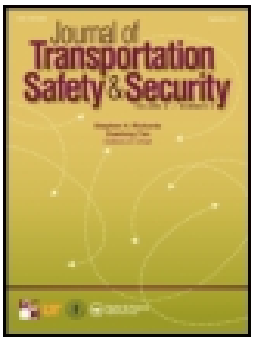

Journal of Transportation Safety & Security

ISSN: 1943-9962 (Print) 1943-9970 (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Patterns of rainy weather crashes: Applying rules mining

Subasish Das, Anandi Dutta & Xiaoduan Sun

To cite this article: Subasish Das, Anandi Dutta & Xiaoduan Sun (2019): Patterns of rainy weather crashes: Applying rules mining, Journal of Transportation Safety & Security, DOI: 10.1080/19439962.2019.1572681

To link to this article: https://doi.org/10.1080/19439962.2019.1572681

Published online: 26 Feb 2019.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2019.1572681

Patterns of rainy weather crashes: Applying rules mining

Subasish Das[a], Anandi Dutta[b], and Xiaoduan Sun[c]

[a]Texas A&M Transportation Institute, Texas A&M University, College Station, Texas, USA; [b]Computer Science and Engineering Department, Texas A&M University, TAMU, College Station, Texas, USA; [c]Civil Engineering Department, University of Louisiana, Lafayette, Louisiana, USA

ABSTRACT

Driving in rainy weather is considered as one of most hazardous conditions for driving. There is a need for appropriate countermeasures focusing on the reduction of these crashes, but measuring the key factors under such conditions is very challenging. With a humid subtropical climate, the annual precipitation in Louisiana is about 64 inches, twice above the national average. Approximately 11% of total crashes in Louisiana happen during rainy weather, and nearly 25% of total fatal crashes happen in rainy weather annually. This study applied association rules mining to discover crash patterns during rainy weather with Louisiana crash data (2004–2011). The findings showed that "single-vehicle run-off road crash" is predominant during rainy weather and is associated with grade-curve aligned roadways, curved roadways, and roadways with no streetlights at night. In rainy condition, no injury and sideswipe crashes are also significant in numbers. Moderate injuries are dominant in single-vehicle crashes. Roadways with poor illumination are associated with straight, level aligned roadways in rainy weather crashes. Drivers (age 15 – 44) are vulnerable in run-off crashes when the roadways had poor illumination and curves during rainy condition. The findings of this study will be beneficial for safety practitioners.

KEYWORDS
road safety; data mining; association rules; market basket analysis; nonparametric statistical method

1. Introduction

Rainy weather is considered as one of the most hazardous conditions for driving. Drivers tend to adapt their driving behavior to adjust the conditions presented by inclement weather. Depending on the surroundings, drivers drive more vigilantly by keeping longer headways, reducing operating speeds, or being more cautious (Das, Brimley, Lindheimer, and Pant, 2016). Due to the visual obstruction from rainfall and loss of surface friction, most vehicles slow down during rainfall, but crashes still occur, which

CONTACT Subasish Das s-das@tti.tamu.edu Associate Transportation Researcher, Texas A&M Transportation Institute, 3135 TAMU, College Station, TX 77843-3135, USA.
Color versions of one or more figures in the article can be found online at www.tandfonline.com/utss.

may be associated with various contributing factors. Measuring the added risk and identifying key crash contributing factors during rainy weather has been challenging. The total reported number of traffic crashes in rainy weather in Louisiana is 11,398 in 2011 and 10,204 in 2010. Approximately, 11% of yearly total crashes in Louisiana happened during rainfall, and nearly 25% of total fatal crashes in Louisiana happened in rainy weather in the state. To reach "Destination Zero Deaths" set by Louisiana Highway Safety Strategies, it is critical to reduce the number of crashes and crash severity under the rainy condition (Louisiana Highway Safety Commission, 2017).

There are several ways to identify crash risk factors. The parametric models work well if the assumptions and model format are accurate to reflect the underlying relationships between dependent and independent variables. Violation of any assumption could lead to flawed or at least inadequate estimations. Nonparametric statistical methods like data mining techniques have been receiving increased attention from researchers in traffic safety because of no-predefined assumptions. Additionally, factor identification task involves categorical data analysis. Crash frequency analysis and crash severity analysis are two major transportation safety research areas, which have been extensively studied. Lord and Mannering (2010) conducted a comprehensive review of state-of-the-art crash frequency studies and their limitations. Savolainen, Mannering, Lord, and Quddus (2011) conducted a similar study on crash-injury severities. Mannering and Bhat (2014) summarized analytic methods used in these two transportation research areas and provided future directions. Interested readers can consult these studies to understand the breadth and depth of traffic safety analysis.

This method is concerned with the identification of interesting patterns from a massive data set. This research aims to understand the ramifications of rainy condition on safety from a perspective of roadway environment and driver characteristics. Due to the nature of the data and associated research objectives, association rules mining is a good fit for the analysis.

2. Literature review

It is known that inclement weather plays a key role in crash occurrence due to reduction of friction and low visibility. In past, there are limited number of studies that incorporated inclement weather in traffic safety analysis. During the last few years, there has been a surge of studies that incorporate inclement weather in the safety analysis. Rainfall is considered as one of the most significant inclement weather events. The current literature review focuses on the traffic safety studies related to rainy weather. In a study conducted in 1971, Campbell (1971) examined the extent of the wet pavement crash scenarios in the U.S. Haghighi-Taleb (1973) investigated the association between rainfall and crashes in two cities (Huddersfield and London). The findings showed that moderate and heavy rainfall have similar effects on crash rates. Satterthwaite (1976) assessed the seasonal and weather effects on crash frequencies in California. The findings of this study showed that weather is one of the major contributing factors affecting crash frequencies. Sherretz and Farhar (1978) used rainfall data and crash reports for seven southern Illinois cities to determine the relationship between rainfall and traffic crashes. This study determined a statistically significant linear trend between mean number of crashes and amount of rainfall. Brodsky and Hakkert (1988) examined the added risk of a crash during rainy weather. They used data from two sources: the traffic injury (fatal and nonfatal) crash file in Israel (1979–1981), and 1983 to 1984 fatal crash data in the U.S. using Fatality Analysis Reporting System (FARS). An empirical study was conducted in providing the evidence of crash risk during and following rain events on Calgary and Edmonton, Canada (Andrey & Yagar, 1993). The findings showed that the overall crash risk during rainfall conditions was found to be 70% higher than normal. Eisenberg (2004) investigated the association between precipitation and traffic crashes in the U.S. during 1975–2000. The results showed that 1 cm of precipitation increases the fatal crash rate by 3% if exactly 2 days have passed since the last precipitation and by about 9% if more than 20 days have passed. Keay and Simmonds (2006) examined the impact of rainfall on daily crashes in the metropolitan area of Melbourne, Australi,a during 1987 to 2002. The findings showed that rainfall occurring after a dry spell has an enhanced effect on the volume-normalized crash counts as the spell duration increases. Jung, Qin, and Noyce (2010) examined the effects of rainfall on the severity of single-vehicle crashes on Wisconsin interstate highways utilizing polychotomous response models. The results revealed that rainfall intensity, wind speed, roadway terrain, driver's gender, and safety belt were statistically significant. Mills, Andrey, and Hambly (2011) conducted a study to examine the rainfall related collision risk by using police records and comprehensive insurance claim data for Winnipeg, Canada, over the period 1999 to 2001. Both data sets showed similar results—precipitation increases the risk of injury collision and risk of injury relative to corresponding dry weather control periods. The results of the study conducted by Sun, Hu, Habib, and Magri (2011) indicated a higher crash risk and a higher injury risk during rain. The risk potentials vary depending upon the type of highway, location of the highway, time of day, crash severity, and crash characteristics. Xu, Wang, and Liu (2013) developed separate crash risk prediction models for different weather conditions using crash data, weather data, and traffic data on the I-880N freeway in California. The findings showed that the traffic flow characteristics contributing to crash risk were different across different weather conditions. The speed difference between upstream and downstream stations was also found to be significant. Ahmed, Abdel-Aty, Lee, and Yu (2014) examined the viability of using airport weather information in real-time road crash risk assessment in locations with inclement weather and recurrent fog problems. Jaroszweski and McNamara (2014) used city-wide weather radar approach to rainfall quantification and matched-pairs analysis to examine the influence of rainfall on crashes in the U.K. cities of Manchester and Greater London during 2008 to 2011. Theofilatos and Yannis (2014) provided a systematic review of the effect of traffic and weather characteristics on roadway safety. Black, Villarini, and Mote (2017) used daily precipitation data and crash data from six U.S. states (Arkansas, Georgia, Illinois, Maryland, Minnesota, Ohio) for the period 1996 to 2010. A matched pair analysis showed that there is a statistically significant increase in crash and injury rates during rainfall days. Jackson and Sharif (2016) used fatal crash data and geospatial data to examine the temporal and spatial distribution of rainfall related crashes in Texas from 1982 to 2011. Results indicate that rain is a significant contributor in few counties but at less than 95% confidence in some of the wetter counties. The study conducted by Das, Brimley, Lindheimer, and Zupancich (2017) shows that the likelihood of a crash increase during periods of rainfall, despite the tendency for less traffic and for lower speeds to prevail during these times. Ghasemzadeh and Ahmed (2017) used SHRP-2 Naturalistic Driving Study (NDS) data to identify patterns in driver behavior and performance in rainy weather conditions. In a follow-up study, the impact of heavy rain on speeding and headway behavior was examined (Ahmed & Ghasemzadeh, 2018). Wu, Abdel-Aty, and Lee (2018) conducted another visibility and inclement weather-related study using data from two regions in Florida. Lee, Chae, Yoon, and Yang (2018) used the rainfall and traffic crash data (during 2007–2015) for Seoul by using structural equation modeling. The findings show that compact cars, young drivers, female drivers, heavy rain, deep water, and roads with a long drainage length have high likelihood with an increase in the level of crash severity. Omranian, Sharif, Dessouky, and Weissmann (2018) used crash-based matched-pairs analysis approach to examine the impact of rainfalls on the crash risk on Texas roadways. The overall findings show that rainfall increases crash risk by about 57% in Texas.

