SoftwareX 11 (2020) 100493

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

RuralSpeedSafetyX: Interactive decision support tool to improve safety

Subasish Das [a,*], L.D. White [b]

[a] Associate Transportation Researcher, Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States of America
[b] Research Specialist IV, Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States of America

ARTICLE INFO

Article history:
Received 11 November 2019
Received in revised form 4 March 2020
Accepted 28 April 2020

Keywords:
Rural roadway
Operating speed
Safety performance
Decision support tool

ABSTRACT

The research on relationships among vehicle operating speed, roadway design elements, weather, and traffic volume on crash outcomes will greatly benefit the road safety profession in general. If these associationships are well understood and characterized, existing techniques and countermeasures for reducing crash frequencies and injuries could potentially improve, and the opportunity for new methodologies addressing and anticipating crash occurrence would naturally ensue. The software developed in this study examines the prevailing operating speeds on a large scale and determines how vehicle operating speeds and different speed measures interact with roadway characteristics and weather condition to influence the likelihood of crashes. The developed interactive decision support tool, named as RuralSpeedSafetyX, incorporates Washington and Ohio data containing the expected total crashes from the final models to show segment-level annual crash analysis. The tool is transferable, and it has adaptability options for newer data sets.

Code metadata

Current code version	1.04
Permanent link to code/repository used of this code version	https://github.com/ElsevierSoftwareX/SOFTX_2019_350
Legal Code License	MIT License
Code versioning system used	git
Software code languages, tools, and services used	R, RStudio, R packages (shiny,)
Compilation requirements, operating environments & dependencies	R
If available Link to developer documentation/manual	https://github.com/subasish/RuralSpeedSafetyX
Support email for questions	s-das@tti.tamu.edu

Software metadata

Current software version	1.04
Permanent link to executables of this version	https://ruralspeedsafety.shinyapps.io/rss_sdi/
Legal Software License	List one of the approved licenses
Computing platforms/Operating Systems	iOS, Linux, OS X, Microsoft Windows, Unix-like
Installation requirements & dependencies	R, Several R Packages (shiny)
If available, link to user manual — if formally published include a reference to the publication in the reference list	https://github.com/subasish/RuralSpeedSafetyX
Support email for questions	s-das@tti.tamu.edu

1. Introduction

Current crash prediction methods – such as those in the Highway Safety Manual (HSM) – consist of safety performance functions (SPF) and crash modification factors (CMF). These models use traffic volume and few geometric variables to estimate crash counts. One of the most significant limitations of the

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), l-white@tamu.edu (L.D. White).

https://doi.org/10.1016/j.softx.2020.100493

HSM – and quantitative safety performance research in general – is placement of less emphasis on speed and weather measures. It is generally expected that a vehicle's operating speed during crash impacts butes injury severity of crash victims and that speed differential between drivers contributes to the potential of the frequency of crashes. However, beyond these general relationships, there is minimal consistent evidence for speeds (i.e., posted, average operating, or other relevant speed measures) affecting annual crash frequency, although intuitively speed clearly plays a major role in safety. Another key issue is missing is the inclusion of weather data in the HSM of SPFs. There is an urgent need for research to explore new data and better understand how to effectively quantify highway safety on a daily, hourly, or another short-term basis to overcome the limitations of the current methods.

This study collected data from three sources for Ohio and Washington: (1) Highway Safety Information System (HSIS), (2) the National Performance Management Research Dataset (NPMRDS) Version 1, and (3) National Oceanic and Atmospheric Administration (NOAA) weather data. The project team conflated these databases to develop a database suitable for model development. The project team developed SPFs for total (K - fatal, A — incapacitating injury, B — non-incapacitating injury, C — minor injury, and O — property damage only) crashes, fatal and injury (KABC) crashes, for Washington and Ohio separately for different facility types.

This study developed an interactive decision support tool that provides segment level expected number of annual crashes in heatmap format. The tool was developed using open source R software and its 'Shiny' framework [1,2]. The tool is reproducible (i.e. transferable) and it can be extended by using new variables and additional data. The speed data from NPMRDS is available for all states. Similarly, NOAA precipitation data is also available for all states. Other states can use NPMRDS and NOAA sources for these data to be included in the state specific SPF development process. Once the SPFs are developed, the tool framework and the supporting codes can be used to reproduce similar decision support tool for other states. This transferability characteristic is one of the unique contributions of this study. The tool is currently hosted at: https://ruralspeedsafety.shinyapps.io/rss_sdi/.

