Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–11

DOI: 10.1177/03611981211037891
journals.sagepub.com/home/trr
SAGE

Identifying Patterns of Key Factors in Sun Glare-Related Traffic Crashes

Subasish Das[1], Xiaoduan Sun[2], Bahar Dadashova[1], M. Ashifur Rahman[2], and Ming Sun[2]

Abstract

Sun glare is one of the major environmental issues contributing to traffic crashes. Every year, many traffic crashes in the United States are attributed to sun glare. However, quantitative analysis of the influence of sun glare on traffic crashes has not been widely undertaken. This study used traffic crash narrative data for 7 years (2010–2016) from Louisiana to identify crash reports that provided evidence of drivers indicating sun glare as the primary contributing factor of the crashes. Additional geometry and traffic information was collected to identify the list of key crash-contributing factors. This study used cluster correspondence analysis to perform the data analysis. After performing several iterations, six clusters were identified that provided additional insight in relation to sun glare-related crashes. The six clusters are associated with mixed (business and residential) localities, intersection-related crashes on U.S. roadways, single-vehicle crashes on residential two-lane undivided roadways, curve-related crashes on parish roadways in residential localities, interstate-related crashes in open country localities, and curve-related crashes in open country localities. The findings of the current study can add insights to the ongoing safety analysis on sun glare-related crashes.

Keywords

crash data, safety, safety performance and analysis

Sun glare occurs when the sun is low on the horizon, usually an hour after dawn and before dusk. If the sun glare is acute in nature, drivers usually experience a temporary dazzling sensation that may turn into impairment of vision (*1*). The safety of all road users may be negatively affected by this impairment. It can be noted that most modern-day vehicles have sun visors installed. However, distraction caused by sun glare can occur if the driver makes an unexpected turn, shifting their vision range with the sun close to the immediate horizon (*2*). The National Highway Traffic Safety Administration reported sun glare as one of the safety concerns in several hundred fatalities (*3*). Sun glare can be considered a significant safety concern, and additional research is needed to assess its nature and impact on road safety.

In many cases, the total number of traffic crashes attributable to sun glare is under-reported. This is mostly as a consequence of the nature of police crash reports. In many cases, the law enforcement official considers sun glare as one of many causes, but not as the sole cause, of a crash, because of the overwhelmingly high number of drivers who report that they are temporally blinded by sun glare. The objective of this paper is to evaluate the impact of sun glare on road safety in Louisiana as a case study. To perform the data analysis, this study collected 7 years (2010–2016) of traffic crash data and police-reported crash narratives from Louisiana. As crash data contain a wide range of information, this study selected variables of importance based on the information value of the variable categories. The current study applied cluster correspondence analysis to perform the analysis in a way to select the association between key contributing factors for sun glare-related crashes.

The rest of the paper is described as follows. First, a brief literature review is provided. Then, in the

[1]Texas A&M Transportation Institute, Bryan, TX
[2]Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA

Corresponding Author:
Subasish Das, s-das@tti.tamu.edu

methodology section, data preparation and exploratory data analysis are described with a brief introduction to the theory of cluster correspondence analysis. The next section describes the results of the modeling with discussions. The last section is conclusions, which describe overall findings, limitations, and future directions.

Literature Review

Although earlier studies have identified direct exposure to sunlight as an important contributory factor in crashes, with the ability to impair visual performance by causing temporary blindness (*4*, *5*), only limited research has been performed.

Mitra obtained sunrise and sunset data from the National Oceanic and Atmospheric Association (NOAA) to identify glare-related crashes and compute daytime windows with the worst possible glares (*6*). Using ANOVA and chi-square test for proportions, the statistical tests confirm the strong influence of sun glare on traffic collisions. Statistical evidence also shows that the effects of glare are worse in early spring and early fall, and less during summer. In a follow-up study using data from signalized intersections in Tucson, Arizona, Mitra assessed how sun glare affects intersection safety (*7*). A comparison analysis was performed to differentiate crashes that are caused by sun glare (morning and evening) from those unaffected by glare. Sunrise and sunset data from the NOAA were used to compute windows with the worst possible glare scenarios. The granular-level analysis suggests that odds of glare-related crash incidence are higher in east- and west-bound compared with north- and south-bound directions. The harmful effects of glare are found to be low during the summer months. The results also show that sun glare affects rear-end and right-angle crashes at signalized intersections.

Hagita and Mori analyzed crashes shortly before and after sunset in China (*8*). Traffic crash rates shortly after sunset were found to be higher than at any other time, whereas the rate shortly before sunset was found to be lower. The results also show that pedestrian crash counts shortly after sunset were higher than at any other time. To evaluate the degrees of sun glare effects, Li et al. proposed the use of the publicly accessible Google Street View (GSV) panorama images (*9*). They used a deep learning tool—convolutional neural network algorithm—on the segmented GSV images to predict obstruction-related safety concerns attributable to sun glare. Additionally, this study predicted the time windows of sun glare by calculating the sun positions and the relative angles between drivers and the sun for the locations in Cambridge, Massachusetts. The results showed precise prediction accuracies.