The essential method behind transportation safety analysis is to identify the relationship between a large variety of variables and crash occurrence or crash severity. To achieve this goal, a variety of methods have been applied and accepted by transportation researchers. Those methods include logistic regression(Al-Ghamdi, 2002; Dissanayake & Lu, 2002), decision trees (Chung, 2013; Figueira, Pitombo, de Oliveira, & Larocca 2017; Khan, Bill, & Noyce, 2015; Saha, Alluri, & Gan, 2015), support vector machines (Li, Lord, Zhang, & Xie, 2008; Chen, Zhang, Qian, Tarefder, & Tian, 2016), text mining (Brooks, 2008; Brown, 2016), multiple correspondence analysis (Das & Sun, 2017; Das, Brimley, Lindheimer & Pant, 2016; Das, Avelar, Dixon, & Sun, 2018; Jalayer, Pou-Rouholamin, & Zhou, 2018), association rules mining (Geurts, Thomas, & Wets, 2005; Marukatat, 2007; He et al., 2008; Pande, & Abdel-Aty, 2009; Kumar, & Toshniwal, 2016; Weng, Zhu, Yan, & Liu, 2016; Ait-Mlouk, Gharnati, & Agouti, 2017; Nitsche, Thomas, Stuetz, & Welsh, 2017; Das, Dutta, Jalayer, Bibeka, & Wu, 2018; Das, Dutta, Avelar, Dixon, Sun, & Jalayer, 2018), association rules negative binomial miner (Das, Minjares-Kyle, Avelar, Dixon, & Bommanayakanahalli, 2017), and deep learning (Gibert, Patel, & Chellappa, 2017; Das, Dutta, Dixon, Minjares-Kyle, & Gillette, 2018).

In the recent years, many studies examined association rules mining in traffic crash analysis. Geurts, Thomas, and Wets (2005) used association rules mining to obtain a descriptive analysis of "black" zones. In this study frequent item sets are generated to identify crash circumstances that frequently occur together to find out which factors explain the occurrence of the crashes in "black" zones. In his study, Marukatat (2007) applied association rules to real traffic-crash data collected from local police stations. This study found out candidate rules offered significant insights into the phenomena of safety improvement. He et al. (2008) conducted multidimensional association rules model for the freeways of China. This study also presented preventive measures in reducing crashes. Pande and Abdel-Aty (2009) used association rules mining in safety analysis by generating closely associated rules. Kumar and Toshniwal (2016) used association rules mining to characterize crash locations. Weng, Zhu, Yan, and Liu (2016) used rules mining approach to identify word zone-related crash patterns. Ait-Mlouk, Gharnati, and Agouti (2017) used this technique to generate insights and sufficient knowledge to enable decision makers to make the right decision to avoid dangerous routes and improve road safety. Nitsche, Thomas, Stuetz, and Welsh (2017) used 1056 junction crashes in the U.K. to determine robust precrash scenario patterns by using association rule mining. Das, Minjares-Kyle, Avelar, Dixon, and Bommanayakanahalli (2017) used the second Strategic Highway Research Program's (SHRP-2) Roadway Inventory Database (RID) crash data for Florida rural roadways to investigate improper passing related crashes by using association rules negative binomial (NB) miner. Das, Dutta, Jalayer, Bibeka, and Wu (2018) applied association rules 'Eclat' algorithm to determine the significant rules from wrong way crashes in Louisiana. Das et al. (2018) applied supervised association rules to determine key patterns in pedestrian crashes.

Many recent studies applied data fusion techniques in developing a data set that can examine the association between crash and rainfall. However, the success of the data fusion depends on rigorous quality check. The generalized findings from these studies fluctuate due to the deviation from the scenario specific assumptions. This study places emphasis on examining the effect of rainfall on the crash occurrence based on the crash database of Louisiana. The study is limited to rainfall-related patterns with crash outcomes, rather than the effect of wet pavement, visibility, and the reduction of friction. Unlike earlier studies, this project is able to identify the patterns of contributing factors during the crash occurrences in the rainy environment. The current study is unique in the methodological aspect as none of the prior studies have applied association rules mining in determining significant patterns in rainfall related crashes.

3. Methodology

3.1. Association rules mining

Data mining is the process of identifying valid and understandable patterns in the data set. It helps in extracting and refining valuable knowledge from large data sets. Data mining involves machine learning, statistical knowledge, modeling concepts, and database management. It is important to note that data mining is concerned only with relationships among variable categories. The methods can be classified into two main sections: descriptive and predictive. Association rules mining, a descriptive analytics, discovers significant rules showing variable category conditions that occur frequently together in a data set.

Many algorithms can be used to discover association rules from data to extract useful patterns. Apriori algorithm is one of the most widely used and famous techniques for finding association rules (Agrawal, Imielinski, & Swami, 1993). Due to the explorative and eloquent nature, intelligible representation and visualization of the found patterns and models are essential for the successful mining process to make the results easy to understand. One important feature of the technique is that no variables are assigned as dependent or independent. The apriori algorithm for searching association rules is easy to interpret, and the computations used are straightforward.

A frequent itemset generation algorithm digs out frequently occurring itemsets, subsequences, or arrangements from large data sets. Frequent itemsets mining has been applied in many branches of science, for example, social science, web mining, and bioinformatics. A set of definitions are given here before demonstrating the method with an example. Let $I = \{i_1,$

i_2, … i_m} be a set of items (e.g., a set of crash categories for a particular crash record) and $C = \{c_1, c_2, \ldots, c_n\}$ be a set of database crash information (transaction) where each crash record c_i contains a subset of items chosen from I. A set of items is referred to as an itemset. An itemset that contains k items is considered as a k-itemset.

An association rule can be expressed as X → Y, where X and Y are disjoint itemsets. Here, X is known as the antecedent and Y is the consequent. The strength of the association rule can be measured in terms of the values of support, confidence, and lift. The equations of support are listed in Eq. 1–Eq. 3 (Dutta, 2016).

$$S(X) = \frac{\sigma(X)}{N} \tag{1}$$

$$S(Y) = \frac{\sigma(Y)}{N} \tag{2}$$

$$S(X \rightarrow Y) = \frac{\sigma(X \cap Y)}{N} \tag{3}$$

Where,

$\sigma(X)$ = Number of incidents with X antecedent
$\sigma(Y)$ = Number of incidents with Y consequent
$\sigma(X \cap Y)$ = Number of incidents with both X antecedent and Y consequent
N = Total number of incidents
S(X) = Support of antecedent
S(Y) = Support of consequent
S(X → Y) = Support of the association rule (X → Y)

The equations of confidence and lift are listed in Eq. 4 and 5. Confidence measures the reliability of the inference of a generated rule. A higher confidence for a X → Y indicates that presence of Y is highly visible in the transactions having X. The lift of the rule makes an association with the frequency of co-occurrence of the antecedent and the consequent to the expected frequency of co-occurrence.

$$C(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X)} \tag{4}$$

$$L(X \rightarrow Y) = \frac{S(X \rightarrow Y)}{S(X).S(Y)} \tag{5}$$

Where,

C(X → Y)= Confidence of the association rule (X → Y)
L(X → Y)= Lift of the association rule (X → Y)

The more the lift value exceeds from 1, the stronger the dependency becomes. It is desirable for rules to have a large confidence factor, high level of support, and a lift value greater than 1. Because some events of interest in traffic safety analysis are very rare (e.g., fatal crashes), the support for some rules of interest could be quite low. It essentially means that lift value is more important for determining strength of an association rule than the other two criteria. Hence, in the present application the rules should be evaluated based on the "lift" values. The rules "discovered" by the algorithm still need to have support greater than a minimum threshold.