2. Problems and background

Although speed is considered a major contributing factor of roadway crashes, research findings are not consistent. While some studies have found that higher speeds are associated with an increased likelihood of collisions, other studies have found the opposite, stating that higher speeds are associated with a lower probability of collisions. A few studies have established statistical models using both operating speed and crash data.

Abdel-Aty and Radwan [3] studied speed by capturing the magnitude of speeding relative to the posted speed limit. This speeding indicator variable was shown to affect the crash involvement of male and young drivers. The preliminary analysis of a study conducted by Taylor et al. [4] based in the United Kingdom revealed that, for the compiled dataset, the average speed was negatively related to crash frequency. The authors attributed this finding to the difference in road quality at the road segments sampled; therefore, they created homogenous groups through which the effects of road quality on the relationship between collisions and speed could be captured. Pei et al. [5] showed that crash risk is negatively associated with average speed when controlling for distance exposure, which goes against research that argues that roadway segments designed for higher speeds should deliver better road safety performance. Yu et al. [6] used a Bayesian inference method to model crashes using one year's worth of crash data on I-70 in Colorado. Their model included real-time weather, traffic, and road geometry variables and indicated that the weather condition variables play a significant role in the crash occurrence. Gargoum and El-Basyouny [7] conducted a study of urban two-lane streets in which they attempted to model the relationship between average speed and crash counts while considering effects from confounding factors. They found that the standard deviation of speed seemed to be negatively related to crash frequencies. In a recent study by Yu et al. [8], the impacts of aggregation approaches on relationship analyses were investigated based on the advanced traffic sensing data of urban expressway systems in Shanghai. Another recent study conducted by Banihashemi et al. [9] found that the severity of crashes (a ratio of FI crashes to total crashes) increased as the speed differential increased. Based on the differing findings regarding the relationship between speed (both operating speed and speed variability) and crash risks across the literature, there is an opportunity to further advance this debate.

Examining free-flow speeds on curved highways in rural New York State presented that drivers did not reduce their speeds sufficiently on curves in the presence of wet-pavement conditions [10]. The researchers concluded that drivers did not recognize that pavement friction is lower on wet pavement compared to the dry pavement.

Jackson and Sharif [11] found that rain is a contributor to crashes in few counties but at less than 95 percent confidence in some of the wetter counties. Mayora and Pina [12] analysed ten years of crash data from two-lane rural roads on the Spanish National Road System and estimated a skid threshold. The results showed that pavement friction improvement yielded significant reductions in wet-pavement crash rates averaging around 68 percent. Najafi et al. [13] used New Jersey crash data and pavement condition data to develop regression models to examine the effect of friction on the rate of wet- and dry-condition vehicle crashes for various types of urban roads.

The literature review reveals that very few studies used both operating speed and weather data to understand the relationships among vehicle operating speed, roadway design elements, weather, and traffic volume on crash outcomes.

3. Software framework

3.1. Data sources

The two primary databases were conflated: (1) The National Performance Management Research Dataset (NPMRDS) and (2) The Highway Safety Information Systems (HSIS) data. Later, the project team assigned the precipitation data from the National Oceanic and Atmospheric Administration (NOAA) on the conflated segments.

NPMRDS: Since July 2013, the Federal Highway Administration (FHWA) procured NPMRDS to support Freight Performance Measurement (FPM) and Urban Congestion Report programs. The NPMRDS includes probe vehicle-based travel time data (for both passenger and freight vehicles) at 5-min intervals for all National Highway System (NHS) facilities. The first version of the NPMRDS is known as 'Version 1' or 'HERE NPMRDS'. The recent version is known as 'Version 2' or 'INRIX NPMRDS', which provides data from January 1, 2017. The NPMRDS data consists of a static GIS file and a database file. The GIS shapefile containing static roadway information was used to relate the travel time information to each traffic message channel or Traffic Message Channel (TMC) segment. The GIS shapefile was provided for visualizing and geo-referencing the NPMRDS data to different maps. The TMC file contains TMC segment geometry information. A database