The literature review shows that there is a need for an in-depth investigation to mitigate the current research gaps in identifying the patterns of the associations between key contributing factors by acquiring a comprehensive sun glare-related traffic crash data.

Methodology

Data Integration

This study obtained 7 years (2010–2016) of crash data from the Louisiana Department of Transportation and Development. The dataset has three major subfiles: (1) crash file, (2) vehicle file, and (3) roadway inventory file. The roadway inventory file contains information about crash location, roadway type, traffic volume, segment length, and other relevant geometric information. As "glare" or "sun glare" are not coded as key factors in the structured crash dataset, this study used "crash narrative" data to identify sun glare-related crashes. A set of keywords (glare, glaring, sunglare, sunglaring, sun glare, sun glaring) was used to identify the crash narratives associated with glare-related crashes. A manual effort was performed to identify sun glare-related crashes by using time of the day of crash occurrence and manual reading of the crash narrative reports. The final dataset contains 1,450 sun glare-related crashes in Louisiana.

Exploratory Data Analysis

The selection of variables is an important step in conducting crash data analysis. As the crash dataset contains a wide range of different types of variables (e.g., numerical, integer, nominal, categorical, or ordinal), it is important to determine the variables that can provide intuitive knowledge about the crash occurrence. The preliminary data-collection process includes crash (e.g., day of week), roadway (e.g., presence of intersection), driver (e.g., driver age), and vehicle (e.g., vehicle type) variables in the analysis. It was noticed that there are significant percentages of missing information in driver- and vehicle-level data. The study is designed for the exploration of crash and roadway-related variables only. This study used both literature review findings and variable importance method, using a random forest algorithm, to identify the key variables associated with crashes. The variables with high information value were later used for the modeling. Table 1 lists the distribution of the categories of the selected variables by crash severity type. It is interesting to see that none of the sun glare crashes are associated with fatal crashes. Severe crashes are also comparatively lower than other severity types. Complaint crashes show a slightly higher percentage during fall and spring seasons. Saturday shows a higher percentage of severe crashes than the other days of the week. Cloudy weather is associated with more severe and moderate crashes than other weather conditions. The severity of intersection-related crashes shows around

Table 1. Distributing of Key Categories by Crash Severity

Variable	Category	Severe (A)	Moderate (B)	Complain (C)	No injury (O)
Season	Fall	0.5	7.2	24.9	67.4
	Spring	0.3	8.2	19.8	71.7
	Summer	0.0	5.3	22.5	72.2
	Winter	0.4	6.4	21.5	71.6
Day of week	Friday (FR)	0.0	8.5	23.4	68.2
	Saturday (SA)	2.8	9.4	22.4	65.4
	Sunday (SU)	0.0	11.4	21.4	67.1
	Monday (MO)	0.0	6.1	25.0	69.0
	Tuesday (TU)	0.8	5.5	19.8	73.8
	Wednesday (WE)	0.4	5.5	24.1	70.0
	Thursday (TH)	0.0	6.9	21.1	72.0
Weather	Clear	0.4	7.2	22.9	69.6
	Cloudy	1.7	3.5	19.0	75.9
	Rain	0.0	0.0	0.0	100.0
	Fog/smoke	0.0	0.0	0.0	100.0
	Other	0.0	0.0	0.0	100.0
Intersection	No	0.4	6.5	21.0	72.1
	Yes	0.5	7.6	24.4	67.6
Access control	Full control	1.5	6.1	18.2	74.2
	No control	0.3	7.1	23.1	69.5
	Partial control	1.2	5.8	17.2	75.9
Highway	Interstate	1.3	6.3	15.2	77.2
	U.S. highway	0.5	6.6	23.9	69.0
	State highway	0.6	6.9	23.6	69.0
	City street	0.2	7.4	23.2	69.2
	Parish road	0.3	6.7	21.3	71.7
	Toll road	0.0	0.0	50.0	50.0
	Not reported	0.0	11.1	22.2	66.7
Land use	Business continuous (cont.)	0.0	6.8	24.4	68.7
	Industrial	0.0	5.7	11.4	82.9
	Residential	0.6	6.0	18.7	74.6
	Residential scattered (scatt.)	1.4	6.9	27.4	64.4
	Mixed	0.2	7.6	22.0	70.2
	Open country	2.7	9.5	23.0	64.9
	School/playground	0.0	5.0	50.0	45.0
Roadway facility	2-way no seperation (sep.)	0.4	7.0	23.3	69.3
	2-way with barrier (barr.)	0.0	13.0	15.2	71.7
	2-way with sep.	0.6	6.2	21.4	71.7
	One-way road	0.0	5.0	22.8	72.3
	Other	0.0	14.3	21.4	64.3
Alignment	Curve-level	0.0	7.1	14.3	78.6
	Curve-level-elevation (elev.)	0.0	0.0	30.0	70.0
	Dip, hump-straight	0.0	0.0	0.0	100.0
	Hillcrest-curve	0.0	0.0	0.0	100.0
	Hillcrest-straight	0.0	7.7	30.8	61.5
	On grade-curve	0.0	16.7	0.0	83.3
	On grade-straight	0.0	20.0	30.0	50.0
	Straight-level	0.5	6.8	22.7	70.1
	Straight-level-elev.	0.0	12.1	30.3	57.6