3.2. Exploratory data analysis

To identify important contributing factors for rainy weather crashes (crashes that are defined as rainy in the weather variable) in Louisiana, a large data set containing eight years of crash records (2004–2011) was obtained from the Louisiana Department of Transportation and Development (LADOTD). The data was stored as an unsorted and unmanageable format in Microsoft Access database tables. Every crash record has many variables, and the detailed information is stored in separate data tables such as crash, driver, and vehicle. Cross-table in-depth analysis would be useful to find out hidden crash patterns, and association rules mining provides promise in exploring these patterns. The data preparation task involved the following steps:

- Combine eight years of crash data from LADOTD police reported crashes.
- Merge driver and roadway condition data with the crash data by matching with the crash identification number.
- Prepare the rainy weather crash database by filtering rainy weather in the weather condition variable.

Traffic crash databases contain many variables that can make the outcomes of the rules uninterpretable. The crash records with missing and questionable information were removed. For example, 300 crashes have driver age listed as 200, which were removed. The final database contains 58,288 rainy weather–related crash records occurred on all types of roadway functional classes (interstate to local roadway). The general distribution of the important variables is listed in Table 1. It is important to note crash record includes information on all vehicles and drivers as well as occupants. This study only considers at-fault (driver responsible for a crash occurrence) driver information for the final analysis.

To focus on the meaningful analysis, a set of key variables are selected such as the information on crash timing (day of the week), roadway

Table 1. Distribution of rainy weather crashes by key variables.

Categories	Frequency	Percentage	Categories	Frequency	Percentage
Alignment			Collision Type		
Straight-level	39,546	67.85	Rear end	11,350	19.47
Curve-level	11,835	20.30	Right angle	5,096	8.74
Straight-level-elevated	750	1.29	Sideswipe- same direction	1,962	3.37
On grade-straight	1,999	3.43	Single vehicle	32,357	55.51
On grade_curve	2,231	3.83	Left turn- opposite direction	1,316	2.26
Curve-level-elevated	783	1.34	Left turn- angle	1,869	3.21
Hillcrest-straight	801	1.37	Left turn- same direction	755	1.30
Hillcrest-curve	230	0.39	Head-on	1,108	1.90
Dip, hump-straight	94	0.16	Right turn- opposite direction	127	0.22
Dip, hump-curve	19	0.03	Right turn- same direction	382	0.66
Lighting			Sideswipe- opposite direction	1,966	3.37
Daylight	34,660	59.46	Day of week		
Dark- continuous street light	1,473	2.53	Weekend	26,141	44.85
Dark- no street light	19,170	32.89	Weekday	32,147	55.15
Dark- street light at intersection Only	974	1.67	Gender		
Dusk	957	1.64	Male	37,542	64.41
Dawn	1,054	1.81	Female	20,746	35.59
Driver Age			Severity		
15–24	18,977	32.56	No Injury	32,613	55.95
25–34	13,107	22.49	Complaint	18,983	32.57
35–44	9,501	16.30	Moderate	5,913	10.14
45–54	8,121	13.93	Severe	345	0.59
55–64	4,780	8.20	Fatal	434	0.74
64–75	2,289	3.93			
75 plus	1,513	2.60			

characteristics (alignment, lighting), human factor (driver gender and age), and crash characteristics (crash severity and collision type). The variable selection procedure primarily considered findings from the past studies. A variable is taken out from further consideration if one of its attributes exceeds over 70% (determined by examining number of rules by using larger number of variables) in frequency. If these variables are not taken out from analysis, redundant rules with high lift values will be generated. For example, pavement type is not considered for the next level of variable selection as one of its attributes (asphalt) exceeds 70% in frequency distribution. Additionally, this study is focused on rainfall-related patterns with crash outcomes, rather than the effect of wet pavement, visibility, and friction reduction. Thus, the removal of these variables from analysis would not affect the research goals.

Random forest algorithm was later applied to determine the significant factors from the primary list of variables. As the primary concern of this study is to identify patterns from rainy weather crashes, variable selection methodology is not described in this study. For the association rules, there are various settings required to be altered for significant findings. The minimum support and confidence are essential to generate the important rules. After a significant number of trials and errors, the minimum support for the rules was considered as 1% with the minimum confidence of 60%. One percent of minimum support means that no item or set of items will be considered frequent for the first analysis if it does not appear in at least 583 traffic crashes (1% of total 58,288 crash records). It may be rather argued that the choices for the values of these parameters are subjective, which is partly true. However, a trial-and-error experiment indicates that setting minimum support too low will result in exponential growth of the number of items in the frequent item sets. In contrast, by choosing a support parameter too high, the algorithm will be capable of generating a small number of rules. The minimum confidence value of 60% indicates that a rule is considered reliable when the consequent of the rule occurs at least six out of ten times that the antecedent appears. By choosing different confidence values, a trial-and-error experiment showed that this parameter value gives rather stable results concerning the amount of rules generated by the algorithm. The purpose of postprocessing the association rules set is to identify the subset of interesting rules in a generated set of noteworthy rules.

4. Results and discussion

The association rules were generated in this study by using 'arules' package in software R (Hahsler, Buchta, Gruen, & Hornik, 2018). The final analysis demonstrates that the data set has 58,288 rows with 43 items/attributes and a density of 0.163. The frequency of the items is shown in Figure 2. The top five frequent items in the dataset are alignment = straight-level, driver gender = male, lighting = daylight, severity = no injury, and collision type = single vehicle. The frequency of the rules generated for different itemsets, and the statistics of support, confidence and lift are displayed in Figure 1 and Table 2, respectively. Rules are created for two phases. When

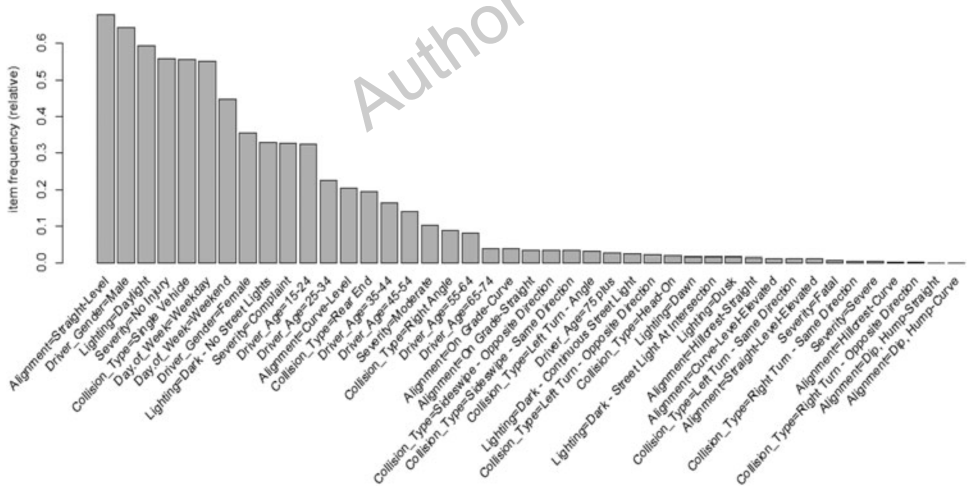

Figure 1. Item frequency plot.

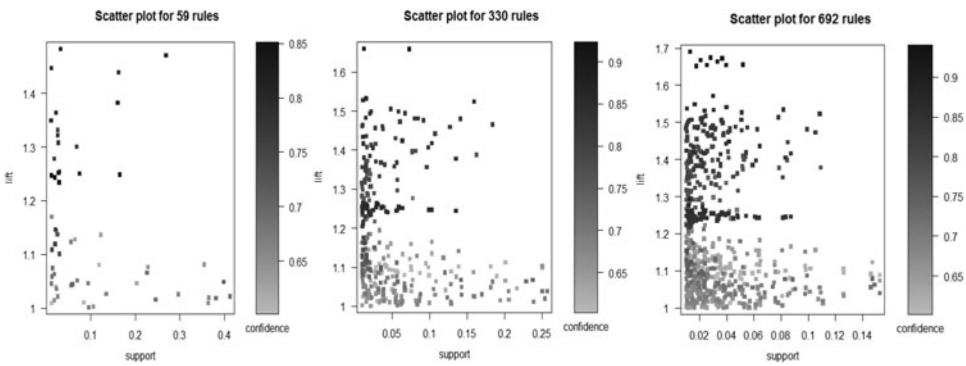

Figure 2. Scatter plot of the generated rules (a) 2-itemsets, (b) 3-itemsets, (c) 4-itemsets. (a) 2-itemsets; (b) 3-itemsets; (c) 4-itemsets.

Table 2. Summary chart of the association rules mining.

