Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis

Subasish Das

To cite this article: Subasish Das (2020): Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2020.1802362

To link to this article: https://doi.org/10.1080/23249935.2020.1802362

Accepted author version posted online: 27 Jul 2020.
Published online: 06 Aug 2020.

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2020.1802362

Identifying key patterns in motorcycle crashes: findings from taxicab correspondence analysis

Subasish Das

Texas A&M Transportation Institute, Roadway Safety Program, Bryan, TX, USA

ABSTRACT

Due to the absence of protective structural surrounding and advanced restraints like the motorists, motorcyclists are considered vulnerable roadway users like pedestrians and bicyclists. Per vehicle mile traveled in 2016, motorcyclist fatalities occurred 28 times more frequently than passenger car occupant fatalities. The identification of the patterns and associations between key contributing factors can help in determining strategies for motorcycle-related crash reduction. In addition to current endeavors, there is a need for newer directions in study design with newer data sources and methods. Determining the groups of core factors helps address motorcycle crashes more effectively. This study used seven years (2010–2016) of motorcycle crash data in Louisiana to determine the key relationships between the influencing factors by using Taxicab Correspondence Analysis (TCA). The analysis showed that TCA presents a dimension-reduced map of the variable categories by developing several clusters.

ARTICLE HISTORY
Received 11 November 2019
Accepted 21 July 2020

KEYWORDS
Motorcycle crashes; taxicab correspondence analysis (TCA); patterns and trends; dimension reduction; contributing factors

Introduction

Motorcycle crashes lead to a high rate of traffic fatalities and serious injuries. In 2017, motorcyclist fatalities in the U.S. occurred 27 times more often than passenger car occupant fatalities per vehicle mile traveled. In 2017, there were 5,172 motorcycle crash-related fatalities recorded in the U.S. – a 16% increase from the number of motorcyclists killed in 2009 (NHTSA 2019). The identification of the patterns and associations between key contributing factors can help develop strategies to reduce motorcycle-related crashes. The traditional approach of motorcycle crash studies is the establishment of a relationship between key contributing factors and crash occurrence, sorted by injury level. One key deficit of the developed models is that these models depend on general inferences and aggregate measures. Additionally, these methods must consider several assumptions to perform the analysis.

The major task of highway safety analysis is to identify the highly associated factors for targeted crash types. The Taxicab Correspondence Analysis (TCA) method is a pattern recognition tool that investigates the significance of variable attributes by identifying

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M Transportation Institute, Roadway Safety Program, Bryan, TX 77807, USA

the co-occurrence of variable groups from within a complex dataset with a wide range of variables. The TCA method considers arbitrary data sets as a mixture of points in a high dimensional space. This technique aims to simplify a complex data structure to meaningful knowledge in the form of trends and patterns. Instead of depicting the relationships between a set of dependent variables and crash outcomes, TCA is used to analyze the associations between multiple variable attributes for pattern recognition. It is beneficial to understand patterns of a complex dataset, so the methodological framework is often comprehensive. In this context, the main goal of this study is to determine the relationships between and the significance of key variables by using seven years (2010–2016) of Louisiana motorcycle crash data. Specifically, the study assesses the association between factors and describes the significance and inference of clusters and risk factors.

Literature review

Many studies have examined the patterns and trends of motorcycle crashes. These studies can be divided into three broader groups: (1) vulnerability measures, (2) key contributing factors, and (3) crash severity analysis.

Vulnerability measures

The major task of highway safety analysis is to identify the most highly associated factors for targeted crash types. Motorcyclists have less protection, and they maintain higher speeds than other non-motorcycle motorists, so vulnerability measures are particularly high in motorcycle-related injuries. In an early study, Peek-Asa, McArthur, and Kraus (1999) examined the frequency of head injuries for non-standard helmet use among motorcyclists. The study used statistical tools such as Chi-squares and variance analysis in order to differentiate among groups using the medical records of the motorcyclists injured, which were collected from the available crash reports. The study concluded that the frequency and severity of head injuries were greater among motorcyclists using non-standard helmets than those using standard helmets or those with no helmet. Cohn et al. (2004) studied the impression of crashes on motorcyclists and the consequences of crashes after a year of the injury. The study utilized various data like demographic data, health condition, helmet usage, and employability of 94 patients for a period of six months. In another study, Khor et al. (2017) used multiple logistic analysis to explore the relationship between helmet use and C-spine injuries as a result of motorcycle crashes. The study gathered information including demographics, vital signs, Abbreviated Injury Scale (AIS), Injury Severity Score (ISS), and specific injuries for all motorcycle collisions between 2007 and 2014 involving either a driver or passenger. The research found that the use of helmet decreases the likelihood of head injuries and casualties; however, no connection with C-spine injury was identified.

De Rome (2005) developed a variety of plans and policies that aimed to solve road safety problems associated with motorcycles as a type of transport. Specifically, the policies aimed to reduce motorcycle crashes and casualties. A model for the development of motorcycle safety plans was developed with the aim of providing a systematic structure and process in the development of plans to promote all stakeholders' ownership of priorities and countermeasures. Furthermore, Ogle and Tillotson (2008) observed that the use of helmets was effective in reducing the danger and incidence of both head wounds and deaths in motorbike crashes, as well as in decreasing hospitalization and healthcare expenses. They also considered some of the statements against using helmets, including the individual's right to choose to wear a helmet.

Key contributing factors

The assessment of contributing factors for motorcycle crashes is difficult due to a lack of consistent real-time environment information and other information needed to understand the causation patterns of motorcycle crashes. Kostyniuk and Nation (2005) examined the trends and patterns of motorcycle crashes from 2001 to 2005 with the objective of augmenting motorcycle-related fatalities in Michigan. Similarly, Ryb et al. (2009) concentrated on the impact of age on the consequence of injured individuals in a Maryland data set linking hospital discharge documents and police reports. This study examined whether this mortality rise reflected alterations in collision frequency or case fatality levels in any specific age category. In another study, Medina and Soto (2011) aimed to define their perception of the relationship between highway condition and motorcycle safety. They conducted a 39-segment highway review method using correlation, ANOVA, and multiple regression analyses, along with a motorcycle drivers' study. Results of the research suggested that the primary road components connected with motorcycle collision rates were the type and width of cross-section, junction density, published speed limit, presence of on-street parking, pavement faults, and housing development.

Naumann et al. (2010) investigated on the recent high rise of motorcycle crashes. Weighted ratios and rates were calculated, and the number of potential motorcycle crashes was estimated using linear regression. The research showed a growing trend, and preliminary estimates suggest that, due to present developments, these casualties could almost double from 170,000 to 320,000 annual injuries from 2008 to 2020. Correspondingly, Eustace, Indupuru, and Hovey (2011) used Ohio crash information from 2003 to 2007 to explore the likelihood of a motorcyclist being seriously wounded in a collision and the concerning risk variables. The findings indicated that the risk of fatality rises considerably when the following conditions apply: female rider, speeding, impairment, driving without a helmet, single rider collisions or non-intersection place, collision on horizontal bends or graded sections, and on main roads.

Safety training is one potential countermeasure that could be used to reduce motorcycle crashes. Shaheed and Gkritza (2014) examined whether states that require driver's education have a lower risk of motorcycle collisions than the states where it is not a requirement. In addition, they analyzed Iowa motorcycle crash data from 2001 to 2008 by using a latent class strategy to explore variables influencing collision seriousness and resulting in one-vehicle motorcycle crashes. The findings of the assessment indicated an important connection between the serious outcomes of collision injury and crash-specific variables such as speed, run-off highway, no helmet, impaired riding, and others. Wu et al. (2018) investigated which types of motorcycle crashes constitute the greatest safety risk to riders. Chawla, Karaca, and Savolainen (2019) conducted factor identification by using the publicly available Federal Highway Administration (FHWA) Motorcycle Crash Causation Study (MCCS).

Crash severity outcomes

Several studies have quantified how different factors are associated with injury severity of motorcycle crashes. Using full Bayesian formulation, Cheng et al. (2017) developed five models representing different correlations of weather conditions and common motorcycle crash injuries and compared best-fit measures for four different levels of severity. Waseem, Ahmed, and Saeed (2019) found a need to lower speed limits on highways with a greater motorcycle ratio, distinguish motorcycles from heavy cars, remove fixed items from the roadside, and reduce risky conduct of motorcyclists. Chung and Song (2018) used Korean motorcycle crash data from 2009 to determine the critical factors that affect motorcycle crash severity based on categorical principal components analysis (CatPCA) and nonlinear canonical correlation analysis (NLCCA). Rappole et al. (2019) conducted a study on key contributing factors associated with the U.S. Army personnel rider fatalities. Das et al. (2018b) used a deep learning technique to examine five years (2010–2014) of Louisiana at-fault motorcycle rider-involved crashes. The developed model can predict severity types with high precision (94% accuracy), which is not typically achieved using a statistical method or machine learning algorithm.