33% of crashes resulting in injuries. For segment or non-intersection-related crashes, this percentage is 27%. Access control and highway types do not show significant differences among the percentage distributions. Open country locations are associated with more severe crashes than the other locality types. On-grade crashes are disproportionately higher in percentages than other alignment types.

Cluster Correspondence Analysis

Correspondence analysis, an unsupervised machine learning method, can explore two-way and multi-way tables that contain association between the rows and columns from datasets with a wide range of categorical variables. Many recent transportation engineering studies (*10–25*) have applied dimension-reduction methods to

Figure 1. Coordinates of the variable categories and key clusters.

solve engineering problems. Cluster correspondence analysis, a variant of the correspondence analysis framework, utilizes both dimension reduction and cluster analysis for nominal data. This approach concurrently allocates individuals to clusters and optimal scaling measures to the variable categories. To determine the nature of the underlying cluster structures, this method outperforms other correspondence analysis techniques. A brief overview of cluster correspondence analysis is described here, which is mostly based on the work conducted by Velden et al. (*26*).

Initially, the data can be associated with n entities (e.g., drivers involved in sun glare-related crashes) for p categorical variables (for example, roadway alignment). One can express it by a super indicator matrix $\mathbf{Z}$ with $n \times Q$ dimension, where $Q = \sum_{j=1}^{p} q_j$. By using an indicator matrix $\boldsymbol{Z}_K$, the user can develop a tabular format to cross-tabulate cluster memberships with the nominal or categorical variables such as $\boldsymbol{F} = \boldsymbol{Z}'_K \boldsymbol{Z}$, where $\boldsymbol{Z}_K$ is the $n \times K$ indicator matrix indicating cluster membership. Application of the correspondence analysis framework to this matrix populates optimal scaling values for rows (as clusters) and columns (as categories). The clusters are optimally separated in relation to the distributions over the categorical variables in the two-dimensional plane. Similarly, the categories differing distributions over the clusters can be expressed as:

$$max\emptyset_{clusca}(\boldsymbol{Z}_K, \boldsymbol{B}^*) = \frac{1}{p} trace \boldsymbol{B}^{*'} \boldsymbol{D}_z^{-1/2} \boldsymbol{Z}' \boldsymbol{M} \boldsymbol{Z}_K \boldsymbol{D}_Z^{-1} \boldsymbol{Z}'_K \boldsymbol{M} \boldsymbol{Z} \boldsymbol{D}_z^{-1/2} \boldsymbol{B}^* \quad (1)$$

where

$\boldsymbol{M} = \boldsymbol{I}_n - \mathbf{1}_n \mathbf{1}'_n / n$

$\boldsymbol{B} = \sqrt{np} \boldsymbol{D}_z^{-\frac{1}{2}} \boldsymbol{B}^*$

$\boldsymbol{D}_K = \boldsymbol{Z}'_K \boldsymbol{Z}_k$, a diagonal matrix with cluster sizes

$\boldsymbol{D}_z$ is a diagonal matrix so that $\boldsymbol{D}_z 1_Q = \boldsymbol{Z}' 1_n$

An open-source software R package *clustrd* (*27*) was used in performing the analysis. Based on the preliminary results, the Calinski–Harabasz measure has been found as the suitable measure. This measure, also known as the valence ratio criterion, is the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. The performance can be expressed by this measure. This measure is used for the application of k-means clustering to complete clustering for different k measures.

