Source Database System for Safety Data Management and Analysis in RITI Communities in Washington State*.

Washington State Office of Financial Management. 2019. *Population density and land area criteria used for rural area assistance and other programs*. Retrieved from Office of Financial Management: Washington Data and Research: https://www.ofm.wa.gov/washington-data-research/population-demographics/population-estimates/population-density/population-density-and-land-area-criteria-used-rural-area-assistance-and-other-programs

Washington Traffic Safety Commission, 2013. *Washington State Strategic Highway Safety Plan 2013*. Washington State Department of Transportation

Wu, Z., Sharma, A., Mannering, F. L., & Wang, S. 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. *Accident Analysis & Prevention, 54*, 90-98.

Xiao, S., Liu, X. C., & Wang, Y. 2015. Data-driven geospatial-enabled transportation platform for freeway performance analysis. *IEEE Intelligent Transportation Systems Magazine*, *7*(2), 10-21.

Zeng, Z.Q., Zhu, W.B., Ke, R.M., Ash J., Wang, Y.H., Xu J.P., Xu, X.X., 2017. *A generalized nonlinear model-based mixed multinomial logit approach for crash data analysis*. Accident Analysis and Prevention, 99, 51-65.