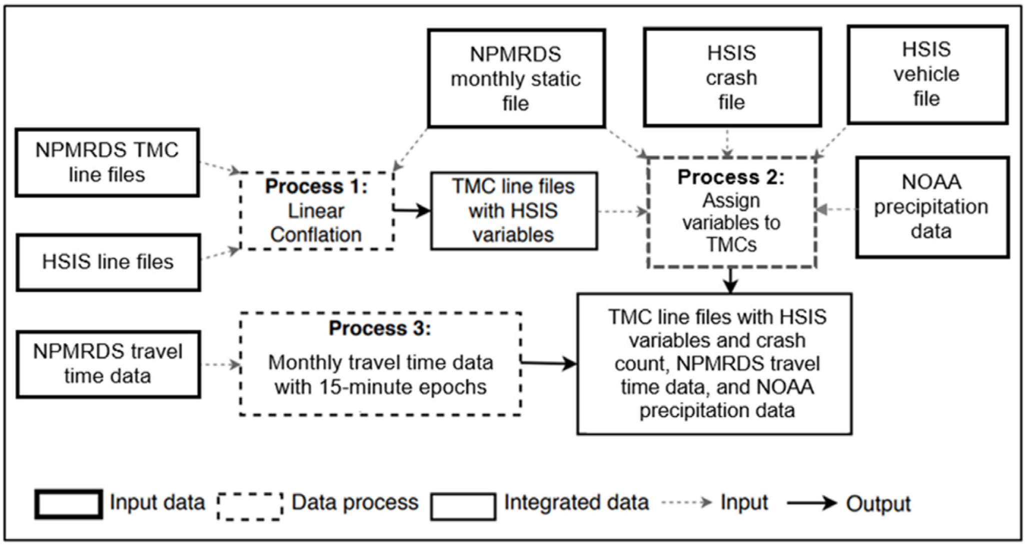

Fig. 1. Data conflation.

containing a set of files includes average travel times of passenger car, truck, and both passenger car and truck for identified roadways geo-referenced to TMC segment IDs.

HSIS: The HSIS data is a multi-state safety database that contains crash, roadway inventory, and traffic volume data for several States (Washington, California, Minnesota, Michigan, Illinois, Maine, Ohio, North Carolina, and Utah). Typical data include variables such as collision type, vehicle category, gender and age of occupants, contributing factors, severity type, and lighting condition. Traffic volume data contain annual average daily traffic (AADT) data. Roadway data information on roadway cross-section and the type of roadway includes the number of lanes, lane width, shoulder width and type, median width, rural/urban designation, and functional classification.

3.2. Data conflation

Given the list of the data sources and the purpose of the data analysis, the project team developed conflated datasets by integrating information from different sources. The 2015 NPMRDS Static Files were generally produced on a quarterly basis. There are three different Static Files for 2015: January–June (2014Q3), July–October (2015Q3), and November–December (2015Q4). For example, 2015Q4 has 650 additional TMCs in Washington rural NHS roadway networks. In an exploration of the three Static Files, researchers found that over 95 percent of the NPMRDS TMCs are the same in the rural areas of the states across the three NPMRDS Static Files. Fig. 1 shows the data conflation flowchart. Two databases (NPMRDS and HSIS for 2015) were used in this study to develop the conflated database for two focus states (Ohio and Washington). The 2015 annual precipitation data from NOAA weather stations were conflated on the HSIS roadway segments. The HSIS segments with all geometric variables, crash information, and precipitation data were later conflated to the TMCs. Total segment lengths (both directions) of Ohio rural two-lane and rural multilane roadways are 1,907 miles and 1,621 miles, respectively. Total segment lengths (both directions) of Washington rural two-lane and rural multilane roadways are 3,552 miles and 521 miles respectively. It is important to note that this conflation framework is transferable to other HSIS states as well any other state's linear roadway network.

3.3. Safety performance functions

Separate models were developed for total (all) and fatal and injury crashes. Experience with the regression-based calibration of SPFs and CMFs using total (all), fatal and injury crash indicates that the calibration coefficients often vary among model types for common variables. Some of this variation is likely since geometric elements often have a different effect on all crashes than on fatal and injury crashes. When crash frequency varies systematically from county to county, district to district, and state to state because of formal and informal differences in the reporting threshold, the use of severity-based crash data to build severity-based crash prediction models may yield inaccurate results about the variable influence. Thus, the researchers developed models for two severity levels to understand the difference in variable effects. Except for curve length and radius, the interaction between the variables was not considered. As noted by Srinivasan and Bauer [14], interactions are not usually considered during SPF development. The authors mentioned that there is no easy way to identify which interactions are important and how they should be included in a model unless there is some theoretical reason for including certain interactions. Interested readers can consult Das et al. [15] for the details of the SPFs. The SPF framework for rural two-lane roadways is described below as an example.