The literature review identifies a critical gap in ongoing motorcycle crash studies. Many studies have focused either on factor identification or severity analysis. Analysis aiming to identify patterns of contributing factors has been explored less. The present study aims to mitigate the current research gap by applying an innovative data mining method.

Methodology

Taxicab correspondence analysis (TCA)

Jean-Paul Benzećri (1992) developed a multivariate statistical technique, called correspondence analysis (CA). There are several books that contain further information about CA if readers desire more detail (Benzećri 1992; Murtagh 2005; Greenacre and Blasius 2006; Le Roux and Rouanet 2010; Hjellbrekke 2018). Choulakian recently proposed an improved version of CA, called taxicab correspondence analysis (TCA), in a series of papers (Choulakian 2006a; Choulakian 2006b; Choulakian 2013). Various forms of CA analysis have been used in transportation safety studies (Das and Sun 2015; Das and Sun 2016; Jalayer, Pour-Rouholamin, and Zhou 2018; Das et al. 2018a; Das et al. 2019; Das and Dutta 2020; Das et al. 2020a; Das et al. 2020b; Das et al. 2020c; Das, Tran, and Theel 2020d).

CA is based on Euclidean distance. On the other hand, the concept of TCA is based on taxicab distance or Manhattan city block distance. Consider, $X = (x_1, x_2, \ldots.., x_n)$ and $Y = (y_1, y_2, \ldots.., y_n)$ and a vector $\boldsymbol{v} = (v_1, v_2, \ldots.., v_n)$ to evaluate these distances:

$$\text{Euclidean Distance} = ED(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2 [\text{with } L_2 \text{ Norm} = ||v||_2 = \sqrt{\sum_{i=1}^{n} (v_i)^2}]} \tag{1}$$

$$\text{Taxicab Distance} = TD(X, Y) = \sum_{i=1}^{n} |x_i - y_i| [\text{with } L_1 \text{ Norm} = ||v||_1 = \sum_{i=1}^{n} |v_i|] \tag{2}$$

The concept of singular value decomposition (SVD) is important to understand the core concept of different variants of CA. The analysts can consider a real matrix A, which can be decomposed as $M\Lambda^{1/2}N'$, with Λ the diagonal matrix of the real non-negative eigenvalues of AA' in which M the orthogonal matrix of the corresponding eigenvectors, and N the matrix of eigenvectors of $A'A$ (with constraints $M'M = I$ and $N'N = I$).The k-rank matrix can be written as:

$$a_{ij} = \sum_{i=1}^{k} \sqrt{\lambda_\alpha} m_{i\alpha} n_{i\alpha} \tag{3}$$

Choulakian (2013) used a recursive optimization process to build the SVD outcomes. In TCA, the distances can be considered as L_∞ metrics. The distance can be calculated by $max_{i\in(1,n)}|x_i - y_i|$ [with L_∞Norm $= v_\infty = max_{i\in(1,n)}|v_i|$]. The analysts can find the solution by recursively applying the optimization problem on the residuals in the form of taxicab singular value decomposition (TSVD). The relations between rows and columns of N are summarized by the χ^2 statistics. The independence can be formulated as $N_0 = nT_0 = nrl'$. The change from independence can be estimated as:

$$\begin{aligned}\chi^2 = n\theta^2 &= n \sum_i \sum_j \left(\frac{(t_{ij} - t_{i.}t_{.j})^2}{t_{i.}t_{.j}} \right) \\ &= n \sum_i \sum_j \frac{d_{ij}^2}{t_{i.}t_{.j}} [\text{with} (r-1) \times (l-1) \text{degrees of freedom}]\end{aligned} \tag{4}$$

TCA can be described asTaxicab SVD of the data table $D = T - rl'$ by considering the table's profiles, respectively $R = D_r^{-1}D$ for the rows and $L = D_l^{-1}D$ for the columns. The solution considers the residuals from the previous factors for each new step (a clear distinction from the conventional CA approach). The equations can be written as:

$$T = p_r p_c' + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} B_\alpha C_\alpha' \tag{5}$$

Elementwise the formula becomes:

$$t_{ij} = t_{i.}t_{.j} + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} B_{i\alpha} C_{j\alpha} \tag{6}$$

After conducting the transformation:

$$n_{ij} = nr_i l_j (1 + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} b_{i\alpha} c_{j\alpha}) \tag{7}$$

Data description

Data integration

The current study collected seven years (2010–2016) of traffic crash data from the Louisiana Department of Transportation and Development (LADOTD). The database contains information at three levels: (1) crash level data (each row indicates crash level information), (2) vehicle data (each row indicates driver level information), and (3) roadway inventory (each row indicates inventory information for each crash). It is important to note that the entries in vehicle data are greater than the entries in crash level data as vehicle data contains information of each involved vehicle in a crash. The crash identification numbers of vehicle databases are later merged with crash level and roadway inventory databases to develop a comprehensive database. Altogether, 14,084 motorcycles and motorcycle riders were identified in this study. Figure 1 illustrates the flowchart of the study design.

Exploratory data analysis

Figure 2 shows the number of motorcycle crashes from 2010 to 2016 in each parish of Louisiana. The figure shows that six parishes (Caddo, Calcasieu, Lafayette, East Baton Rouge, Orleans, and Jefferson) experience at least 700 motorcycle crashes per year. Five of these parishes are in the southern part of Louisiana. The Orleans Parish has the highest number of motorcycle crashes (1,865 crashes).

A distribution of crash severities based on districts shows that three districts have different patterns. KAB (K = fatality, A = incapacitating injury or severe, B = non-incapacitating injury or moderate) percentages are normally lower than CO percentages in most of the districts. District 58 (Chase) shows that there is a higher percentage of KAB crashes than CO (C = possible injury or complaint, and O = no injury or property damage only or PDO) crashes. However, District 58 represents only 67 motorcycle crashes. District 4 (Bossier City) and District 61 (Baton Rouge) show that percentages of KAB and CO crashes are similar. These two plots provide a brief narration of the spatial distribution of motorcycle crashes and severities. The anomalies and over-representation require more rigorous exploratory analysis on different spatial scales. Future studies can take a deeper look at these anomalies (Figure 3).

Due to the presence of a wide variety of variables, crash data analysis is complex in nature. This study developed a precise database based on key contributing factors by

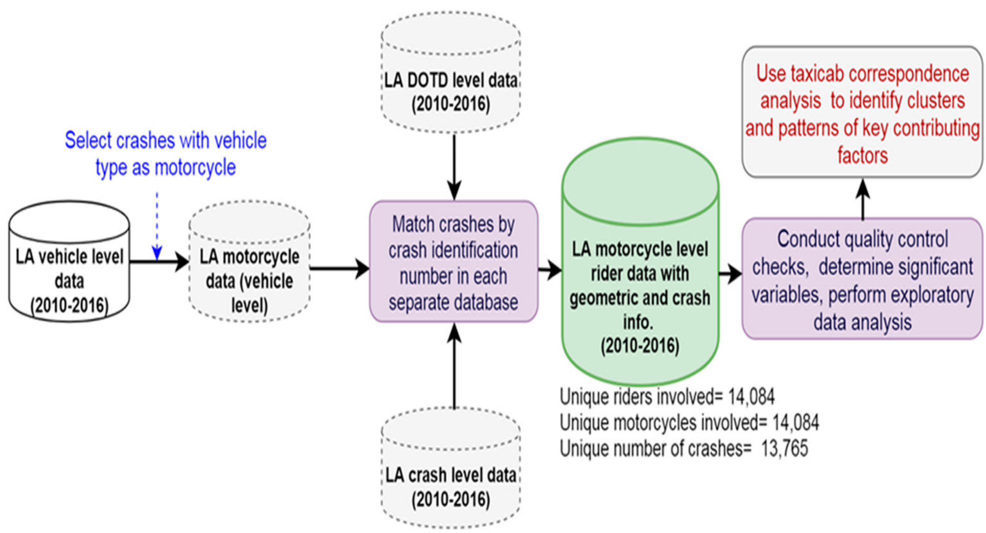

Figure 1. Flowchart of the study design.