Results

Cluster Analysis Results

This study conducted the k-means runs randomly multiple times to gain an optimal number of clustering. After conducting the optimization technique, the final cluster has been fixed at six. The objective criterion value of the analysis is 3.9608. Figure 1 shows a two-dimensional visualization known as biplot. This plot generalizes a simplistic preview of two-variable scatterplots with the text labels of the variable categories. One can project individual subject points into this biplot illustration and identify the variability within and between clusters. In the framework of correspondence analysis, the origin indicates the mean profile, and all other coordinates depict variations from this mean profile. Four clusters (cluster

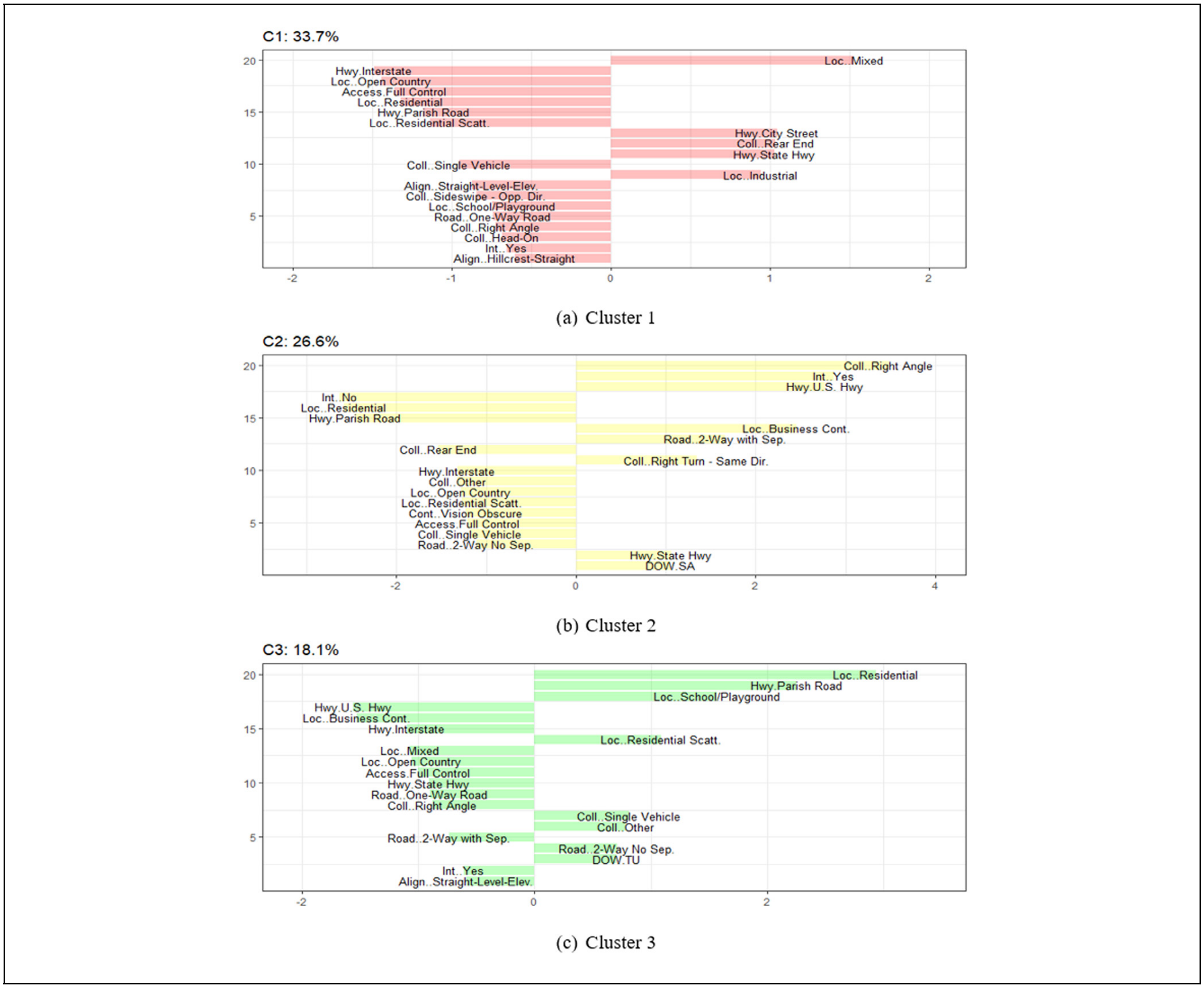


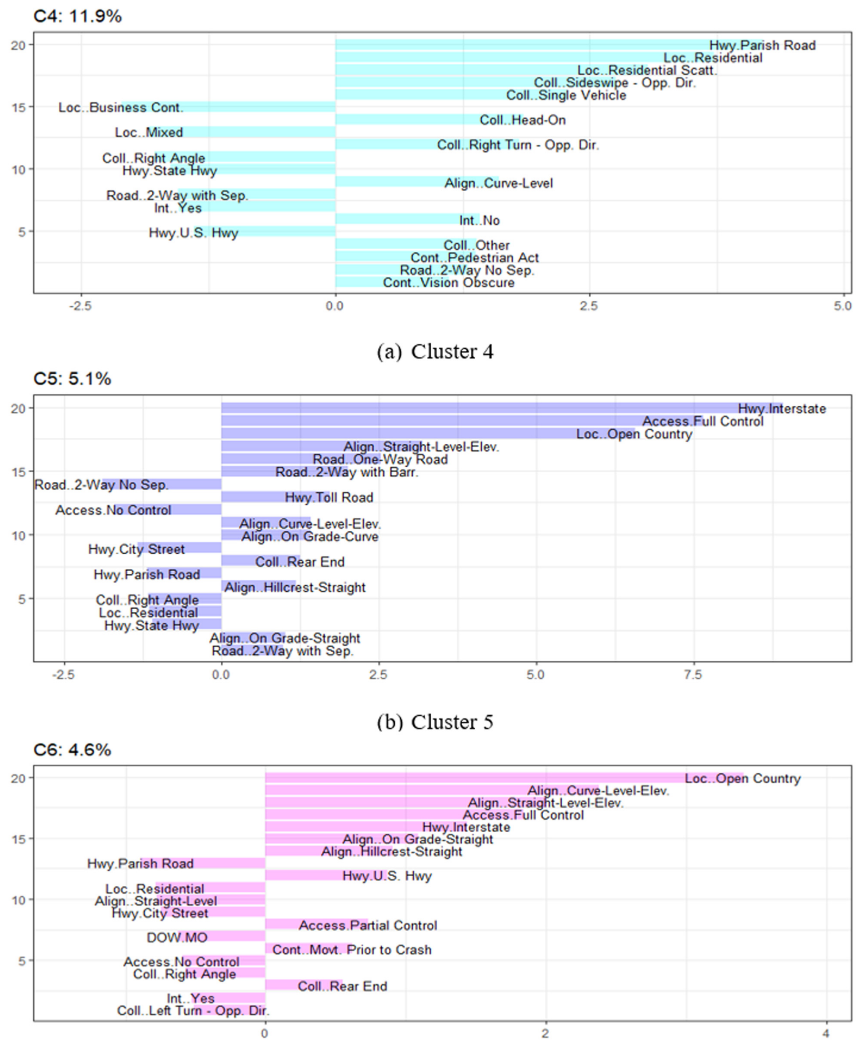
Figure 2. Top 20 largest standardized residuals per cluster (cluster 1–3): (a) standardized residuals in cluster 1; (b) standardized residuals in cluster 2; (c) standardized residuals in cluster 3.

1, 2, 3, and 4) are on the right side of the y-axis. The first two clusters are in quadrant 1, and the other two are in quadrant 4. The categories are closely placed for all of these four clusters. The centroid of cluster 6 is located in quadrant 2. The centroid of cluster 5 is far from all of the remaining clusters.

Figures 2 and 3 distinctly confirm the graphical illustration shown in Figure 1. It is important to establish the number of optimum clusters. Figure 3 displays both the centroids of the clusters (in blue ellipses) and variable categories. Table 2 shows the key measures associated with the cluster centroids, including the sample size (i.e., number and percentage of crashes within each cluster) and coordinates. The first four clusters contain information for 90% of the data. The least information is associated with cluster 6 (only 4.6% of data).