				Lift $\geq$ 1								
Support$_{min}$ = 1%, Cofidence$_{min}$ = 60%				Support			Confidence			Lift		
Minlen	Maxlen	Rules (all)	Rules (Lift $\geq$ 1)	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
1	1	2	2	0.644	0.661	0.678	0.644	0.661	0.678	1.000	1.000	1.000
2	2	81	59	0.010	0.109	0.412	0.600	0.715	0.849	1.002	1.147	1.482
3	3	423	330	0.010	0.054	0.257	0.601	0.722	0.922	1.001	1.176	1.660
4	4	871	692	0.010	0.033	0.152	0.600	0.721	0.939	1.000	1.190	1.691
5	5	767	622	0.010	0.022	0.092	0.601	0.723	0.939	1.001	1.207	1.693
6	6	217	179	0.010	0.015	0.036	0.600	0.720	0.928	1.000	1.217	1.934
7	7	6	6	0.011	0.011	0.012	0.631	0.740	0.843	1.072	1.249	1.471
All	All	2,367	1,890	0.010	0.034	0.678	0.600	0.722	0.939	1.000	1.194	1.934

the minimum threshold of the lift is selected as one, the total counts of the rules are lowered. For 1-itemset value, only two rules are produced; for 7-itemsets value, total counts of the rules are 6. The maximum number of rules is generated for 4-itemsets. In most cases it is very difficult to interpret itemsets with higher numbers. In this study, the result analysis was limited to 4-itemsets for easier interpretation.

The first 50 rules of 2-itemsets are listed in Table 3. From the first four rules, it is seen that rainy weather crashes involving either curve-aligned roadways or dark with no street lights are mostly single-vehicle run-off-road (ROR) crashes. The National Highway Traffic Safety Administration (NHTSA) study also showed that (Liu and Subramanian, 2009) ROR crashes are more likely to occur during inclement weather condition. When looking at the interpretation of the association rules, the top 13 rules listed in Table 5 express that the top rules with higher lift mostly relate to collision type and temporal characteristics (e.g., daylight). The variable category single-vehicle crash is found as "consequent" for top four rules. It may indicate that crashes in rainy weather are associated with run-off crashes due to the visual obstruction. The second rule {Lighting = Dark - No Street Lights} => {Collision Type = Single Vehicle} is associated with

Table 3. First fifty rules for 2-itemsets.

No.	Antecedent		Consequent	Supp.	Cofid.	Lift	Count
1	Alignment = On Grade-Curve	>	Collision_Type = Single Vehicle	0.031	0.823	1.482	1807
2	Lighting = Dark - No Street Lights	>	Collision_Type = Single Vehicle	0.268	0.816	1.47	15621
3	Alignment = Curve-Level-Elevated	>	Collision_Type = Single Vehicle	0.011	0.803	1.447	641
4	Alignment = Curve-Level	>	Collision_Type = Single Vehicle	0.162	0.799	1.439	9443
5	Collision_Type = Rear End	>	Lighting = Daylight	0.16	0.822	1.383	9326
6	Driver_Age = 75 plus	>	Lighting = Daylight	0.021	0.811	1.364	1224
7	Collision_Type = Left Turn - Same Direction	>	Lighting = Daylight	0.01	0.803	1.35	583
8	Collision_Type = Sideswipe - Same Direction	>	Severity = No Injury	0.025	0.745	1.332	1457
9	Collision_Type = Left Turn - Angle	>	Lighting = Daylight	0.025	0.786	1.322	1457
10	Collision_Type = Sideswipe - Same Direction	>	Lighting = Daylight	0.026	0.778	1.308	1515
11	Collision_Type = Right Angle	>	Lighting = Daylight	0.068	0.773	1.301	3964
12	Collision_Type = Left Turn - Opposite Direction	>	Lighting = Daylight	0.017	0.76	1.278	991
13	Driver_Age = 65–74	>	Lighting = Daylight	0.029	0.746	1.255	1690
14	Collision_Type = Left Turn - Angle	>	Alignment = Straight-Level	0.027	0.85	1.252	1574
15	Collision_Type = Right Angle	>	Alignment = Straight-Level	0.074	0.848	1.25	4313
16	Collision_Type = Rear End	>	Alignment = Straight-Level	0.165	0.847	1.249	9618
17	Collision_Type = Left Turn - Same Direction	>	Alignment = Straight-Level	0.011	0.846	1.247	641
18	Collision_Type = Left Turn - Opposite Direction	>	Alignment = Straight-Level	0.019	0.843	1.243	1107
19	Collision_Type = Sideswipe - Same Direction	>	Alignment = Straight-Level	0.028	0.837	1.234	1632
20	Lighting = Dawn	>	Collision_Type = Single Vehicle	0.012	0.65	1.171	699
21	Driver_Age = 75 plus	>	Alignment = Straight-Level	0.02	0.778	1.147	1166
22	Collision_Type = Sideswipe - Opposite Direction	>	Lighting = Daylight	0.023	0.68	1.144	1341
23	Collision_Type = Sideswipe - Opposite Direction	>	Driver_Gender = Male	0.025	0.732	1.137	1457
24	Collision_Type = Rear End	>	Day.of_Week = Weekday	0.122	0.627	1.137	7111
25	Severity = Moderate	>	Collision_Type = Single Vehicle	0.064	0.626	1.128	3730
26	Driver_Age = 55–64	>	Lighting = Daylight	0.055	0.668	1.124	3206
27	Lighting = Dark - Continuous Street Light	>	Alignment = Straight-Level	0.019	0.76	1.12	1107
28	Lighting = Dark - Street Light At Intersection	>	Alignment = Straight-Level	0.013	0.753	1.109	758
29	Driver_Age = 65–74	>	Alignment = Straight-Level	0.029	0.747	1.101	1690
30	Day.of_Week = Weekday	>	Lighting = Daylight	0.355	0.643	1.081	20692
31	Collision_Type = Rear End	>	Severity = No Injury	0.118	0.604	1.08	6878
32	Driver_Gender = Female	>	Lighting = Daylight	0.228	0.64	1.077	13290
33	Collision_Type = Head-On	>	Driver_Gender = Male	0.013	0.692	1.075	758
34	Lighting = Dark - No Street Lights	>	Driver_Gender = Male	0.226	0.687	1.066	13173
35	Lighting = Dark - Continuous Street Light	>	Driver_Gender = Male	0.017	0.685	1.064	991
36	Lighting = Dawn	>	Driver_Gender = Male	0.012	0.682	1.059	699
37	Severity = Moderate	>	Driver_Gender = Male	0.069	0.677	1.052	4022
38	Severity = No Injury	>	Alignment = Straight-Level	0.398	0.712	1.049	23199
39	Driver_Age = 45–54	>	Driver_Gender = Male	0.094	0.674	1.047	5479
40	Severity = Complaint	>	Lighting = Daylight	0.011	0.674	1.046	641
41	Lighting = Dark - Street Light At Intersection	>	Driver_Gender = Male	0.203	0.622	1.046	11832
42	Driver_Age = 55–64	>	Alignment = Straight-Level	0.058	0.709	1.045	3381
43	Driver_Age = 55–64	>	Driver_Gender = Male	0.055	0.672	1.043	3206
44	Alignment = On Grade-Straight	>	Lighting = Daylight	0.021	0.618	1.04	1224
45	Collision_Type = Sideswipe - Same Direction	>	Driver_Gender = Male	0.022	0.667	1.036	1282
46	Alignment = Curve-Level	>	Driver_Gender = Male	0.134	0.661	1.027	7811
47	Day.of_Week = Weekend	>	Driver_Gender = Male	0.296	0.661	1.026	17253
48	Alignment = On Grade-Curve	>	Driver_Gender = Male	0.412	0.694	1.022	24015
49	Lighting = Daylight	>	Alignment = Straight-Level	0.025	0.658	1.022	1457
50	Alignment = Straight-Level	>	Lighting = Daylight	0.412	0.608	1.022	24015

15,621 crashes. The consequent of this rule verifies that single-vehicle crashes under rainy weather are closely associated with roadway segments without light at night. Another study showed that crashes happening in signalized intersections with bad weather and dark street lighting had a significantly higher probability of severe injury (Abdel-Aty, 2003). Rules 14 to 19 in Table 5 indicate the relationship between collision type and straight-level crashes. Nearly 68% of the rainy weather crashes happened in straight-level roadways. As the variable *alignment* is skewed to a particular category (straight-level), more rules are associated with this category. The

Table 4. Association rules for 3-itemsets.