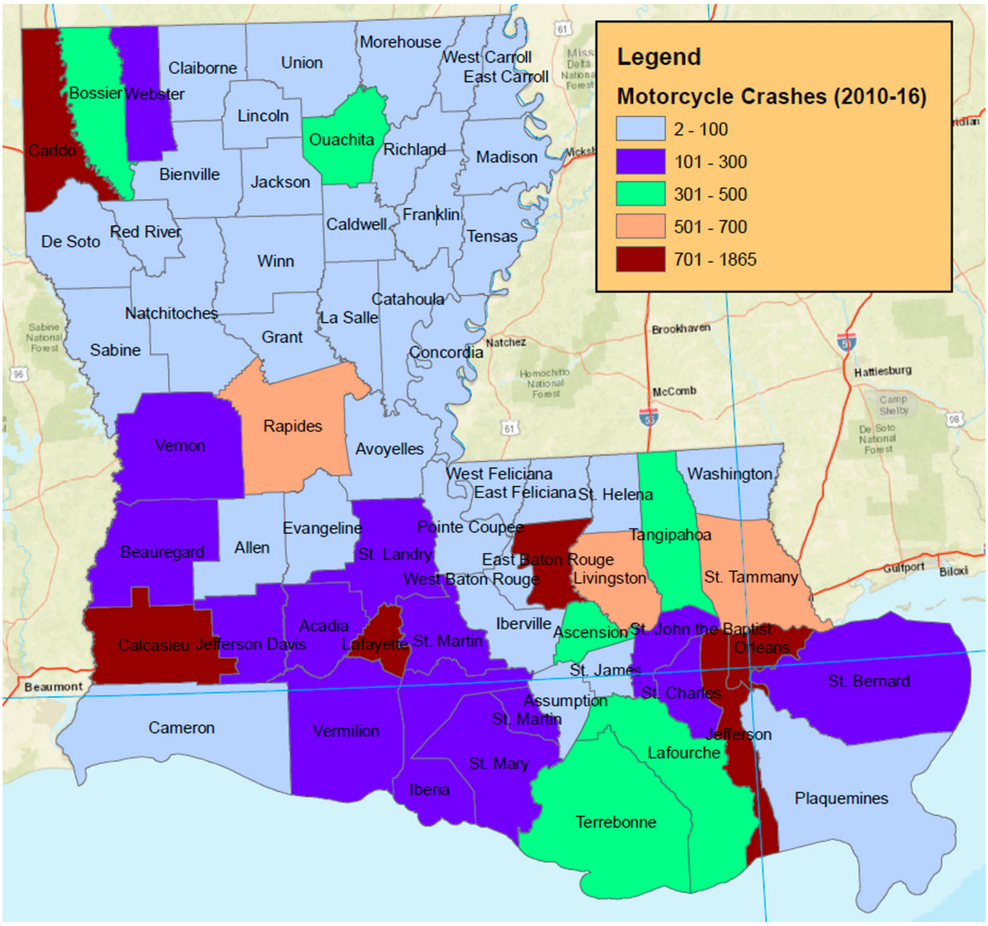

Figure 2. Motorcycle crashes by Parishes.

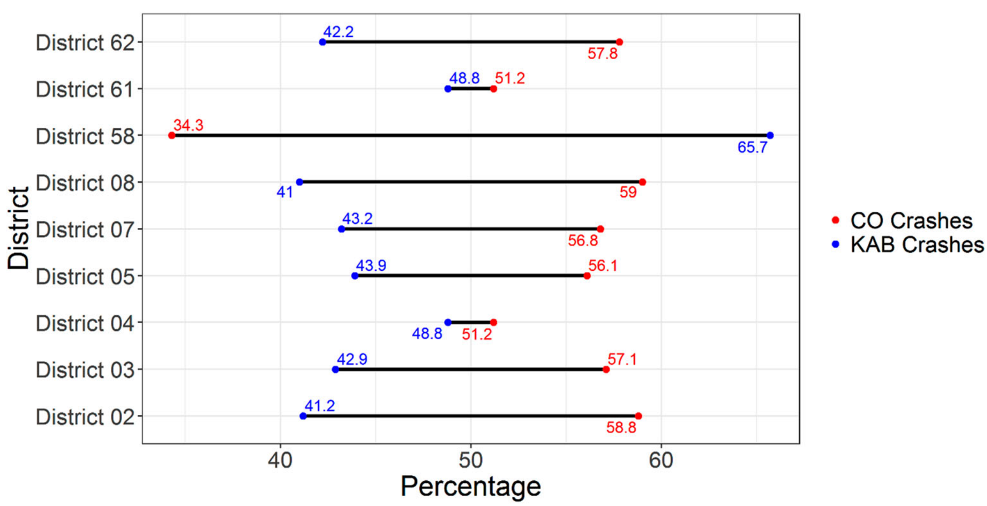

Figure 3. Percentage distribution of KAB and CO crashes by districts.

Table 1. Crash related variables.

Variable Category	Perc (%)
Day of Week	
Mon_Thurs	48.4
Fri	16.1
Sat_Sun	35.5
Hour ([b])	
1 am–6 am (01:00–06:59)	4.1
7 am–12 pm (07:00–12:59)	20.4
1 pm–6 pm (13:00–18:59)	48.8
7 pm-12 am (19:00–00:59)	26.7
Season ([b])	
Spring	30.5
Summer	25.7
Autumn	27.2
Winter	16.6
Collision Type (Collision)	
Single Motorcycle (Non-Coll w M/V)	30.5
Rear End	26.1
Rt Angle	18.6
Lt Turn	15.0
Sideswipe	8.0
Rt Turn	1.9
Location ([b])	
Segment	64.0
Intersection	36.0
Posted Speed Limit (PSL)	
25 mph (40 km/hr) or less	10.0
30–35 mph (48–56 km/hr)	25.0
40–45 mph (64–72 km/hr)	33.1
50–55 mph (80–88 km/hr)	24.7
60–65 mph (97–105 km/hr)	3.7
70–75 mph (113–121 km/hr)	3.5
Locality ([b])	
Business	31.6
Mixed	29.5
Residential	27.1
Open Country	11.8

Variable Category	Perc (%)
Roadway Type	
Two Way-No Physical Sep	65.1
Two Way-Physical Sep	27.2
One Way	7.8
Roadway Relation (RoadRel)	
On Roadway	82.9
Beyond Shoulder	15.4
Shoulder	1.8
Access Control (AccessControl)	
No Control	88.0
Full Control	7.0
Partial Control	5.0
Alignment ([b])	
Straight-Level	80.9
Curve-Level	16.1
On Grade	3.0
Traffic Control Type (TrafficControl)	
White Dashed Line	32.1
Yellow no Passing Line	10.7
No Control	13.5
Yellow Dashed Line	25.4
Green Sig on	7.8
Stop/Yield Sign	5.3
Red Sig on	5.2
Lighting ([b])	
Daylight	75.3
Dark-Continuous Street Lts	13.0
Dark-No Street Light	9.0
Dark-Str Lts-Intersect Only	2.7
Weather ([b])	
Clear	85.9
Cloudy	10.7
Rain/Fog/Sleet/Snow	3.4

Note: [a]Variable code, which is used in TCA analysis. [b]Variable code is same as the variable name.

removing non-pertinent variables and crash entries with incomplete information for various significant variables (for example, crash entries with variable attributes listed as others, unknown, or missing were removed). This approach reduced the dataset size by 10%. The final dataset contains information of 12,657 motorcycle riders who were involved in crashes. This study used research findings from past studies to determine a key set of variables for this analysis. The final dataset contains 24 variables with 92 attributes in total.

Table 1 lists the descriptive statistics of the crash-related variables. Motorcycle crashes that take place on the weekend are slightly over-represented compared to weekday motorcycle crashes. Daytime crashes represent nearly 75% of all motorcycle crashes in Louisiana. Spring and autumn are the worst seasons for motorcycle crashes.

These two seasons represent around 58% of all motorcycle crashes. In collision types, single motorcycle crashes have a higher percentage (around 31% of motorcycle crashes). The

Table 2. Vehicle and motorcycle rider related variables.

Variable Category	Perc (%)
Rider Gender	
Male	94.6
Female	5.5
Rider Age (MCAge)	
15–24	16.4
25–34	23.0
35–44	19.4
45–54	21.2
55–64	15.3
> 65	4.7
Severity Level (MCInj)	
Fatal	3.4
Severe	6.3
Moderate	33.2
Complaint	34.3
No Injury	22.9
First Harmful Event (FirstHarmEv)	
Collision with Vehicle	74.9
Run-off-Road (ROR) Right	12.8
Overturned/Rollover	6.9
ROR Left	3.1
Crossed Median/Centreline	2.4
Intoxication from All Involved in Crashes (AlcInvol)	
No	93.8
Yes	6.2
Number of Vehicles Involved	
Motorcycle only	26.6
Motorcycle and One Vehicle	68.0
Multiple Vehicles	5.4
Number of Occupants (NumOccupant)	
Motorcyclist only	92.1
With Pillion Riders	7.9
Rider Condition (MCCond)	
Normal	68.4
Inattentive	29.4
Impaired	2.2
Prior Movement (PriorMovement)	
Proceeding Straight Ahead	66.3
ROR	14.7
Stopped	12.4
Making Turn	6.7
Rider Violation (Violation)	
No Violations	56.9
Careless Operation	29.3
Failure to follow rules	11.2
Speeding	2.1
Improper Operation	0.4

Note: [a]Variable code, which is used in TCA analysis.

other three influential collision types are the rear end, right angle, and left turn. Two-way roadways, especially roadways with no physical separation, seem more prone to motorcycle crashes. Road departure motorcycle crashes (i.e. road relation as 'beyond shoulder' or 'shoulder') represent around 17% of all crashes. Approximately 88% of crashes occurred on roadways with no access control. Crashes on curved roadways comprise 18% of all crashes. Motorcycle crashes also have a higher representation in segment related crashes. Around 80% of motorcycle crashes happened on roadways with a speed limit in between 30 and 55 mph. Around 22% of the crashes happened in the dark with different lighting conditions. Maneuvering a motorcycle on rainy days is difficult, so it follows that 3.4% of crashes occurred in rainy conditions.