The six plots in Figures 2 and 3 display the 20 variable categories for each cluster with the highest standardized residuals (positive or negative). A positive (negative) residual indicates the category has an above (below) average frequency within the cluster. To interpret the results, the categories' positive residual means are usually explained. An advantage of this approach is the ability to generate "in cluster" proportions of the categories.

Cluster 1. This cluster has five categories with positive residual means: mixed location, industrial area, city street, state highway, and rear-end collision (Figure 2*a*). This indicates an association of sun glare-related rear-end collisions on city streets or state highways in industrial areas. The findings are in line with Mitra (*7*).



(a) Cluster 4

(b) Cluster 5

(c) Cluster 6

Figure 3. Top 20 largest standardized residuals per cluster (cluster 4–6): (a) standardized residuals in cluster 4; (b) standardized residuals in cluster 5; (c) standardized residuals in cluster 6.

Table 2. Location of the Cluster Centroids

Cluster	Sample size (%)	Sum of squares	Dimension 1	Dimension 2
Cluster 1	489 (33.7)	0.0226	0.0056	0.0041
Cluster 2	386 (26.6)	0.0230	0.0062	0.0215
Cluster 3	263 (18.1)	0.0145	0.0078	−0.0148
Cluster 4	172 (11.9)	0.0199	0.0077	−0.0367
Cluster 5	74 (5.1)	0.0336	−0.0878	−0.0030
Cluster 6	66 (4.6)	0.0169	−0.0302	0.0018

Cluster 2. This cluster has eight categories with positive residual means: right-angle collision, intersection, U.S. highway, business area, 2-way road with separation, collision while turning right in the same direction, state highway, and Saturday (Figure 2*b*). This indicates that sun glare-related right-angle collisions are associated

with intersections on 2-way roads with separation on either state or U.S. highways. These collisions were also associated with business areas on Saturdays.