Different variable combinations and various model forms were examined to identify the best possible relationship between the number of crashes and independent variables. The model presented below (for rural two-lane roadways) was developed by findings from several preliminary regression analyses. This model form includes variables that are intuitive, in-line with previous findings and best fits the data:

$$N = Len \times e^{b_0 + b_{aadt}\ln(AADT)} \times CMF_{lw} \times CMF_{hc} \times CMF_{sdif} \times CMF_{svar1} \times CMF_{svar2} \times CMF_{sff} \times CMF_{int} \times CMF_{prec} \quad (1)$$

with,

$$CMF_{lw} = e^{b_{lw}(w_l - 12)}$$

$$CMF_{hc} = 1.0 + b_{hc}\left(\frac{L_c}{L}\right)$$

$$CMF_{sdif} = e^{b_{sd}(SpdDiff)}$$

$$CMF_{svar1} = e^{b_{sv1}(I_{svar1})}$$

$$CMF_{svar2} = e^{b_{sv2}(I_{svar2})}$$

$$CMF_{sff} = e^{b_{sff}(SFF)}$$
$$CMF_{int} = e^{b_{int} I_{int}}$$
$$CMF_{prec} = e^{b_{prec}(p_{prec})}$$

where:

N	=	Predicted annual average crash frequency
Len	=	Segment length, miles
$AADT$	=	Average annual daily traffic, vehicles per day
CMF_{lw}	=	Crash Modification Factor (CMF) for lane width
CMF_{hc}	=	CMF for horizontal curve
CMF_{sdif}	=	CMF for speed difference between weekend and weekday
CMF_{svar1}	=	CMF for variance in hourly operating speeds
CMF_{svar2}	=	CMF for variance in monthly operating speeds
CMF_{sff}	=	CMF for free-flow speed
CMF_{int}	=	CMF for presence of an intersection on the segment
CMF_{prec}	=	CMF for precipitation
w_l	=	Average lane width in both directions (ft)
L_c	=	Total length of all horizontal curves on the segment
$SpdDiff$	=	Percent difference of operating speeds between weekend and weekday
I_{svar1}	=	Indicator variable for high variance in hourly operating speeds within a day (= 1 if hourly standard deviation is > 1 mph; = 0 otherwise)
I_{svar2}	=	Indicator variable for high variance in monthly operating speeds within a year (= 1 if monthly standard deviation is > 1 mph; = 0 otherwise)
SFF	=	Free-flow speed, mph
I_{int}	=	Indicator variable for intersection presence (= 1 if present; = 0 otherwise)
p_{prec}	=	Percent of days with precipitation.
b_j	=	Calibrated coefficients (j = hc, sd, svar1, svar2, sff, int, prec).

The inverse dispersion parameter, K (which is the inverse of the overdispersion parameter λ), is allowed to vary with the segment length. The inverse dispersion parameter is calculated using:

$$K = L \times e^{k} \quad (2)$$

where,

K	=	Inverse dispersion parameter.
k	=	Calibration coefficient for inverse dispersion parameter.

4. Implementation

The implementation of the developed decision support tool is based on R software [1] and R shiny [2] framework. The server is composed of two components: R studio Server, and shiny server. The tool is hosted as a shinyapps.io webpage. The interactive web applications are developed using R codes. The codes are uploaded in GitHub and are fully reproducible (i.e. transferable). The major contribution of this work is that the developed decision support interactive tool is scalable and transferable with options for data downloading. The software is based on the developed SPFs; however, the tool can be transferable for any new modelling as well as the calibration of the SPFs. To reproduce similar decision support tool for other states, users need to follow few simple steps: (1) conflate NPMRDS and state specific roadway networks by following the framework mentioned in Section 3.2, (2) assign precipitation data from NOAA on the conflated segments, (3) develop speed measures on different roadway facilities using NPMRDS data, (4) develop state specific SPFs for different roadway facilities using similar geometric variable, speed measures, and precipitation data to determine the expected crashes, (5) assign expected crashes on the roadway network shapefile, and (6) use 'RuralSpeedSafetyX' coding framework to develop similar tool for other states.