This study focused on rider related information. Table 2 lists percentage distributions of vehicle and motorcycle rider related attributes. According to the police reported KABCO scale, there is a 3.4% fatality rate for riders involved in motorcycle crashes, while there is only a 0.45% fatality rate for riders in motorcycle and other vehicle crashes all together. Louisiana data shows that male was the dominant gender group involved in motorcycle crashes. Around 95% of the motorcycle riders involved in crashes were male. The 25–34-year-old age group experienced the highest percentage of motorcycle crashes, followed by the 45–54-year-old age group. Next, in terms of crash involvement, the third highest group was the 35–44-year-old age group. In summary, the 25–55-year old motorcycle riders contribute 63% of all motorcycle crashes.

Around 10% of motorcycle riders were involved in fatal and severe crashes. Around 44% of the riders were involved in some sort of violation. More than half of the riders involved in crashes were in normal condition. Involvement with other vehicles also influenced motorcycle crashes. In Louisiana, 6.2% of riders and drivers were intoxicated in the motorcycle crash database. Inattentive riders were involved in 30% of the crashes. Around 8% of the riders had additional occupants in their motorcycles. Occupant injury and other associations were not explored in the current study.

Results and discussions

As a distribution-free method, TCA does not require any prior assumptions like other conventional statistical models. In recent years, different variants of CA (e.g. TCA or MCA) have gained popularity among researchers from various domains, such as traffic safety engineering. The sparsity in a data set can be defined as the percentage of zero abundances. Some of the common outliers in these datasets are rare observations, zero-block structure, and relatively high valued cells. There is a need to manage an abundance of data in crash datasets and generate insights in the presence of outliers. TCA, a sturdy-robust-resistant variation of CA, is a suitable tool for this analysis. This study used two phases to perform the analysis. In Phase 1, analysis was performed based on the locations of the attributes in the TCA plot. In Phase 2, individual row (i.e. consideration of each rider) level analysis was performed to identify the trends of key variables such as rider age, rider severity, number of vehicles involved, and types of access control.

Phase 1: clusters based on attribute locations

This method generates a graphical display, known as a TCA plot, which is useful for the general audience as it shows an n-dimensional display of the attribute locations to cut through the clutter of complex data dynamics. The display on two dimensions, also known as biplot, shows the co-occurrence of the variable attributes in a two-dimensional space, where proximity in the space indicates the similarity of the attributes. The complete TCA plot (see Figure 4) provides a general overview of the location of the attributes of the variables. Majority of the variables are shown in black, and variables that are suitable for Phase 2 analysis are shown by different colored texts.

A broad category difference can be found by analyzing the presence of the variables with reference to either the x-axis or the y-axis (Das and Dutta 2020). The value of the coordinates and mass (relative weightage of the attributes) measures of the key attributes are listed in Appendix (Table A1). Figure 4 identifies several key insights:

- The x-axis divides the age group into two broad categories (young and aging riders in one group showing in the lower side of the plot and middle-aged riders in another group showing in the upper side of the plot).
- The x-axis also divides access control into two groups: no access control in the upper side and full and partial control in the lower side. It is also seen that location of high posted speed limits (60 mph and above) are closer to full access control roadways.
- The y-axis divides the involvement of vehicles into two groups: the right side shows 'motorcycle only' crashes and the left side indicates 'multiple-vehicle' crashes.

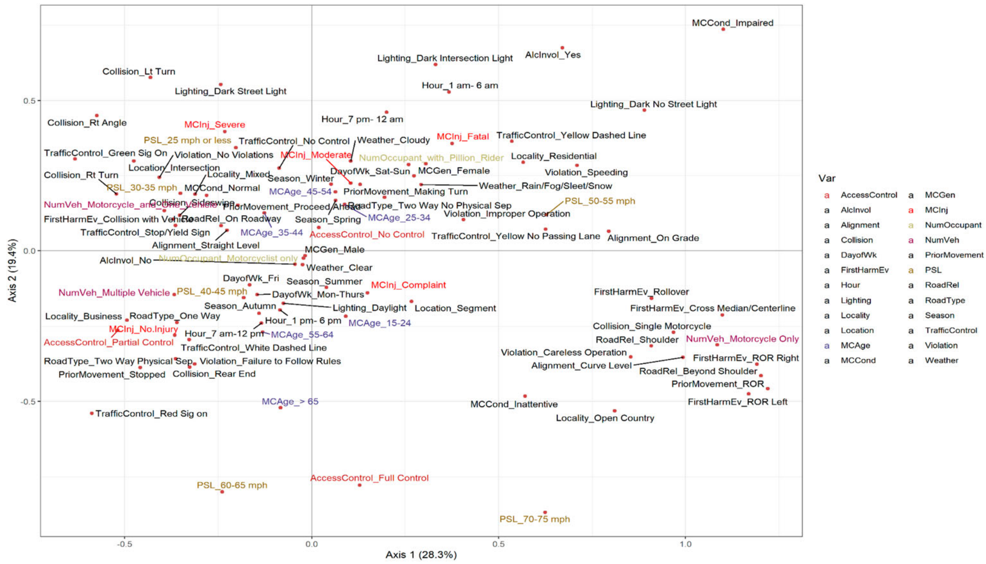

Figure 4. Principal MCA plot (based on column or variable based information) for the variable categories.

- Some of the attributes are far from the other groups of clusters of the variable attributes. These attributes are roadways with a posted speed limit of 60 mph and above (7.2% of all crashes), full access control roadways (7% of all crashes), riders that are older than 65 (4.7% of all riders), and a red traffic signal (5.2% of all crashes).
- Due to the lower occurrences and limited association with other attributes, these attributes represent lower co-occurrence in motorcycle crashes.

Clusters in the upper right (Figure 5a)

Cluster 1 (Var_Cluster01). The attributes in this cluster include dark with intersection lighting, dark with no lighting, nighttime, alcohol impairment, roadways with a yellow dashed line, and fatal injury. Usually, alcohol impairment occurs at night. This cluster indicates that fatal crashes are associated with two factors: intersection and alcohol impairment. These attributes describe fatal impaired crashes that occur mostly at intersections during the night-time and at segments with no lighting at night. Wu et al. (2018) also found that night-time and impaired crashes are key contributors to motorcycle crashes. Other studies (Saeed et al. 2019) show that alcohol-impaired crashes can be further explored by considering neighborhood effects and spatial effects. Countermeasures such as lighting at night and enforcement can be considered potential countermeasures.

Cluster 2 (Var_Cluster02). The attributes in this cluster are female riders with pillion riders, inclement (rain/fog/sleet/snow) weather, and weekend. The cluster indicates that the presence of pillion riders during inclement weather increases the risk of motorcycle crashes of the female riders. It is important to notice that this cluster is close to Cluster 1, which represents attributes associated with fatal crashes. Multiple studies (Rifaat, Tay, and de Barros 2012 and Savolainen and Mannering 2007) observed that female riders were more likely to