Cluster 3. This cluster has eight categories with positive residual means: residential area, parish road, located near a school or playground, residential scattered area, 2-way road with no separation, Tuesday, single-vehicle collision, and other collision type (Figure 2*c*). This cluster indicates an association of single-vehicle collisions on 2-way parish roads with no separation and residential or school areas. The results are in agreement with other previous studies (*6–8*).

Cluster 4. This cluster has 12 categories with positive residual means: parish road, residential area, sideswipe collision from the opposite direction, single-vehicle collision, head-on collision, collision while turning right in the opposite direction, curve-level alignment, no intersection, pedestrian action as a contributing factor, obscured vision as a contributing factor, 2-way road with no separation, and other collision types (Figure 3*a*). This cluster indicated an association among many variables. Single-vehicle collisions, head-on collisions, sideswipe collisions from the opposite direction, and collisions turning right from the opposite direction were associated with curve-level alignment and no intersection. They were also associated with both pedestrian action and impaired vision as a contributing factor. Other studies also found similar results (*6*, *8*).

Cluster 5. This cluster has 13 categories with positive residual means: interstate, full control, open country area, straight-level elevated alignment, one-way road, 2-way road with a barrier, toll road, curve-level elevated alignment, alignment on grade curve, rear-end collision, hillcrest straight alignment, alignment on grade straight, and 2-way road with separation (Figure 3*b*). This indicates an association between sun glare-related rear-end collisions on interstate or toll road one-way roads or 2-way roads with a barrier or separation and alignment on grade curve or grade straight, hillcrest straight alignment, or curve-level elevated alignment. It is also associated with the open country area and full access control.

Cluster 6. This cluster has 11 categories with positive residual means: open country area, curve-level elevated alignment, straight-level elevated alignment, full control, interstate, alignment on grade straight, hillcrest straight alignment, U.S. highway, partial control, movement before crash as a contributing factor, and rear-end collision (Figure 3*c*). This indicates an association between rear-end collisions in open country areas on the interstate or U.S. highway and curve-level elevated alignment, straight-level elevated alignment, alignment on grade straight, or hillcrest straight alignment. It is also associated with either partial or full access control and movement before the crash.

Discussion

The color heatmap format of Tables 3 and 4 provides a quick glance of the percentage distribution of the categories by clusters and proportion odds of the categories by clusters. The darkest shading indicates the cluster with the most crashes for that category. Tables 3 and 4, along with Figures 1 to 3, provide techniques for identifying the primary categories within a cluster and their associated implications.

Table 3 provides a brief overview of the distribution of the variable categories in each cluster. Instead of normalizing the categories by variable group, this approach effectively represents which category is dominant in each cluster type. The highest number in each row is displayed with a darker red color. Table 4 lists the proportion odds of the categories when compared with the original dataset of sun glare-related crashes.

The key findings from these tables are discussed below:

- Cluster 1 represents the sun glare-related crashes that occur on business/industrial and mixed localities. The odds values for the majority of the categories are not drastically higher. The other categories with slightly higher odds are segment-related crashes, state highways, city streets, cloudy/foggy as weather conditions, and two-way roadways with no separations.
- Cluster 2 shows higher odds for rain-related crashes. The odds ratios are not drastically higher in this cluster. The odds of intersection and U.S. highways are also higher in this cluster.
- Cluster 3 and cluster 4 show higher odds for residential locality-related crashes. Other categories with higher odds for these two clusters are fog/smoke and rain as weather conditions, and curve, hillcrest, and on grade alignments.
- Cluster 5 is associated with high proportions of crashes on full access control roadways. Around 78% of crashes in this cluster occurred on full access control roadways. For the original data, this percentage is 54%. The proportion odds show a high value of 15.45. The other categories with higher odds value for this cluster are interstate roadways, toll roadways, and two-way roadways with a barrier. The properties of other categories indicate that this cluster is mostly associated with interstate-related sun glare-related crashes. It is