The tool contains a dashboard with various dropdown lists of steps to evaluate risk scoring at the segment level (direction specific). Users have the flexibility of selecting several options. The beta version has the following drop down and selection options (see Fig. 2):

- Year: 2015
- State: WA, and OH
- County: Counties in Each State
- Facility Type:
  - Rural Interstate
  - Rural Two-Lane
  - Rural Multilane
- Severity: All, and Fatal and Injury

The risk scoring of the segment is based on the expected annual crashes on the segment. The expected number of crashes is a combination of observed crashes (historical crashes) and predicted crashes (crash estimates from SPFs) with the use of weighting factors as described in the HSM. The tool offers several query options and provides estimates of expected annual crashes (total and fatal/injury) at different geographic scales, such as state, county, and facility type. The estimates are graphically displayed in a color-coded map (yellow indicates low number of annual crashes and red indicated high number of annual crashes) and are available for download in CSV format. The users need to follow few simple steps:

- Select options from drop-down panels
- Select 'Severity'
- Click 'Refresh Map' (will take some time to load the map and associated data)
- Zoom in/out (segments have hovering option that shows important features of the segment)
- Click 'Download Data' to download the data in CSV format

After selecting the options, the user needs to click the "Refresh Map" button to generate the interactive map. For example, by selecting "Year: 2015; State: WA; County: All Counties; Facility: All; Severity: All", a heatmap based on number of crashes will be generated (see Fig. 2). The map can be seen at other smaller spatial units also. For example, selection of "Year: 2015; State: WA; County: Yakima County; Facility: All; Severity: All" will allow the user to develop the map and associated data at the county level (see Fig. 3). Selection of "Year: 2015, State: WA; County: Whitman County; Facility: Two-lane; Severity: All" will allow the user to create the map and associated data at facility level of a county (see Fig. 4). Users can also go to another level down by creating the map using fatal and injury expected crashes by selecting options from 'Severity'.

The interactive map has a hovering option. Users can see associated data on a segment by dragging the mouse to that segment (see Fig. 5).

There are boundary conditions for each of the variables. If the input value exceeds, the tool will produce excessive expected crashes or risk scores. In such cases, the users need to update the data input correctly.

5. Conclusion

This study investigated the prevailing operating speeds and weather data on NHS roadways and quantified how traffic speed and weather condition interact with roadway characteristics to affect the likelihood of crashes. The inclusion of speed information expanded upon the existing state of practice by incorporating operational data as risk variables through statistical models that include speed measures to quantify highway safety risk and anticipate crash occurrence. Another contribution of the current study is the development of SPFs by direction — which is significantly different from the HSM practice. Additionally, this study has three other unique contributions:

Fig. 2. Interface of RuralSpeedSafetyX. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

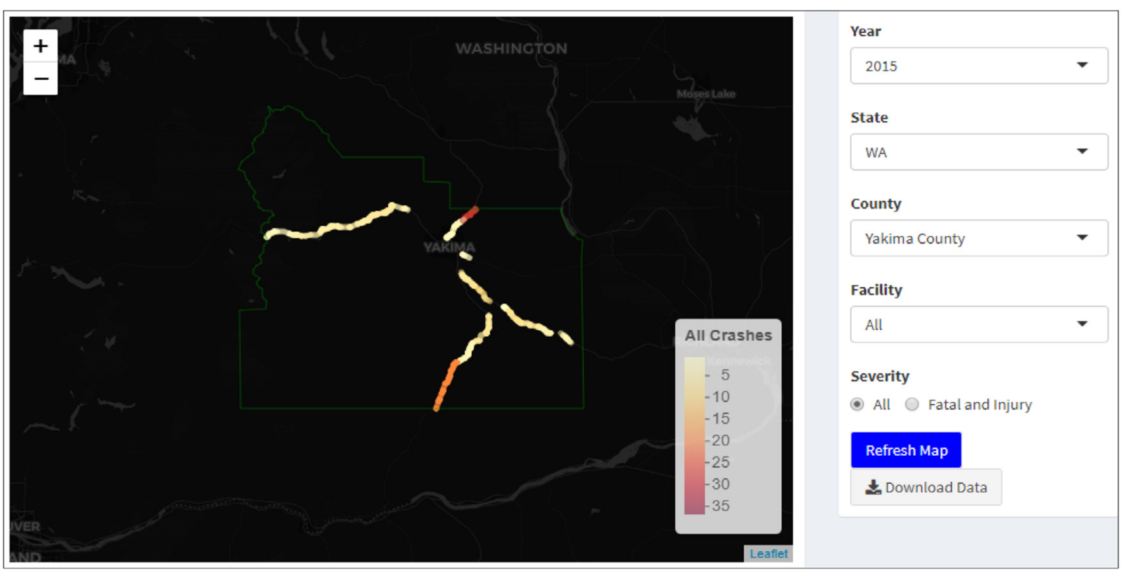

Fig. 3. Selection at county level.