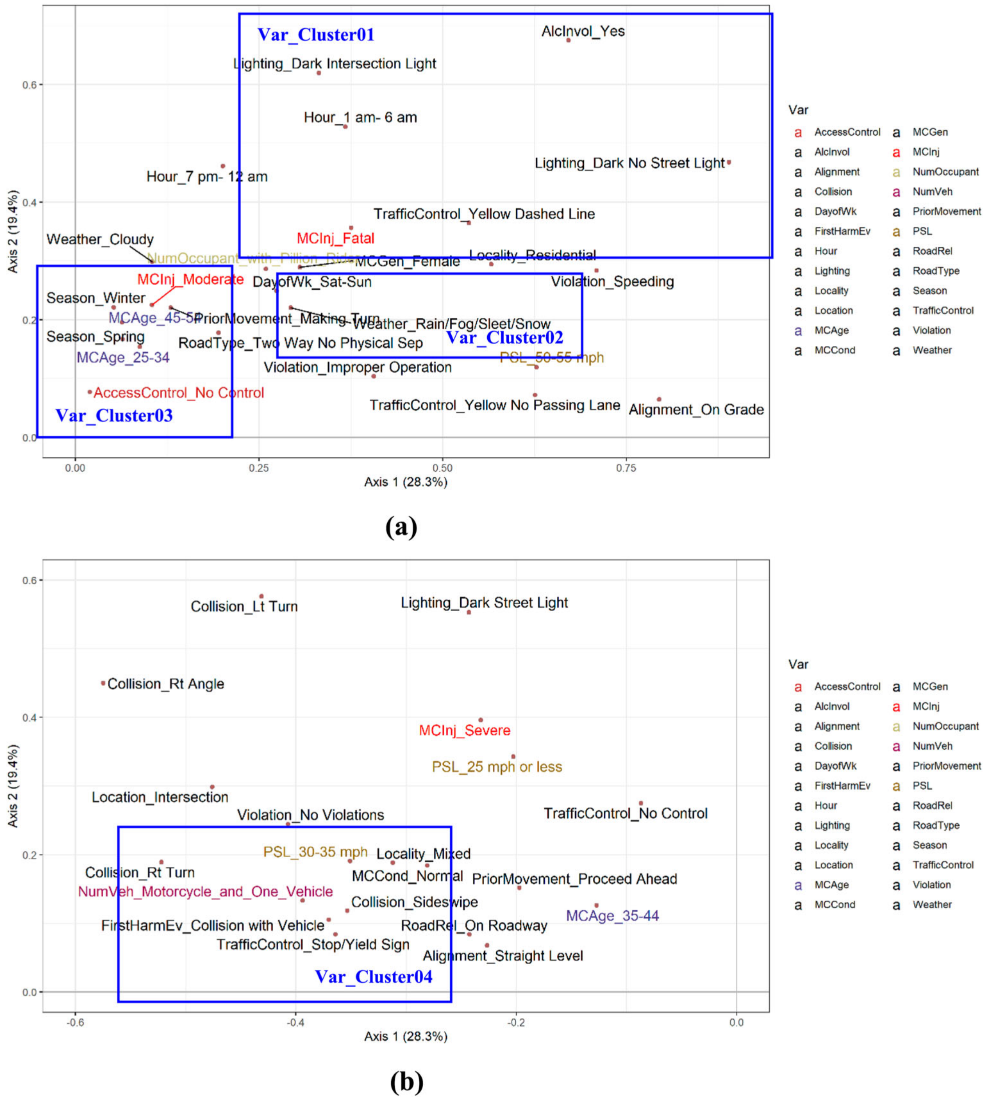

Figure 5. MCA plot (based on column or variable based information) of the upper right and upper left.

be involved in crashes with a high severity level than male riders. However, the findings of Chung, Song, and Yoon (2014) contradicted this finding. In the study of Quddus, Noland, and Chin (2002), gender did not significantly affect the outcomes of the injury severity (fatal vs. not fatal); however, results showed that women suffered more severe injuries than men when they were injured. This unique finding requires additional investigation in understanding the role of other latent features.

Cluster 3 (Var_Cluster03). The attributes in this cluster are cloudy weather, winter or spring as season, rider age groups (25–34 years, and 45–54 years), moderate injury, making a turn as prior movement, roadways with no access control (i.e. non-freeways), and two-way roadways with no separation. This cluster mostly shows the attributes that are dominant on rural two-lane undivided roadways with no access control. Turn movement is an issue on these roadways. These crashes are also associated with moderate injuries. Eustace, Indupuru, and Hovey (2011) also showed similar findings regarding this association. Improvement on two-way roadways with no separation will be beneficial in reducing crashes associated with this cluster attributes.

Clusters in the upper left (Figure 5b)

Cluster 4 (Var_Cluster04). The attributes in this cluster are mixed locality (residential and business), sideswipe and right turn collisions, collision with a vehicle as the first harmful event, unsignalized intersection, lower posted speed limit (30–35 mph), no violation by the rider, and normal rider condition. The cluster shows that not-at-fault riders are involved in multiple-vehicle turning or sideswipe crashes on low-speed unsignalized intersections. This cluster mostly highlights the association between

Clusters in the lower left (Figure 6a)

Cluster 5 (Var_Cluster05). The attributes in this cluster are one-way roadway with a white dashed line, no rider injury, a roadway with physical separation, roadways with partial access control, business locality, multiple-vehicle crashes, stopped as prior movement, rear-end crash, and rider failure to follow rules as the violation. This cluster illustrates the association of the violations (e.g. failure to follow the rules) with intersection related incidents and rear-end collisions. Chawla, Karaca, and Savolainen (2019) also showed that traffic violation contributes significantly to motorcycle crashes.

Cluster 6 (Var_Cluster06). The attributes in this cluster are male riders, weekday, clear weather, age group 55–64 years, daytime, autumn season, no alcohol impairment, no pillion riders, roadways with 40–45 mph posted speed limit. The majority of these attributes (for example, male riders = 95%, no occupants = 93%, non-impaired = 94%, clear weather = 86%, daytime = 70%) are clustered together as these attributes are highly representative in motorcycle crash data. The other attributes (55–64 years old = 15%, autumn = 27%, 40–45 mph roadways = 33%) indicate that these attributes mostly occur with highly representative attributes. The common attributes in the cluster were found to be significant by other studies as well (Wu et al. 2018; Chawla, Karaca, and Savolainen 2019; Xin et al. 2019).

Clusters in the lower right (Figure 6b)

Cluster 7 (Var_Cluster07). The attributes in this cluster are young (15–24 years) riders, complaint injury, summer season, and segment related crash. This cluster shows that young rider crashes are associated with complaint injury and segment related crashes. It is common that summer months are associated with high motorcycle exposure with more teen riders on the road (ROSPA 2017). It is also obvious that the severity type of this cluster is associated with the young age of the riders. It can be explained that riders generally use popular motorbike routes during summer for recreational purpose. Rider licensing program improvement and rider education can help in lowering young rider motorcycle crashes. Additional enforcement along the popular motorcycle routes during summer months can also be beneficial.

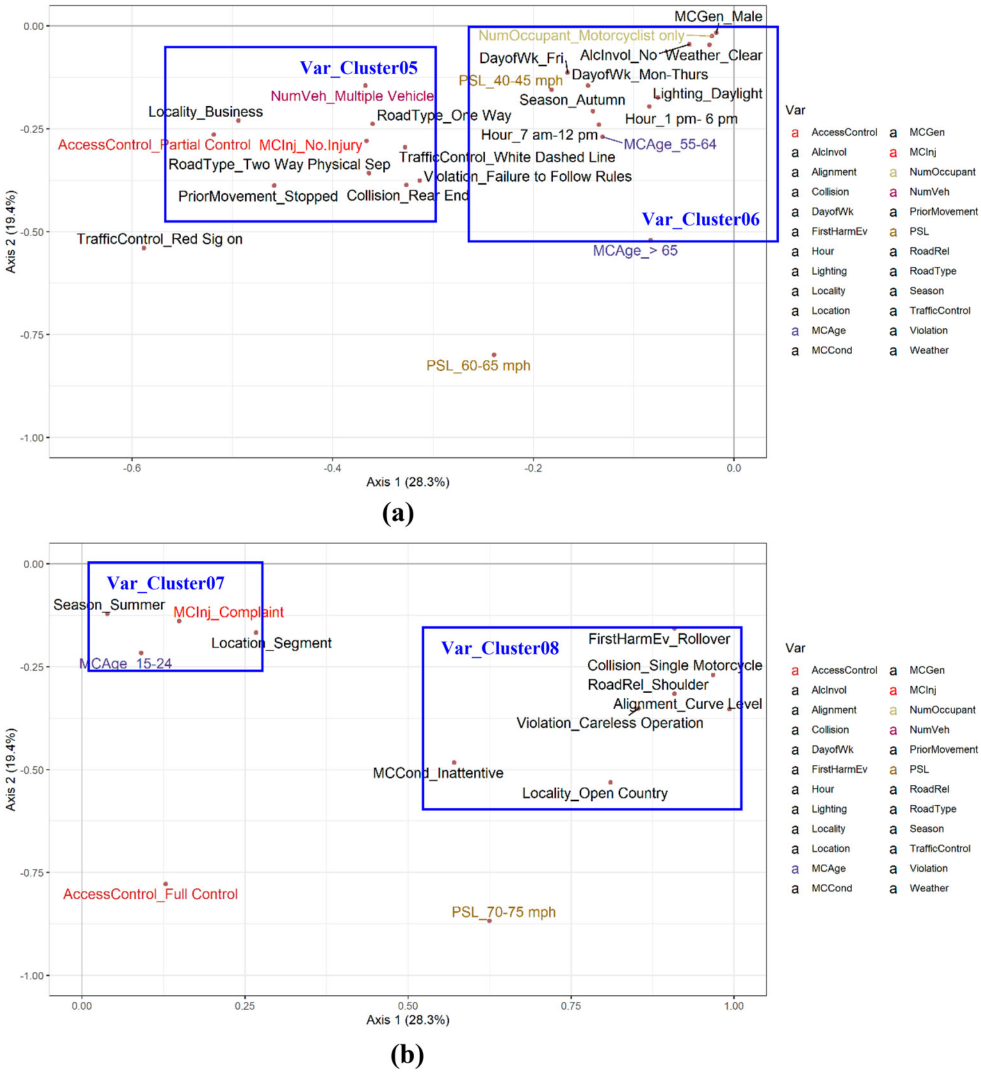

Figure 6. MCA plot (based on column or variable based information) of the lower right and lower left.