No.	Antecedent		Consequent	Supp.	Conf	Lift	Count
1	Alignment = On Grade-Curve, Lighting = Dark - No Street Lights	>	Collision_Type = Single Vehicle	0.013	0.922	1.660	752
2	Alignment = Curve-Level, Lighting = Dark - No Street Lights	>	Collision_Type = Single Vehicle	0.073	0.921	1.659	4,237
3	Day.of_Week = Weekend, Alignment = On Grade-Curve	>	Collision_Type = Single Vehicle	0.016	0.851	1.533	907
4	Alignment = On Grade-Curve, Driver_Age = 15–24	>	Collision_Type = Single Vehicle	0.012	0.848	1.528	675
5	Lighting = Dark - No Street Lights, Driver_Age = 35–44	>	Collision_Type = Single Vehicle	0.048	0.836	1.506	2,807
6	Alignment = Curve-Level, Severity = Moderate	>	Collision_Type = Single Vehicle	0.022	0.832	1.499	1,311
7	Alignment = Curve-Level, Driver_Age = 15–24	>	Collision_Type = Single Vehicle	0.059	0.832	1.499	3,418
8	Lighting = Dark - No Street Lights, Driver_Age = 25–34	>	Collision_Type = Single Vehicle	0.067	0.830	1.495	3,921
9	Alignment = On Grade-Curve, Driver_Gender = Male	>	Collision_Type = Single Vehicle	0.021	0.829	1.494	1,218
10	Lighting = Dark - No Street Lights, Driver_Gender = Female	>	Collision_Type = Single Vehicle	0.085	0.822	1.481	4,935
11	Lighting = Dark - No Street Lights, Driver_Age = 15–24	>	Collision_Type = Single Vehicle	0.092	0.820	1.477	5,353
12	Lighting = Dark - No Street Lights, Driver_Gender = Male	>	Collision_Type = Single Vehicle	0.184	0.814	1.466	10,712
13	Lighting = Dark - No Street Lights, Driver_Age = 45–54	>	Collision_Type = Single Vehicle	0.037	0.811	1.461	2,172
14	Alignment = On Grade-Curve, Driver_Gender = Female	>	Collision_Type = Single Vehicle	0.011	0.811	1.461	618
16	Alignment = Curve-Level, Driver_Age = 35–44	>	Collision_Type = Single Vehicle	0.028	0.809	1.457	1,606
17	Alignment = Curve-Level, Driver_Age = 25–34	>	Collision_Type = Single Vehicle	0.038	0.807	1.454	2,216
18	Alignment = Curve-Level, Driver_Gender = Male	>	Collision_Type = Single Vehicle	0.108	0.801	1.443	6,266
21	Collision_Type = Right Angle, Driver_Gender = Female	>	Alignment = Straight-Level	0.030	0.858	1.264	1,743
22	Collision_Type = Left Turn - Angle, Driver_Gender = Male	>	Alignment = Straight-Level	0.017	0.852	1.256	1,017
23	Collision_Type = Rear End, Driver_Age = 15–24	>	Alignment = Straight-Level	0.059	0.851	1.255	3,438
24	Collision_Type = Rear End, Driver_Age = 25–34	>	Alignment = Straight-Level	0.037	0.851	1.255	2,160
25	Collision_Type = Left Turn - Opposite Direction, Driver_Gender = Male	>	Lighting = Daylight	0.010	0.746	1.255	588
26	Collision_Type = Right Angle, Severity = Complaint	>	Alignment = Straight-Level	0.029	0.851	1.254	1,685
27	Collision_Type = Right Angle, Driver_Gender = Male	>	Lighting = Daylight	0.039	0.745	1.254	2,284
30	Collision_Type = Rear End, Driver_Gender = Female	>	Alignment = Straight-Level	0.064	0.849	1.252	3,731
31	Collision_Type = Right Angle, Severity = No Injury	>	Alignment = Straight-Level	0.036	0.849	1.252	2,096
32	Day.of_Week = Weekend, Collision_Type = Rear End	>	Alignment = Straight-Level	0.062	0.849	1.251	3,595
33	Collision_Type = Rear End, Severity = No Injury	>	Alignment = Straight-Level	0.100	0.846	1.247	5,806
34	Day.of_Week = Weekday, Collision_Type = Rear End	>	Alignment = Straight-Level	0.103	0.846	1.247	6,021
35	Lighting = Daylight, Collision_Type = Right Angle	>	Alignment = Straight-Level	0.057	0.846	1.247	3,334

Table 5. Association rules for 4-itemsets.

No	Antecedent	Consequent	Supp	Conf	Lift	Count
1	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.013	0.939	1.691	749
2	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.028	0.930	1.675	1,626
3	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Severity = No Injury	Collision_Type = Single Vehicle	0.037	0.928	1.673	2,128
4	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Gender = Female	Collision_Type = Single Vehicle	0.021	0.926	1.668	1,209
5	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Gender = Male	Collision_Type = Single Vehicle	0.052	0.919	1.656	3,028
6	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Severity = Complaint	Collision_Type = Single Vehicle	0.025	0.919	1.655	1,467
7	Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.018	0.917	1.652	1,026
8	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.030	0.872	1.572	1,738
9	Lighting = Dark - No Street Lights, Driver_Gender = Female, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.016	0.860	1.549	957
10	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.040	0.856	1.541	2,317
11	Day.of_Week = Weekend, Alignment = On Grade-Curve, Driver_Gender = Male	Collision_Type = Single Vehicle	0.011	0.854	1.538	618
12	Day.of_Week = Weekend, Lighting = Dark - No Street Lights, Severity = No Injury	Collision_Type = Single Vehicle	0.082	0.852	1.535	4,751
13	Day.of_Week = Weekend, Alignment = Curve-Level, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.030	0.851	1.533	1,762
14	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Age = 45–54	Collision_Type = Single Vehicle	0.024	0.850	1.532	1,380
15	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Gender = Female	Collision_Type = Single Vehicle	0.051	0.848	1.527	2,977
16	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Gender = Male	Collision_Type = Single Vehicle	0.108	0.846	1.523	6,320
17	Day.of_Week = Weekend, Lighting = Dark - No Street Lights, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.024	0.842	1.518	1,428
18	Day.of_Week = Weekday, Lighting = Dark - No Street Lights, Severity = No Injury	Collision_Type = Single Vehicle	0.078	0.840	1.514	4,546
19	Alignment = Curve-Level, Severity = Complaint, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.022	0.839	1.512	1,294
20	Lighting = Dark - No Street Lights, Driver_Gender = Female, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.021	0.838	1.510	1,213
21	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.051	0.837	1.508	2,957
22	Alignment = Curve-Level, Severity = Complaint, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.014	0.837	1.507	829
23	Day.of_Week = Weekend, Alignment = Curve-Level, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.013	0.836	1.507	782
24	Day.of_Week = Weekend, Lighting = Dark - No Street Lights, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.036	0.836	1.506	2,102
25	Day.of_Week = Weekend, Alignment = Curve-Level, Severity = Moderate	Collision_Type = Single Vehicle	0.012	0.835	1.504	694
26	Lighting = Dark - No Street Lights, Severity = No Injury, Driver_Age = 55–64	Collision_Type = Single Vehicle	0.011	0.833	1.500	642
27	Alignment = Curve-Level, Driver_Gender = Female, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.022	0.832	1.499	1,264
28	Alignment = Curve-Level, Driver_Gender = Male, Driver_Age = 15–24	Collision_Type = Single Vehicle	0.037	0.832	1.499	2,154
29	Day.of_Week = Weekday, Alignment = Curve-Level, Severity = Moderate	Collision_Type = Single Vehicle	0.011	0.829	1.494	617
30	Day.of_Week = Weekday, Lighting = Dark - No Street Lights, Driver_Age = 35–44	Collision_Type = Single Vehicle	0.024	0.829	1.494	1,379
31	Day.of_Week = Weekend, Alignment = Curve-Level, Severity = Complaint	Collision_Type = Single Vehicle	0.030	0.828	1.492	1,726
32	Alignment = On Grade-Curve, Severity = No Injury, Driver_Gender = Male	Collision_Type = Single Vehicle	0.010	0.827	1.490	589
33	Lighting = Dark - No Street Lights, Driver_Gender = Male, Driver_Age = 25–34	Collision_Type = Single Vehicle	0.046	0.826	1.488	2,708
34	Day.of_Week = Weekend, Lighting = Dark - No Street Lights, Driver_Age = 45–54	Collision_Type = Single Vehicle	0.019	0.826	1.487	1,099
35	Alignment = Curve-Level, Severity = Moderate, Driver_Gender = Male	Collision_Type = Single Vehicle	0.016	0.826	1.487	932

lift values for different types of collisions are almost the same. Among these rules, {Collision Type = Rear End} => {Alignment = Straight-level} displays a higher support value (0.165) with 9,618 frequency count. Middle age (45–54) to older drivers (55 plus) are seen to be crash prone in daytime rather than nighttime in rainy weather. Female drivers are seen in fewer rules than the male drivers. Male drivers are associated with various variable categories with higher lift value than the female drivers. Age and gender are found significant during rainy weather in other studies (Ahmed & Ghasemzadeh, 2018). One particular rule {Lighting = Dark- No Street Light} => {Driver_Gender = Male} shows a high support value of 0.203 with a frequency of 11,832 crashes.