- Developed the conflated dataset with traffic speed, roadway design elements, traffic volume information, and crash frequency for two states (Ohio and Washington).
- Quantified the targeted relationship between crashes and influential variables by developing best-fit models that address the impact of operating speed and weather at roadway segment levels to measure safety risk alongside traditional highway safety variables.
- Developed a scalable, flexible, and transferable decision support tool (RuralSpeedSafetyX) that can be reproduced by using newer datasets.

This study is not without limitations. First, this research used roadway segments based on NPMRDS travel time data TMC segment lengths (some of the segments are quite long compared to other segments). Further examination of the effects of segment length would improve modelling reliability. Second, the current tool is limited annual crash prediction only. Advanced models with granular analysis (e.g., daily crash prediction) can make the decision support framwork more effective.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project 'Rural Speed Safety Project for USDOT SDI' has been funded by the USDOT Safety Data Initiative (SDI). The decision support tool has been developed as a part of this project.

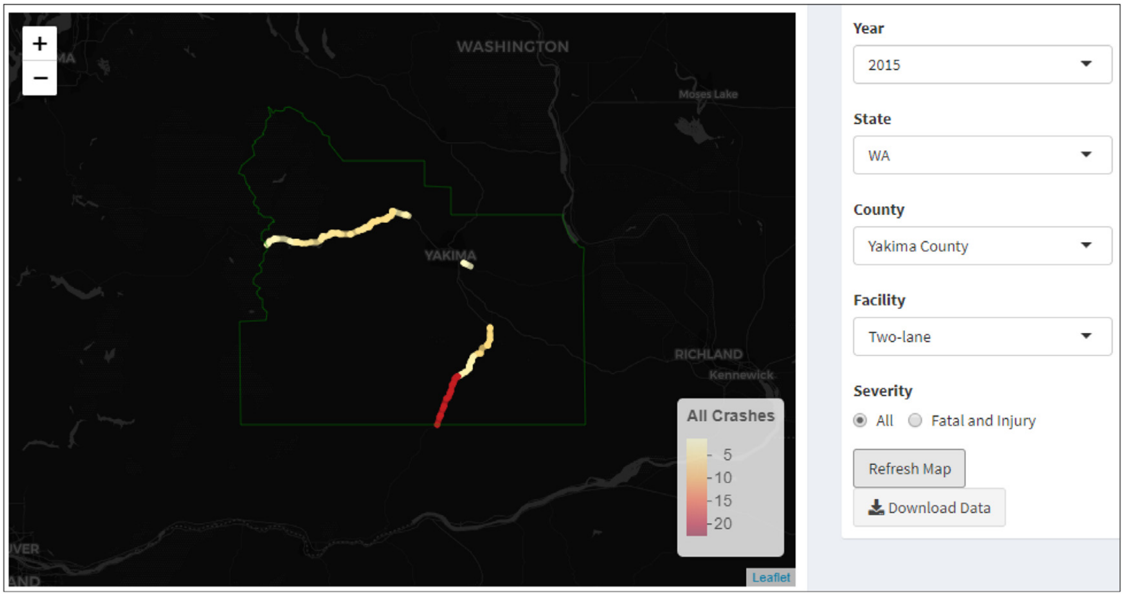

Fig. 4. Selection at county and facility type level.