Table 3. Proportion of Categories in Different Clusters

Variable	Category	All data	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Season	Fall	38.28	42.3	30.3	36.9	44.8	39.2	42.4
Season	Spring	20.21	19	14.2	24.3	31.4	23	15.2
Season	Summer	10.41	10.2	14	6.5	4.7	17.6	13.6
Season	Winter	31.1	28.4	41.5	32.3	19.2	20.3	28.8
DOW	FR	17.1	18.8	16.3	13.3	21.5	8.1	22.7
DOW	MO	17.1	17.8	19.2	20.2	10.5	16.2	6.1
DOW	SA	7.38	4.9	12.4	4.6	6.4	10.8	6.1
DOW	SU	4.83	3.7	5.2	3.8	8.1	2.7	9.1
DOW	TH	20.9	20.7	24.6	19.8	14	24.3	19.7
DOW	TU	16.34	15.5	9.1	22.4	24.4	23	12.1
DOW	WE	16.34	18.6	13.2	16	15.1	14.9	24.2
Int.	No	54.41	61.3	20.2	63.1	82	78.4	69.7
Int.	Yes	45.59	38.7	79.8	36.9	18	21.6	30.3
Access	Full control	4.55	0	0	0	0	70.3	21.2
Access	No control	89.45	94.7	92.5	95.4	95.9	24.3	65.2
Access	Partial control	6	5.3	7.5	4.6	4.1	5.4	13.6
Hwy	City street	31.52	40.7	30.1	37.3	19.8	1.4	13.6
Hwy	Interstate	5.45	0	0	0	0	89.2	19.7
Hwy	Parish road	22.69	13.9	2.1	45.6	75.6	0	4.5
Hwy	State hwy	24.9	32.9	34.2	15.2	3.5	2.7	30.3
Hwy	Toll road	0.14	0	0	0	0	2.7	0
Hwy	U.S. hwy	14.69	11.5	33.4	1.9	0.6	4.1	28.8
Loc.	Business cont.	33.31	37.8	58.3	14.4	1.2	23	24.2
Loc.	Industrial	2.41	4.7	1.6	0.8	0	2.7	3
Loc.	Mixed	31.03	44.4	39.6	18.3	4.1	9.5	27.3
Loc.	Open country	5.1	0	0	0	0.6	64.9	37.9
Loc.	Residential	21.72	12.1	0.5	51	67.4	0	6.1
Loc.	Residential Scatt.	5.03	1	0	10.3	23.3	0	1.5
Loc.	School/playground	1.38	0	0	5.3	3.5	0	0
Seve.	Complaint	22.55	22.5	27.7	19.8	16.9	14.9	27.3
Seve.	Moderate	6.97	7.2	6.2	7.6	7.6	8.1	4.5
Seve.	No injury	70.07	70.1	65.8	72.2	75	75.7	66.7
Seve.	Severe	0.41	0.2	0.3	0.4	0.6	1.4	1.5
Wea.	Clear	95.31	95.3	94.8	97	95.9	93.2	92.4
Wea.	Cloudy	4	4.5	4.1	2.3	2.9	5.4	7.6
Wea.	Fog/smoke	0.14	0.2	0	0.4	0	0	0
Wea.	Other	0.21	0	0	0	1.2	1.4	0
Wea.	Rain	0.34	0	1	0.4	0	0	0
Road.	2-way no Sep.	66.69	74	50.8	79.1	94.8	4.1	53
Road.	2-way with barr.	3.17	2.7	1.3	2.7	2.3	17.6	6.1
Road.	2-way with Sep.	22.21	18.4	36	14.8	2.9	40.5	28.8
Road.	One-way road	6.97	3.9	11.4	1.9	0	33.8	12.1
Road.	Other	0.97	1	0.5	1.5	0	4.1	0
Align.	Curve-level	3.86	3.1	1	3.4	12.2	4.1	6.1
Align.	Curve-level-elev.	0.69	0	0	0	0	5.4	9.1
Align.	Dip, hump-straight	0.14	0	0.3	0.4	0	0	0
Align.	Hillcrest-curve	0.07	0	0	0	0.6	0	0
Align.	Hillcrest-straight	0.9	0	0.5	0.4	1.7	5.4	4.5
Align.	On grade-curve	0.41	0	0	0	1.2	4.1	1.5
Align.	On grade-straight	0.69	0.4	0	0.8	0	4.1	4.5
Align.	Straight-level	90.97	96.3	98.2	94.7	81.4	55.4	59.1

important to note that only 5% of sun glare-related crashes are associated with this cluster.

- Cluster 6 properties are also similar to cluster 5. However, this cluster does not include toll road crashes. The odds of curve-level elevated are disproportionately higher in this cluster.