Cluster 8 (Var_Cluster08). The attributes in this cluster are non-collision with another vehicle (i.e. single motorcycle), curve alignment, run off-road, inattentive rider, open country roadways, rollover crash. The patterns of these crashes indicate that most of these crashes are roadway departure related (i.e. single motorcycle, run off-road, and rollover crashes) crashes on open country roadways. This cluster also shows that these crashes occur on curve aligned roadways. Waseem, Ahmed, and Saeed (2019) suggested several strategies, including reduction of posted speed limits on highways with a greater motorcycle ratio, and dislocation or removal of fixed roadside objects. Xin et al. (2019) showed that sharp (radius $\leq$ 1500 ft) non-reverse curves were the riskiest curve design for motorcyclists, followed by sharp reverse curves and moderate (1500 ft $<$ radius $\leq$ 3000 ft) reverse curves. Other studies (Wu et al. 2018) also found that curve is a significant factor in motorcycle crashes. Improving rider awareness by rider education can be useful to reduce crashes associated with this cluster.

Phase 2: clusters based on locations of individuals

Figure 7 shows the distribution of crash occurrences (each row represents an entry for a motorcyclist involved in a crash) on both axes. Twenty clusters (each cluster is shown by a vertical line on the x-axis) were developed from TCA analysis (based on the x-axis). Figure 7 shows that clusters are present on both sides of the x-axis. The point size indicates the group size of the individuals.

Row level analysis (individual rider level) was conducted using TCA outcomes. The location of the coordinates from the individual level analysis generated 20 cluster groups (not similar to the variable attribute clusters discussed before). In this analysis, four key qualitative variables (rider age, number of vehicles involved, rider severity type, and access control of the facilities) were used to describe the clusters. This analysis was completed by computing the log-odds ratio (LOR) of yes to no for each cluster with respect to the marginal distribution. For example, in the age variable, 1 represents young (15–24 years) riders, and 0 represents not young (greater than 24 years) riders.

$$LOR_i = ln\left(\frac{n_{1i}/n_{0i}}{N_1/N_0}\right)$$

Where, LOR_i = log-odds ratio for the i^{th} cluster for a variable; N_1 = frequency of 'yes' attribute in a variable; N_0 = frequency of 'no' attribute in a variable; n_{1i} = frequency of 'yes' attribute in a variable (for cluster i); n_{0i} = frequency of 'no' attribute in a variable (for cluster i).

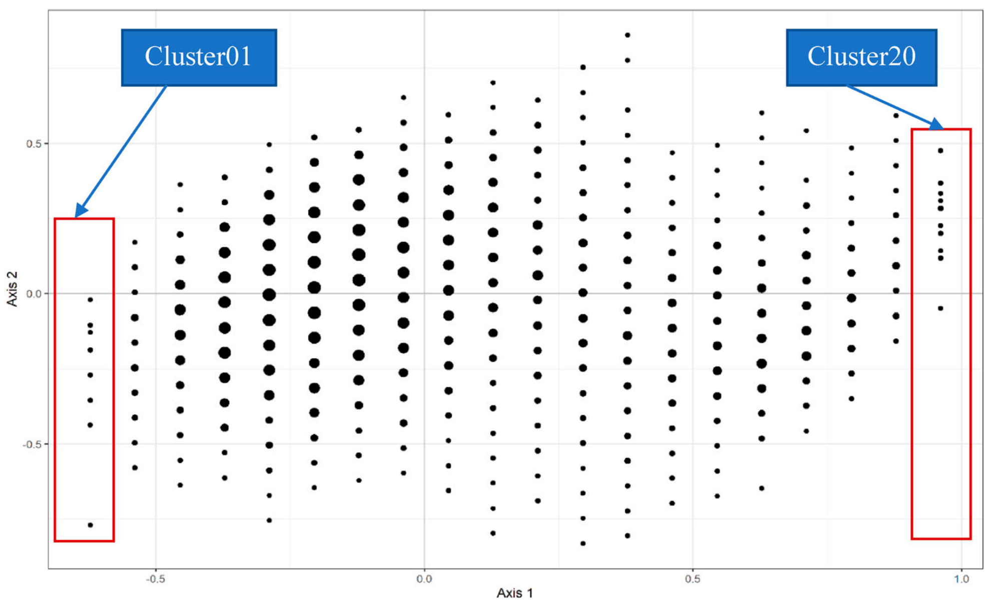

Figure 7. Clusters from row-based analysis (each row represents an entry for a motorcyclist involved in a crash).

Using '*Single vs. Multiple Vehicle Involvement*' as an example, the interpretation of LOR (S = s) can be given by:

- LOR = 0, then the proportion of single-vehicle crashes in *i*th cluster equals the proportion of multiple-vehicle crashes in the sample.
- LOR > 0, then the proportion of single-vehicle crashes in cluster *i* is greater than the proportion of multiple-vehicle crashes in the sample. That is, the cluster *i* is positively associated with multiple-vehicle crashes and negatively associated with multiple-vehicle crashes.
- LOR < 0, then the proportion of multiple-vehicle crashes in cluster *i* is larger than the proportion of single-vehicle crashes in the sample. That is, the cluster *i* is positively associated with multiple-vehicle crashes and negatively associated with single-vehicle crashes.

Some of the key findings from Table 3 are:

- Six clusters (Cluster04, Cluster05, Cluster06, Cluster07, Cluster08, and Cluster09) contain around 63% of all riders.
- Ten clusters show positive LOR measures for young riders. Seven clusters (Cluster08, Cluster09, Cluster10, Cluster15, Cluster16, Cluster18, and Cluster20) show positive LOR values for impaired and KAB crashes. Five clusters (Cluster10, Cluster11, Cluster12, Cluster15 and Cluster17) indicate single-vehicle involvement. This finding indicates that young male riders are more susceptible to KAB and single-vehicle crashes.
- Seven clusters (Cluster10, Cluster11, Cluster12, Cluster13, Cluster14, Cluster15, and Cluster17) show positive LOR values for single-vehicle crashes. Young riders are overrepresented in these clusters, so are the KAB crashes and no access control roadways.
- Eleven clusters have positive LOR values for KAB crashes. These clusters also show a high likelihood of young riders, single-vehicle crashes, and crashes on 'no access control' roadways.
- Seven clusters (Cluster01, Cluster02, Cluster03, Cluster04, Cluster05, Cluster09, and Cluster13) show a high likelihood of freeway crashes (full or partial access control roadways). These clusters show a high likelihood of riders above 24 years old, multiple-vehicle crashes, and CO crashes.
- Each of these clusters has quantitative measures. Explanations can be provided to interpret all of these clusters. For example, Cluster06 has 13.36 percent of all motorcycle riders. For rider age, vehicle involvement, and KAB injuries, LOR values are negative. It indicates that the proportion of 'not young riders' is larger in this cluster than the proportion of young rider. It also indicates that the proportion of KAB crashes is lower than the proportion of CO crashes in this cluster. The same is true for single-vehicle involvement. This cluster shows positive LOR value for 'access control' roadways, which indicates that the proportion of full or partial access control roadway (i.e. freeway) is greater than 'no access control' roadways. The overall interpretation is that this cluster is associated with freeway related single-vehicle crashes that involve riders older than 24 years old with CO injuries.

Table 3. Odds ratio by row-based clusters of individuals and attribute groups.

Cluster	Count	15–24 years vs. 24 years and above riders	Single vs. Multiple Vehicle Involvement	KAB vs. CO Crash	No Access Control vs. Access Control
Cluster01	35			−1.034	−1.884
Cluster02	167	−0.858		−0.629	−1.473
Cluster03	542	−0.847	−3.756	−0.355	−0.512
Cluster04	1102	−0.468	−3.255	−0.307	−0.302
Cluster05	1495	−0.113	−2.967	−0.122	−0.108
Cluster06	1691	−0.022	−1.824	−0.053	0.022
Cluster07	1647	0.170	−1.347	−0.008	0.171
Cluster08	1129	0.121	−0.865	0.076	0.066
Cluster09	829	0.378	−0.218	0.070	−0.110
Cluster10	620	0.090	0.288	0.104	0.080
Cluster11	544	0.145	0.887	−0.004	0.056
Cluster12	461	0.255	1.466	−0.114	0.216
Cluster13	417	−0.014	2.752	0.182	−0.316
Cluster14	390	−0.106	3.947	0.241	0.028
Cluster15	377	0.141	4.520	0.229	0.629
Cluster16	426	0.169		0.331	0.618
Cluster17	373	−0.131	6.139	0.288	1.694
Cluster18	242	0.173		0.562	1.055
Cluster19	121	−0.354		0.593	2.185
Cluster20	49	0.374		0.981	

TCA is a recent addition among various CA methods (for example, multiple correspondence analysis or MCA, Nonlinear Optimal Scaling Method or NOSM, Categorical Principal Component Analysis or CatPCA, Nonlinear Canonical Correlation Analysis or NLCCA). TCA generates a proximity map of the variable categories, or attributes, in a low-dimensional plane by revealing the key patterns or clusters from a multi-dimensional dataset. TCA can also be used in determining the quantitative measures in the form of 'log odds ratio.' This study identified several risk clusters that require additional attention to design suitable countermeasures. The findings of attribute-based and individual-based clusters can be beneficial in reducing motorcycle crashes.