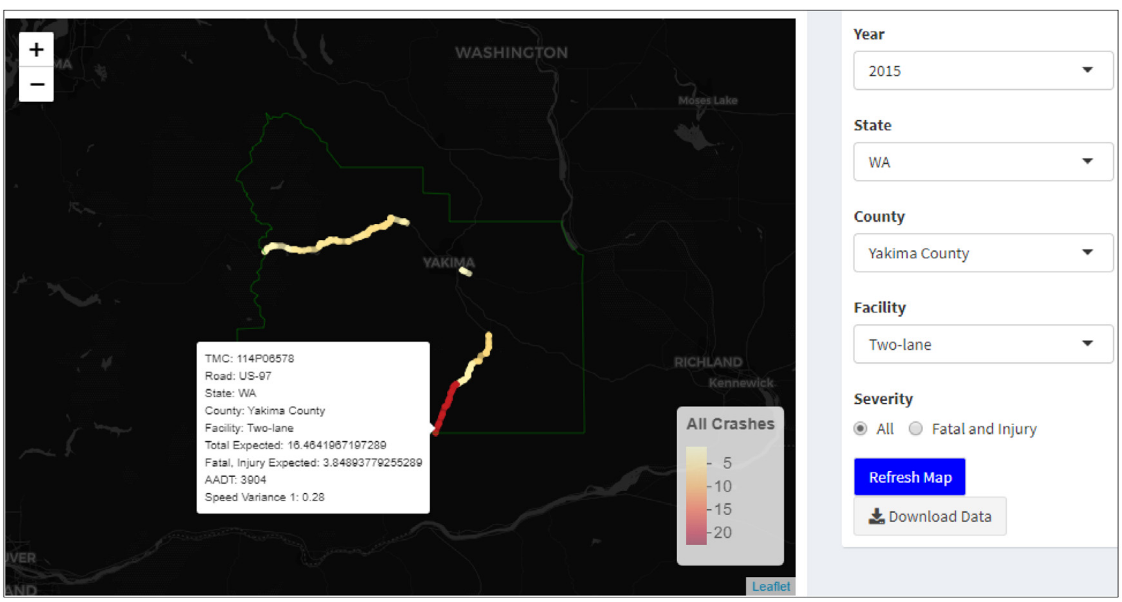

Fig. 5. Hovering option.

We like to thank Paul Teicher, Philip Sung, and David Winter for their support and important feedback. We also like to thank two anonymous reviewers for their excellent comments and suggestions.

References

[1] Open source software R project, R project. 2019, https://www.r-project.org/.
[2] Rstudio, R shiny. 2019, https://shiny.rstudio.com/.
[3] Abdel-Aty M, Radwan A. Modeling traffic accident occurrence and involvement. Accid Anal Prev 2000;32:633–42.
[4] Taylor M, Lynam D, Baruya A. The effects of drivers' speed on the frequency of road accidents. TRL REPORT, 2000, p. 421.
[5] Pei X, Wong S, Sze N. The roles of exposure and speed in road safety analysis. Accid Anal Prev 2012;48:464–71.
[6] Yu R, Abdel-Aty M, Ahmed M. Bayesian random-effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accid Anal Prev 2013;50:371–6.
[7] Gargoum S, El-Basyouny K. Exploring the association between speed and safety: A path analysis approach. Accid Anal Prev 2016;93:32–40.
[8] Yu R, Quddus M, Wang X, Yang K. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. Accid Anal Prev 2018;120:304–31.
[9] Banihashemi M, Dimaiuta M, Zineddin A, Spear B, Smadi O, Hans Z. Using linked SHRP2 RID and NPMRDS data to study speed-safety relationships on urban interstates and major arterials. Washington DC: The 98th TRB Annual Meeting; 2019.
[10] Lamm R, Choueiri E, Mailaender T. Comparison of operating speeds on dry and wet pavements of two-lane rural highways. Transp Res Rec: J Transp Res Board 1990;(1280):199–207.
[11] Jackson TL, Sharif HO. Rainfall impacts on traffic safety: Rain-related fatal crashes in texas. Geomat Nat Hazards Risk 2016;7(2):843–60.
[12] Mayora J, Pina R. An assessment of the skid resistance effect on traffic safety under wet-pavement conditions. Accid Anal Prev 2009;41(4):881–6.
[13] Najafi S, Flintsch G, Medina A. Linking roadway crashes and tire-pavement friction: a case study. Int J Pavement Eng 2015;18(2):119–27.

[14] Srinivasan R, Bauer K. Safety performance function development guide: developing jurisdiction specific SPFs. Report No. FHWA-SA-14-005, 2013.

[15] Das S, Geedipally S, Wu L, Fitzpatrick K, Avelar R, Banihashemi M, et al. Rural speed safety project for USDOT safety data initiative. USDOT final report, 2019.