Many interesting rules are observed in Table 4 for 3-itemsets rules. Drivers with age group in between 15 and 44 appear frequently in the rules generated for 3-itemsets. The first rule {Alignment = On Grade-Curve, Lighting = Dark - No Street Lights} => {Collision Type = Single Vehicle} has the highest lift value. Curve-aligned roadways and roadways with poor illumination are frequently visible in the antecedents of the rules. Another study showed that crash severity in adverse weather conditions causing wet pavement surface was more likely to increase at curves or ramps (Abdel-Aty, Pemmanaboina, & Hsia, 2006). Young drivers associated with curve-aligned roadways and poorly illuminated roadways result in single-vehicle run-off crashes in rainy weather. The antecedents {Lighting = Dark - No Street Lights, Driver_Gender = Female} and {Lighting = Dark - No Street Lights, Driver_Gender = Male} resulted in single-vehicle run-off crashes. The rule associated with female driver has a higher lift value than the male drivers, but the support value is higher in the rules for male drivers with a frequency count of 10,712.

The rule {Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Age = 35-44} => {Collision Type = Single Vehicle} has the highest lift value. The next rule is associated with {Driver_Age = 15-24} in place of {Driver_Age = 35-44}. The second rule has higher support value with frequency of 1,626 than the first rule. Another interesting rule is {Alignment = Curve-Level, Lighting = Dark - No Street Lights, Driver_Gender = Female} => {Collision Type = Single Vehicle}. The similar rule associated with male drivers has a lower lift value with higher support. These rules indicated that curve level and poorly illuminated roadways are crash-prone areas during rainy weather driving. The age-group vulnerable for the combination of these two characteristics is in between 15 to 44. The consequent {Collision Type = Single Vehicle} is present in all top 35 rules for 4-itemsets. In most cases, this consequent is associated with these critical attributes: curve roadways, drivers in age group 15–44, and roadways with no lighting at night. Weekend crashes

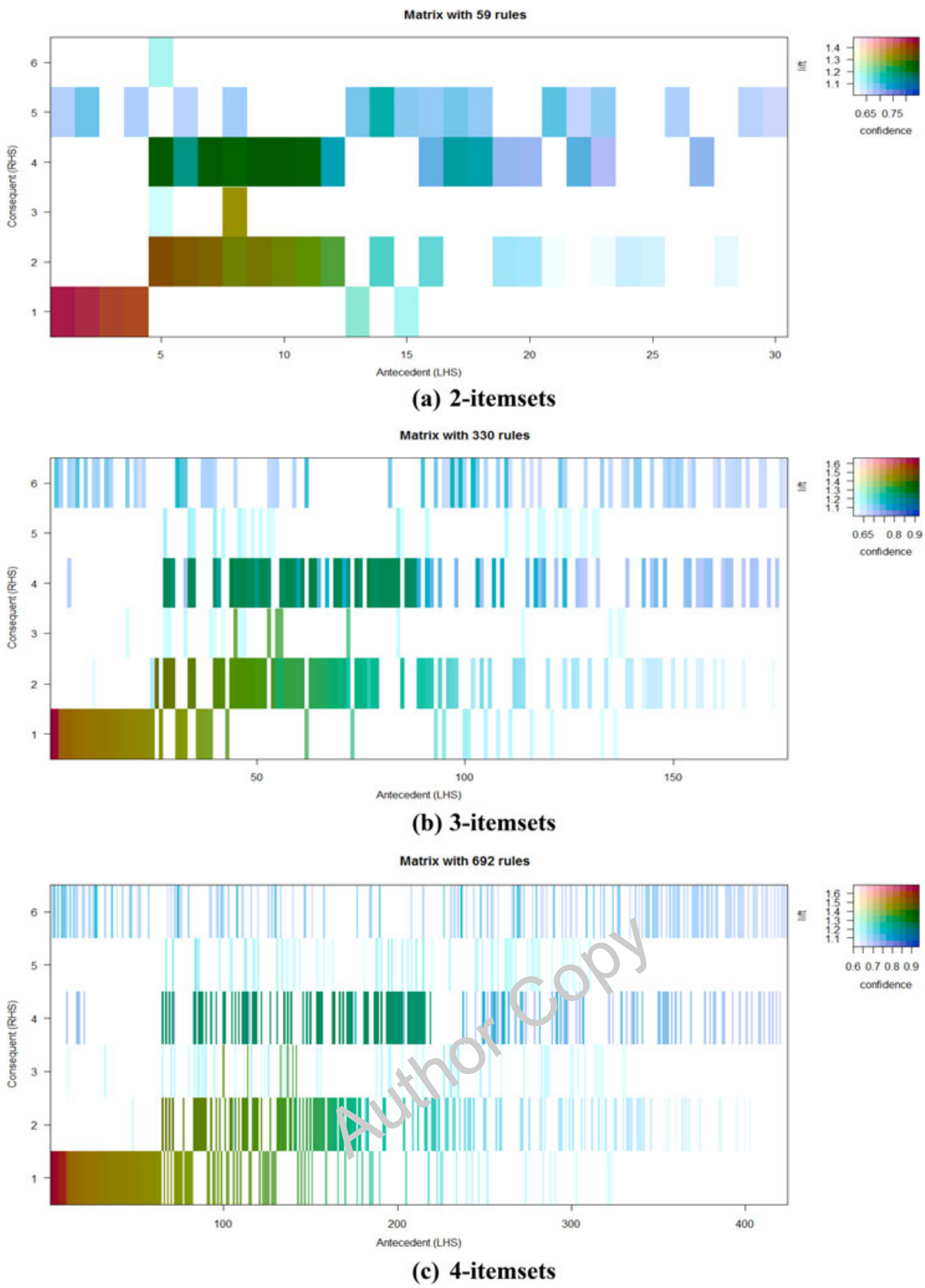

Figure 3. Matrix plot for the generated rules. (a) 2-itemsets; (b) 3-itemsets; (c) 4-itemsets.

have higher lift values than weekday crashes. Although out of 692 rules, consequent {Alignment = Straight-level} is the most frequent (nearly 22% in the rules).

In general, the rules presented above indicate the possible associations for crashes in rainy weather. By analyzing the rules of 2-, 3-, and 4-

itemsets, few variable categories were discovered as the dominant factors for rainy weather crashes. The factors should be carefully investigated by the traffic agencies for considering appropriate countermeasures to avoid crashes and crash severities in hazardous roadways.

An excellent visual representation of the results is also an integral part of the data mining process. Visualization of the results of the association rules is facilitated by the use of 'arulesViz' package (Hahsler & Chelluboina, 2018). Manual inspection of 1,890 rules is not a viable option. A straightforward visualization of association rules involves a scatter plot (Figure 2) with two interest measures on the axes: the support values of the rules are on the x axis and the lift values are on the y axis. The color of the points (light gray to black) is used to indicate the confidence value of each rule. The rules with lower support and lower lift values are higher in frequencies than the rules of higher support and higher lift values.

Matrix-based visualization methods organize the antecedent and consequent itemsets on the x and y axes, respectively. Selected interest measures can be displayed at the joint of the antecedent and consequent for a given rule. An antecedent/consequent combination with no rule keeps the joint area blank. By considering the set of association rules

$$\mathbf{M} = \left\{(a_1, c_1, q_1), \dots (a_i, c_i, q_i), \dots (a_n, c_n, q_n)\right\}$$

Where, a_i is the antecedent, c_i is the consequent, and q_i is the selected interest measure forthe *ith* rule for $i = 1, \dots, n$.

Suppose in the visualization matrix **M**, the set of K unique antecedents and L unique consequents are identified. A '$L \times K$' matrix **M** with one column for each unique antecedent and one row for each unique consequent was also done. Finally, the matrix was populated by setting $\mathrm{M_{ik}} = \mathrm{m_i}$ for $\mathrm{i} = 1, \dots, \mathrm{n}$ and l and k corresponding to the position of a_i and c_i in the matrix. **M** also contains blank cells because many potential association rules will not meet the minimum thresholds for support and confidence. The matrix-based visual plots are illustrated in Figure 3. The number of rows/columns (x and y axis) depends on the number of unique features or attributes in the consequent/antecedent in the set of rules. This figure gives a general idea of how the consequents are varied based on the lift and confidence value with respect to the antecedents. Figure 3 indicates that number of unique features increase with the increase of the itemsets.