Conclusions

Motorcycle riding is considered one of the most intense modes of transportation. Many studies have examined motorcycle crash data in order to investigate the influencing factors. However, the number of motorcycle crashes in the U.S. is still at an unacceptably high level. The traditional approach of investigating the effect of a single factor on the response variable (either count or injury type) is not enough to characterize complex crash dynamics. Very few studies have measured clusters of attributes that contribute to motorcycle crash occurrence. Chung and Song (2018) applied two categorical data analysis methods (CatPCA, and NLCCA) to Korean motorcycle crash data from 2009 to determine the critical factors that affect motorcycle crash severity. The current study is unique from Chung and Song (2018). The current paper makes two major contributions. First, it performed a U.S. based study that included multiple years (2010–2016) of recent motorcycle crash data. Second, it used not only an advanced categorical data analysis method but also developed 'log odds ratio' to provide a quantitative insight of the results. Such insights are not provided in the earlier categorical data analysis using crash data. Thus, this is a timely study showing the opportunities of TCA in analyzing complex and unsupervised (dataset with no predefined response variable) crash data.

To mitigate this research gap, TCA is a viable tool for analyzing complex categorical data in search of meaningful associations between categorical attributes. This method helps in understanding diverse variable categories and produces visual results from the key associations. The motorcycle crash dataset has a small number of cases, so eliminating entries with noise will shrink the dataset significantly. The use of TCA analysis is advantageous because it allows for the removal of noise in the dataset (Das and Dutta 2020). This study prioritizes certain key clusters, as well as target countermeasures listed in this study, which will help authorities to reduce future motorcycle crashes.

This study used a wide variety of roadway and rider characteristics to identify several key risk clusters from attribute level analysis: impaired crashes at roadways with no lighting at night, rural two-lane undivided roadways with no access control, turning related crashes at mixed localities, violation related rear-end crashes, young male riders, roadway departure crashes on curved alignments, and female riders with pillion riders in inclement weather. This study also identified 20 clusters based on the characteristics of individual crashes. Young riders are disproportionately higher among single-vehicle crashes and KAB crashes. Clusters with KAB crashes have a high likelihood of having the following characteristics: single-vehicle crashes, and crashes on 'no access control' roadways. Freeway crashes have a high likelihood of having the following characteristics: riders above 24 years old, multiple-vehicle crashes, and CO crashes. The key groups of the confluence of factors can further be examined to recognize appropriate countermeasures. The log-odds ratio values provide quantitative measures of the key attributes, which can be useful in formulating mitigation strategies (by the safety engineers and policy makers) to make data-driven decisions in improving motorcycle safety.

The current study is not without limitations. First, this study provides limited information on potential countermeasures to reduce motorcycle-specific traffic crashes and associated injuries. There is a need for extensive research to determine the effectiveness of countermeasures based on motorcycle crash data over time. Second, the groups of confluence factors are developed based on only two dimensions, which represent a 49% variance of the complete database. Usage of the NLCCA for multiple planes (using various combinations of axes) could be a potential alternative. The limitations of the current study offer various directions for future research in motorcycle crash investigation.

Disclaimer

The contents of this paper reflect the views of the authors and not the official views or policies of the LADOTD.

Acknowledgments

The author likes to thank three anonymous reviewers for their excellent suggestions. The author also likes to thank Magdalena Theel for a through technical review.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753

References

Benzećri, J. P. 1992. *Correspondence Analysis Handbook*. Boca Raton, FL: CRC Press.

Chawla, H., I. Karaca, and P. Savolainen. 2019. "Contrasting Crash- and Non-Crash-Involved Riders: Analysis of Data from the Motorcycle Crash Causation Study." *Transportation Research Record: Journal of the Transportation Research Board* 2673 (7): 122–131.

Cheng, W., G. S. Gill, T. Sakrani, M. Dasu, and J. Zhou. 2017. "Predicting Motorcycle Crash Injury Severity Using Weather Data and Alternative Bayesian Multivariate Crash Frequency Models." *Accident Analysis & Prevention* 108: 172–180.

Choulakian, V. 2006a. "L1-Norm Projection Pursuit Principal Component Analysis." *Computational Statistics & Data Analysis* 50: 1441–1451.

Choulakian, V. 2006b. "Taxicab Correspondence Analysis." *Psychometrika* 71: 333–345.

Choulakian, V. 2013. "Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis." *Journal of Classification* 30: 397–427.

Chung, Y., and T. Song. 2018. "Safety Analysis of Motorcycle Crashes in Seoul Metropolitan Area, South Korea: An Application of Nonlinear Optimal Scaling Methods." *International Journal of Environmental Research and Public Health* 15 (12): 2702.

Chung, Y., T. J. Song, and B. J. Yoon. 2014. "Injury Severity in Delivery-Motorcycle to Vehicle Crashes in the Seoul Metropolitan Area." *Accident Analysis and Prevention* 62: 79–86.

Cohn, S. M., D. Mishkin, A. Castelblanco, P. Li, C. Popkin, R. Duncan, G. A. Hotz, and Taylor & Francis. 2004. "Outcome of Motorcycle Riders at One Year Post-Injury." *Traffic Injury Prevention* 5 (1): 87–89.

Das, S., S. Ashraf, A. Dutta, and L. Tran. 2020a. "Pedestrians Under Influence (PUI) Crashes: Patterns From Correspondence Regression Analysis." *Journal of Safety Research*. doi:10.1016/j.jsr.2020.07.001.

Das, S., R. Avelar, K. Dixon, and X. Sun. 2018a. "Investigation on the Wrong Way Driving Crash Patterns Using Multiple Correspondence Analysis." *Accident Analysis & Prevention* 111: 43–55.

Das, S., and A. Dutta. 2020. "Extremely Serious Crashes on Urban Roadway Networks: Patterns and Trends." *IATSS Research*. doi:10.1016/j.iatssr.2020.01.003.

Das, S., A. Dutta, K. Dixon, L. Minjares-Kyle, and G. Gillette. 2018b. "Using Deep Learning in Severity Analysis of At-Fault Motorcycle Rider Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2672 (34): 122–134.

Das, S., A. Dutta, and K. Fitzpatrick. 2020b. "Technological Perception on Autonomous Vehicles: Perspectives of the Non-Motorists." *Technology Analysis & Strategic Management*. doi:10.1080/09537325.2020.1768235

Das, S., M. Islam, A. Dutta, and T. Shimu. 2020c. "Uncovering Deep Structure of Determinants in Large Truck Fatal Crashes." *Transportation Research Record: Journal of Transportation Research Board*. doi:10.1177/0361198120931507

Das, S., K. Jha, K. Fitzpatrick, M. Brewer, and T. H. Shimu. 2019. "Pattern Identification from Older Bicyclist Fatal Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2673 (6): 638–649.

Das, S., and X. Sun. 2015. "Factor Association with Multiple Correspondence Analysis in Vehicle–Pedestrian Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2519 (1): 95–103.

Das, S., and X. Sun. 2016. "Association Knowledge for Fatal Run-off-Road Crashes by Multiple Correspondence Analysis." *IATSS Research* 39 (2): 146–155.

Das, S., L. Tran, and M. Theel. 2020d. "Understanding Patterns in Marijuana Impaired Traffic Crashes." *Journal of Substance Use*. doi: 10.1080/14659891.2020.1760381

De Rome, L. 2005. A Model for Motorcycle Road Safety Plans.

Eustace, D., V. K. Indupuru, and P. Hovey. 2011. "Identification of Risk Factors Associated with Motorcycle-Related Fatalities Ohio." *Journal of Transportation Engineering* 137 (7): 474–480.