Figure 4 reveals the easier visualization of large sets of association rules from the grouped matrix. Balloon plots are drawn with antecedent groups as columns and consequents as rows (for 2-, 3-, and 4-itemsets). The color of the balloons (light gray to black) represents the lift value and the size of the balloon shows the aggregated support. The number of antecedents and the most important (frequent) items in the group are displayed as the labels

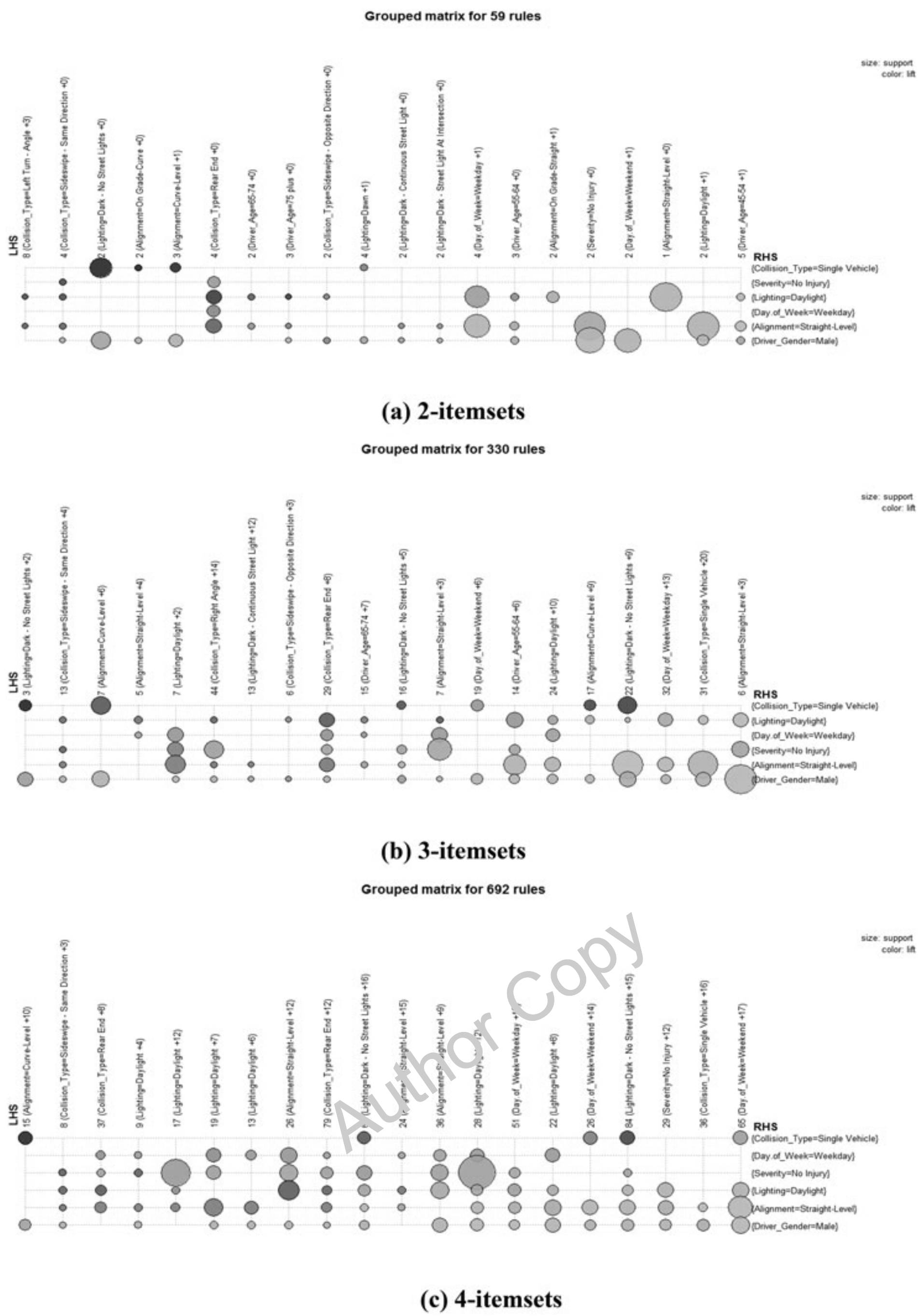

Figure 4. Grouped matrix plot for the generated rules.

for the columns. The association rules reveal certain critical safety attributes: single vehicle, daylight, straight-aligned roadways, and male drivers.

The study shows that the application of association rules mining in a specific environmental condition can help to reveal how drivers' behavior, roadway conditions, and crash temporal characteristics are associated with different collision types and injury severity. These findings are expected to be useful for policy makers to develop better safety policies.

5. Conclusion

The major objective of this study is to develop an efficient way of determining key significant rules that contribute to rainy condition crashes. This article demonstrated that association rules mining technique is one the suitable approach to extract knowledge from the traffic crash data under rainy condition. Some interesting findings are observed for crashes in rainy weather. Some of the findings verify the general perceptions on such types of crashes, and a few findings are quite surprising. The most significant single variable category for the situation is associated with single-vehicle ROR crashes. This crash type is highly associated with the presence of other roadway features such as on grade-curve aligned roadways, curved roadways, and roadways with no streetlights at night. During rainy weather, Property Damage Only and sideswipe (same direction) crashes shows high likelihood of occurrence. For drivers age 55 and older, most of the crashes during rainy weather are associated with daytime. It can be associated with the facts that older drivers usually avoid driving at night during rainy weather. Moderate injuries show high likelihood in single-vehicle ROR crashes, which is also common in ROR crashes due to other reasons. Roadways with poor illumination are associated with straight-level aligned roadways for many rainy weather crashes. Drivers (age group in between 15 to 44) seem to be associated with poor illumination and roadway curves during rainy weather crashes. The findings this study are supported by other study findings. The results provide quantitative support the significant factors associated with rainfall crashes.

This study has several limitations. The current study has not developed an optimization criterion to determine the optimized support and confidence thresholds. Future studies can consider using ant colony optimization or genetic algorithm to determine the optimized values. Another limitation is the use of limited number of variables. Future studies can consider using a larger number of variables to generate more rigorous and latent patterns. Additionally, the results would be more intuitive if there is a comparison between dry weather crashes, which is currently not performed in this study.

By observing the potential patterns in the discovered rules, the results can provide valuable insights into the underlying relationships between risk factors and crashes under particular conditions. Policy makers and safety professionals can use the findings from this study to conduct decision making on the appropriate countermeasures. It is important to adopt weather specific countermeasures (e.g., advisory weather signals or signs, high visible pavement markings in low visibility condition) to minimize the number of crashes occurred during rainy condition. The efforts of highway department to provide safety during rainy condition would be more productive from the findings of the current study.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753
Xiaoduan Sun http://orcid.org/0000-0001-7282-1340

References

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered Probit models. *Journal of Safety Research*, *34* (5), 597–603. doi:10.1016/j.jsr.2003.05.009

Abdel-Aty, M., Pemmanaboina, R., Z., & Hsia, L. (2006). Assessing crash occurrence on urban freeways by applying a system of interrelated equations. *Transportation Research Record*, *1953*(1), 1–9. doi:10.3141/1953-01

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Data mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, *5*(6), 914–925. doi:10.1109/69.250074

Ahmed, M. M., Abdel-Aty, M., Lee, J., & Yu, R. (2014). Real-time assessment of fog-related crashes using airport weather data: A feasibility analysis. *Accident Analysis & Prevention*, *72*, 309–317. doi:10.1016/j.aap.2014.07.004

Ahmed, M. M., & Ghasemzadeh, A. (2018). The impacts of heavy rain on speed and headway Behaviors: An investigation using the SHRP2 naturalistic driving study data. *Transportation Research Part C: Emerging Technologies*, *91*, 371–384. doi:10.1016/j.trc.2018.04.012

Ait-Mlouk, A., Gharnati, F., & Agouti, T. (2017). An improved approach for association rule mining using a multi-criteria decision support system: A case study in road safety. *European Transport Research Review*, *9*(3), 40.

Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*, *34*(6), 729–741. doi:10.1016/S0001-4575(01)00073-2

Andrey, J., & Yagar, S. (1993). A temporal analysis of rain-related crash risk. *Accident Analysis & Prevention*, *25*(4), 465–472. doi:10.1016/0001-4575(93)90076-9

Black, A. W., Villarini, G., & Mote, T. L. (2017). Effects of rainfall on vehicle crashes in six U.S. states. *Weather, Climate, and Society*, *9*(1), 53–70. doi:10.1175/WCAS-D-16-0035.1

Brodsky, H., & Hakkert, A. S. (1988). Risk of a road accident in rainy weather. *Accident Analysis & Prevention*, *20*(3), 161–176. doi:10.1016/0001-4575(88)90001-2

Brooks, B. (2008). Shifting the focus of strategic occupational injury prevention: mining free-text, workers compensation claims data. *Safety Science*, *46*(1), 1–21. doi:10.1016/j.ssci.2006.09.006

Brown, D. E. (2016). Text mining the contributors to rail accidents. *IEEE Transactions on Intelligent Transportation Systems*, *17*(2), 346–355. doi:10.1109/TITS.2015.2472580