Greenacre, M., and J. Blasius. 2006. *Multiple Correspondence Analysis and Related Methods*. New York, NY: Chapman and Hall/CRC.
Hjellbrekke, J. 2018. *Multiple Correspondence Analysis for the Social Sciences*. Boca Raton, FL: Routledge.
Jalayer, M., M. Pour-Rouholamin, and H. Zhou. 2018. "Wrong-Way Driving Crashes: A Multiple Correspondence Approach to Identify Contributing Factors." *Traffic Injury Prevention* 19 (1): 35–41.
Khor, D., K. Inaba, A. Aiolfi, S. Delapena, E. Benjamin, K. Matsushima, A. M. Strumwasser, and D. Demetriades. 2017. "The Impact of Helmet Use on Outcomes after a Motorcycle Crash." *Injury* 48 (5): 1093–1097.
Kostyniuk, L. P., and A. D. Nation. 2005. Motorcycle Crash Trends in Michigan: 2001-2005. Report No. UMTRI-2006-22. Transportation Research Institute, The University of Michigan, Ann Arbor, MI.
Le Roux, Brigitte, and H. Rouanet. 2010. *Multiple Correspondence Analysis*. Vol. 163. Thousand Oaks, CA: Sage.
Medina, A., and J. Soto. 2011. Roadway Factors Associated with Motorcycle Crashes.
Murtagh, F. 2005. *Correspondence Analysis and Data Coding with Java and R*. New York, NY: Chapman and Hall/CRC.
Naumann, R. B., A. M. Dellinger, B. A. West, J. L. Annest, and G. W. Ryan. 2010. "Motorcycle Injuries Across the USA, 2001–2008: Injury Comparisons and Projections." *Injury Prevention* 16. doi:10.1136/ip.2010.029215.297.
NHTSA. 2019. Traffic Safety Facts: Motorcycle. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812492. Accessed: July 2019.
Ogle, A. J., and E. A. Tillotson. 2008. "Should It Be the Law in Wisconsin That All Motorcyclists Are Required to Wear Helmets?" *Journal of Trauma Nursing* 15 (2): 43–46.
Peek-Asa, C., D. L. McArthur, J. F. Kraus, and Elsevier. 1999. "The Prevalence of Non-Standard Helmet Use and Head Injuries among Motorcycle Riders." *Accident Analysis & Prevention* 31 (3): 229–233.
Quddus, M. A., R. B. Noland, and H. C. Chin. 2002. "An Analysis of Motorcycle Injury and Vehicle Damage Severity Using Ordered Probit Models." *Journal of Safety Research* 33: 445–462.
Rappole, C., M. Canham-Chervak, B. Taylor, and B. Jones. 2019. "Factors Associated with Motorcycle Traffic Crash Fatalities among Active Duty U.S. Army Personnel." *Traffic Injury Prevention* 20: 174–181.
Rifaat, S. M., R. Tay, and A. de Barros. 2012. "Severity of Motorcycle Crashes in Calgary." *Accident Analysis and Prevention* 49: 44–49.
The Royal Society for the Prevention of Accidents (RoSPA). 2017. *Road Safety Research: Common Motorcycle Crashes Causes*. Birmingham, UK: RoSPA.
Ryb, G. E., P. C. Dischinger, J. A. Kufera, and T. J. Kerns. 2009. "Trends in Motorcycle Injuries, Deaths and Case Fatality Rates in the U.S.A. Population." *Annals of Advances in Automotive Medicine* 53.
Saeed, T. U., R. Nateghi, T. Hall, and B. Waldorf. 2019. "Statistical Analysis of Area-Wide Alcohol-Related Driving Crashes: A Spatial Econometric Approach." *Geographical Analysis* 1–24. doi:10.1111/gean.12216.
Savolainen, P., and F. Mannering. 2007. "Probalistic Models of Motorcyclists' Injury Severities in Single- and Multi-Vehicle Crashes." *Accident Analysis and Prevention* 39: 955–963.
Shaheed, M. S., and K. Gkritza. 2014. "A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes." *Analytic Methods in Accident Research* 2: 30–38.
Waseem, M., A. Ahmed, and T. U. Saeed. 2019. "Factors Affecting Motorcyclists' Injury Severities: An Empirical Assessment Using Random Parameters Logit Model with Heterogeneity in Means and Variances." *Accident Analysis & Prevention* 123: 12–19.
Wu, K., L. Sasidharan, C. Thor, and S. Chen. 2018. "Crash Sequence-Based Risk Matrix for Motorcycle Crashes." *Accident Analysis & Prevention* 117: 21–31.
Xin, C., Z. Wang, C. Lee, P. Lin, T. Chen, R. Guo, and Q. Lu. 2019. "Development of Crash Modification Factors of Horizontal Curve Design Features for Single-Motorcycle Crashes on Rural Two-Lane Highways: A Matched Case-Control Study." *Accident Analysis & Prevention* 123: 51–59.

Appendix

Table A1. Coordinates and mass values of the variable attributes.

Variable	Attribute	Axis1	Axis2	Mass	Variable	Attribute	Axis1	Axis2	Mass
AccessControl	Full Control	0.129	−0.778	0.003	MCInj	Complaint	0.149	−0.139	0.014
	No Control	0.019	0.077	0.037		Fatal	0.375	0.356	0.001
	Partial Control	−0.519	−0.264	0.002		Moderate	0.104	0.225	0.014
AlcInvol	No	−0.044	−0.045	0.039		No.Injury	−0.366	−0.280	0.010
	Yes	0.671	0.675	0.003		Severe	−0.232	0.396	0.003
Alignment	Curve Level	0.993	−0.353	0.007	NumOccupant	With Pillion Rider	0.259	0.287	0.003
	On Grade	0.795	0.065	0.001		Motorcyclist only	−0.022	−0.024	0.038
	Straight Level	−0.227	0.068	0.034	NumVeh	Multiple Vehicle	−0.367	−0.145	0.002
Collision	Lt Turn	−0.431	0.577	0.006		Motorcycle and One Vehicle	−0.394	0.133	0.028
	Single Motorcycle	0.968	−0.270	0.013		Motorcycle Only	1.084	−0.312	0.011
	Rear End	−0.326	−0.386	0.011	PriorMovement	Making Turn	0.130	0.220	0.003
	Rt Angle	−0.575	0.450	0.008		Proceed Ahead	−0.197	0.152	0.028
	Rt Turn	−0.522	0.189	0.001		ROR	1.220	−0.458	0.006
	Sideswipe	−0.353	0.118	0.003		Stopped	−0.458	−0.388	0.005
DayofWk	Fri	−0.166	−0.114	0.007	PSL	25 mph or less	−0.203	0.343	0.004
	Mon-Thurs	−0.146	−0.145	0.020		30–35 mph	−0.351	0.191	0.010
	Sat-Sun	0.274	0.249	0.015		40–45 mph	−0.182	−0.155	0.014
FirstHarmEv	Collision with Vehicle	−0.370	0.105	0.031		50–55 mph	0.627	0.120	0.010
	Cross Median/Centerline	1.099	−0.213	0.001		60–65 mph	−0.239	−0.800	0.002
	Rollover	0.909	−0.157	0.003		70–75 mph	0.625	−0.868	0.001
	ROR Left	1.168	−0.475	0.001	RoadRel	Beyond Shoulder	1.202	−0.415	0.006
	ROR Right	1.191	−0.377	0.005		On Roadway	−0.242	0.084	0.035
Hour	1 am–6 am	0.367	0.528	0.002		Shoulder	0.908	−0.316	0.001
	1 pm–6 pm	−0.084	−0.196	0.020	RoadType	One Way	−0.360	−0.238	0.003
	7 am–12 pm	−0.135	−0.240	0.008		Two Way No Physical Sep	0.195	0.178	0.027
	7 pm–12 am	0.201	0.461	0.011		Two Way Physical Sep	−0.364	−0.358	0.011
Lighting	Dark Street Light	−0.243	0.553	0.005	Season	Autumn	−0.141	−0.207	0.011
	Dark No Street Light	0.890	0.468	0.004		Spring	0.064	0.167	0.013
	Dark Intersection Light	0.331	0.620	0.001		Summer	0.039	−0.122	0.011
	Daylight	−0.076	−0.174	0.031		Winter	0.052	0.221	0.007
Locality	Business	−0.494	−0.230	0.013	TrafficControl	Green Sig On	−0.633	0.305	0.003
	Mixed	−0.312	0.188	0.012		No Control	−0.087	0.275	0.006
	Open Country	0.811	−0.531	0.005		Red Sig on	−0.588	−0.539	0.002
	Residential	0.566	0.294	0.011		Stop/Yield Sign	−0.364	0.084	0.002

(*continued*)

Table A1. Continued.

Variable	Attribute	Axis1	Axis2	Mass	Variable	Attribute	Axis1	Axis2	Mass
Location	Intersection	−0.476	0.298	0.015		White Dashed Line	−0.328	−0.295	0.013
	Segment	0.267	−0.167	0.027		Yellow Dashed Line	0.535	0.365	0.004
MCAge	> 65	−0.083	−0.521	0.002		Yellow No Passing Lane	0.626	0.072	0.011
	15–24	0.091	−0.217	0.007	Violation	Careless Operation	0.853	−0.352	0.012
	25–34	0.088	0.154	0.010		Failure to Follow Rules	−0.313	−0.376	0.005
	35–44	−0.127	0.126	0.008		Improper Operation	0.406	0.104	0.000
	45–54	0.063	0.196	0.009		No Violations	−0.407	0.244	0.024
	55–64	−0.131	−0.269	0.006		Speeding	0.710	0.284	0.001
MCCond	Impaired	1.101	0.737	0.001	Weather	Clear	−0.025	−0.046	0.036
	Inattentive	0.571	−0.483	0.012		Cloudy	0.104	0.299	0.004
	Normal	−0.281	0.184	0.028		Rain/Fog/Sleet/Snow	0.293	0.220	0.001
MCGen	Female	0.305	0.289	0.002					
	Male	−0.018	−0.017	0.039					