Campbell, M. (1971). The wet-pavement accident problem: Breaking through. *Traffic Quarterly*, *25*, 209–214.
Chen, C., Zhang, G., Qian, Z., Tarefder, R., & Tian, Z. (2016). Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accident Analysis & Prevention*, *90*, 128–139. doi:10.1016/j.aap.2016.02.011
Chung, Y. S. (2013). Factor complexity of crash occurrence: an empirical demonstration using boosted regression trees. *Accident Analysis & Prevention*, *61*, 107–118. doi:10.1016/j.aap.2012.08.015
Figueira, A. da C., Pitombo, C. S., de Oliveira, P. T. M. e S., & Larocca, A. P. C. (2017). Identification of rules induced through decision tree algorithm for detection of traffic accidents with victims: a study case from Brazil. *Case Studies on Transport Policy*, *5*(2), 200–207. doi:10.1016/j.cstp.2017.02.004
Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*(1), 95–103. doi:10.3141/2519-11
Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*(2), 146–155. doi:10.1016/j.iatssr.2015.07.001
Das, S., Brimley, B. K., Lindheimer, T. E., & Pant, A. 2016. *Safety Impacts of Reduced Visibility in Inclement Weather*. Ann Arbor, MI: ATLAS Center, University of Michigan.
Das, S., Brimley, B. K., Lindheimer, T. E., & Zupancich, M. (2017). Association of reduced visibility with crash outcomes. *IATSS Research*, *42*(3), 143–151.
Das, S., Minjares-Kyle, L., Avelar, R., Dixon, K., & Bommanayakanahalli, B. (2017). Improper Passing Related Crashes on Rural Roadways: Using Association Rules Negative Binomial Miner. Transportation Research Board Annual Meeting, Washington D.C.
Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., & Gillette, G. (2018). Using deep learning in severity analysis of at-fault motorcycle rider crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 1–13.
Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2018). Supervised Association Rules Mining on Pedestrian Crashes in Urban Areas: Identifying Patterns for Appropriate Countermeasures. *International Journal of Urban Sciences*, *23*(1), 30–48.
Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: exploration using 'eclat'association rules to promote safety. *International Journal of Transportation Science and Technology*, 7(2), 114–123.
Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention*, *111*, 43–55. doi:10.1016/j.aap.2017.11.016
Dissanayake, S., & Lu, J. (2002). Analysis of severity of young driver crashes: Sequential binary logistic regression modeling. *Transportation Research Record: Journal of the Transportation Research Board*, *1784*(1), 108–114. doi:10.3141/1784-14
Dutta, A. (2016). A smart design framework for a novel reconfigurable multi-processor systems-on-chip (ASREM) architecture. Doctoral Dissertation, Computer Science and Engineering, University of Louisiana at Lafayette.
Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. *Accident Analysis & Prevention*, *36*(4), 637–647. doi:10.1016/S0001-4575(03)00085-X
Ghasemzadeh, A., & Ahmed, M. M. (2017). Drivers' lane-keeping ability in heavy rain. *Transportation Research Record: Journal of the Transportation Research Board*, *2663*(1), 99–108. doi:10.3141/2663-13

Geurts, K., Thomas, I., & Wets, G. (2005). Understanding Spatial Concentrations of Road Accidents Using Frequent Item Sets. *Accident Analysis & Prevention*, *37*(4), 787–799. doi:10.1016/j.aap.2005.03.023

Gibert, X., Patel, V., & Chellappa, R. (2017). Deep multitask learning for railway track inspection. *IEEE Transactions on Intelligent Transportation Systems*, *18*(1), 153–164. doi: 10.1109/TITS.2016.2568758

Haghighi-Taleb, D. (1973). An Investigation into the relationship between rainfall and road accident frequencies in two cities. *Accident Analysis & Prevention*, *5*, 343–349.

Hahsler, M., Buchta, C., Gruen, B., & Hornik, K. (2018). arules: Mining Association Rules and Frequent Itemsets. http://cran.r-project.org/web/packages/arules/arules.pdf.

Hahsler, M., & Chelluboina, S. (2018). arulesViz: Visualizing Association Rules and Frequent Itemsets. http://cran.r-project.org/web/packages/arulesViz/arulesViz.pdf.

He, S., Wang, Y., Sun, Y., Gao, W., Chen, Q., & An, Y. (2008). The Research of Multidimensional Association Rule in Traffic Accidents, 4th International Conference on Wireless Communications, Networking and Mobile Computing, 12-14 October, Dalian, China.

Jackson, T. L., & Sharif, H. O. (2016). Rainfall impacts on traffic safety: rain-related fatal crashes in Texas. *Geomatics, Natural Hazards and Risk*, *7*(2), 843–860. doi:10.1080/19475705.2014.984246

Jalayer, M., Pour-Rouholamin, M., & Zhou, H. (2018). Wrong-way driving crashes: a multiple correspondence approach to identify contributing factors. *Traffic Injury Prevention*, *19*(1), 35–41. doi:10.1080/15389588.2017.1347260

Jaroszweski, D. & McNamara, T. (2014). The influence of rainfall on road accidents in urban areas: A weather radar approach. *Travel Behaviour and Society*, *1*(1), 15–21.

Jung, S., Qin, X., & Noyce, D. A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis & Prevention*, *42*(1), 213–224. doi:10.1016/j.aap.2009.07.020

Keay, K., & Simmonds, I. (2006). Road accidents and rainfall in a large Australian city. *Accident; Analysis and Prevention*, *38*(3), 445–454.

Khan, G., Bill, A., & Noyce, D. (2015). Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity. *Transportation Research Part C: Emerging Technologies*, *50*, 86–96. doi:10.1016/j.trc.2014.10.003

Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, *24*(1), 62–72. doi:10.1007/s40534-016-0095-5

Lee, J., Chae, J., Yoon, T., & Yang, H. (2018). Traffic accident severity analysis with rain-related factors using structural equation modeling – A case study of Seoul City. *Accident Analysis & Prevention*, *112*, 1–10. doi:10.1016/j.aap.2017.12.013

Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*, *40*(4), 1611–1618. doi: 10.1016/j.aap.2008.04.010

Liu, C., & Subramanian, R. (2009). *Factors Related to Fatal Single-Vehicle Run-Off-Road Crashes.* Washington D.C: NHTSA, USDOT.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291–305. doi:10.1016/j.tra.2010.02.001

Louisiana Highway Safety Commission. (2017). *Louisiana Highway Safety Plan- Federal Fiscal Year 2018*. Baton Rouge, LA: Louisiana Highway Safety Commission.

Mannering, F. L., & Bhat, C. (2014). Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research, *1*, 1–22. doi:10.1016/j.amar.2013.09.001

Marukatat, R. (2007). Structure-Based Rule Selection Framework for Association Rule Mining of Traffic Accident Data. Computational Intelligence and Security: Lecture Notes in Computer Science, 4456, 231–239.

Mills, B. N., Andrey, J., & Hambly, D. (2011). Analysis of precipitation-related motor vehicle collision and injury risk using insurance and police record information for Winnipeg, Canada. *Journal of Safety Research*, *42*(5), 383–390. doi:10.1016/j.jsr.2011.08.004

Nitsche, P., Thomas, P., Stuetz, R., & Welsh, R. (2017). Pre-crash scenarios at road junctions: A clustering method for car crash data. *Accident Analysis & Prevention*, *107*, 137–151. doi:10.1016/j.aap.2017.07.011

Omranian, E., Sharif, H., Dessouky, S., & Weissmann, J. (2018). Exploring rainfall impacts on the crash risk on Texas roadways: A crash-based matched-pairs analysis approach. *Accident Analysis & Prevention*, *117*, 10–20. doi:10.1016/j.aap.2018.03.030

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety Science*, *47*(1), 145–154. doi:10.1016/j.ssci.2007.12.001

Saha, D., Alluri, P., & Gan, A. (2015). Prioritizing highway safety manual's crash prediction variables using boosted regression trees. *Accident Analysis & Prevention*, *79*, 133–144.

Satterthwaite, S. (1976). An assessment of seasonal and weather effects on the frequency of road accidents in California. *Accident Analysis & Prevention*, *8*, 87–96. doi:10.1016/0001-4575(76)90002-6

Savolainen, P. T., Mannering, F., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention*, *43*(5), 1666–1676. doi:10.1016/j.aap.2011.03.025

Sherretz, L. A., & Farhar, B. C. (1978). An analysis of the relationship between rainfall and the occurrence of traffic accidents. *Journal of Applied Meteorology*, *17*(5), 711–715. doi:10.1175/1520-0450(1978)017<0711:AAOTRB > 2.0.CO;2

Sun, X., Hu, H., Habib, E., & Magri, D. (2011). Quantifying crash risk under inclement weather with radar rainfall data and matched-pair method. *Journal of Transportation Safety & Security*, *3*(1), 1–14. doi:10.1080/19439962.2010.524348

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, *72*, 244–256. doi:10.1016/j.aap.2014.06.017

Weng, J., Zhu, J., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. *Accident Analysis & Prevention*, *92*, 43–52. doi:10.1016/j.aap.2016.03.017

Wu, Y., Abdel-Aty, M., & Lee, J. (2018). Crash risk analysis during fog conditions using real-time traffic data. *Accident Analysis & Prevention*, *114*, 4–11. doi:10.1016/j.aap.2017.05.004

Xu, C., Wang, W., & Liu, P. (2013). Identifying crash-prone traffic conditions under different weather on freeways. *Journal of Safety Research*, *46*, 135–144. doi:10.1016/j.jsr.2013.04.007