Conclusions

This study applied a relatively new categorical data analysis method that combines cluster analysis and correspondence analysis to determine the key clusters of crash-contributing factors in sun glare-related crashes. In

Table 4. Odds Ratio of the Variable Categories in Different Clusters

Variable	Category	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Season	Fall	1.11	0.79	0.96	1.17	1.02	1.11
Season	Spring	0.94	0.7	1.2	1.55	1.14	0.75
Season	Summer	0.98	1.34	0.62	0.45	1.69	1.31
Season	Winter	0.91	1.33	1.04	0.62	0.65	0.93
DOW	FR	1.1	0.95	0.78	1.26	0.47	1.33
DOW	MO	1.04	1.12	1.18	0.61	0.95	0.36
DOW	SA	0.66	1.68	0.62	0.87	1.46	0.83
DOW	SU	0.77	1.08	0.79	1.68	0.56	1.88
DOW	TH	0.99	1.18	0.95	0.67	1.16	0.94
DOW	TU	0.95	0.56	1.37	1.49	1.41	0.74
DOW	WE	1.14	0.81	0.98	0.92	0.91	1.48
Int.	No	1.13	0.37	1.16	1.51	1.44	1.28
Int.	Yes	0.85	1.75	0.81	0.39	0.47	0.66
Access	Full control	0	0	0	0	15.45	4.66
Access	No control	1.06	1.03	1.07	1.07	0.27	0.73
Access	Partial control	0.88	1.25	0.77	0.68	0.9	2.27
Hwy	City street	1.29	0.95	1.18	0.63	0.04	0.43
Hwy	Interstate	0	0	0	0	16.37	3.61
Hwy	Parish road	0.61	0.09	2.01	3.33	0	0.2
Hwy	State hwy	1.32	1.37	0.61	0.14	0.11	1.22
Hwy	Toll road	0	0	0	0	19.29	0
Hwy	U.S. hwy	0.78	2.27	0.13	0.04	0.28	1.96
Loc.	Business cont.	1.13	1.75	0.43	0.04	0.69	0.73
Loc.	Industrial	1.95	0.66	0.33	0	1.12	1.24
Loc.	Mixed	1.43	1.28	0.59	0.13	0.31	0.88
Loc.	Open country	0	0	0	0.12	12.73	7.43
Loc.	Residential	0.56	0.02	2.35	3.1	0	0.28
Loc.	Residential scatt.	0.2	0	2.05	4.63	0	0.3
Loc.	School/playground	0	0	3.84	2.54	0	0
Seve.	Complaint	1	1.23	0.88	0.75	0.66	1.21
Seve.	Moderate	1.03	0.89	1.09	1.09	1.16	0.65
Seve.	No injury	1	0.94	1.03	1.07	1.08	0.95
Seve.	Severe	0.49	0.73	0.98	1.46	3.41	3.66
Wea.	Clear	1	0.99	1.02	1.01	0.98	0.97
Wea.	Cloudy	1.13	1.03	0.58	0.73	1.35	1.9
Wea.	Fog/smoke	1.43	0	2.86	0	0	0
Wea.	Other	0	0	0	5.71	6.67	0
Wea.	Rain	0	2.94	1.18	0	0	0
Road.	2-way no Sep.	1.11	0.76	1.19	1.42	0.06	0.79
Road.	2-way with barr.	0.85	0.41	0.85	0.73	5.55	1.92
Road.	2-way with Sep.	0.83	1.62	0.67	0.13	1.82	1.3
Road.	One-way road	0.56	1.64	0.27	0	4.85	1.74
Road.	Other	1.03	0.52	1.55	0	4.23	0
Align.	Curve-level	0.8	0.26	0.88	3.16	1.06	1.58
Align.	Curve-level-elev.	0	0	0	0	7.83	13.19
Align.	Dip, hump-straight	0	2.14	2.86	0	0	0
Align.	Hillcrest-curve	0	0	0	8.57	0	0
Align.	Hillcrest-straight	0	0.56	0.44	1.89	6	5
Align.	On grade-curve	0	0	0	2.93	10	3.66
Align.	On grade-straight	0.58	0	1.16	0	5.94	6.52
Align.	Straight-level	1.06	1.08	1.04	0.89	0.61	0.65
Align.	Straight-level-elev.	0.09	0	0.18	1.27	9.47	6.67

addition to a low-dimensional approximation depicting clusters and categories, there is a cluster partitioning of individuals based on the profiles over the categorical variables. Using 7 years (2010–2016) of crash data from Louisiana, this study empirically determined the relative contribution of key factors for different cluster groups. Given that a crash is typically the complex and interrelated result of human, vehicle, roadway, and environmental factors, this study contributes to the current safety literature by identifying high-risk scenarios where sun glare-related crashes are more likely to occur through interactions with other related factors. The research team developed six clusters with associated factors and provided odds ratio measures by the categories. It is anticipated that these findings will provide a deeper understanding of sun glare-related crashes. Most of the countermeasures related to sun glare-related crashes are based on safety precautions taken by the drivers. Drivers who regularly confront sun glare issues in their regular commutes should use anti-glare visors in the car. The use of sunglasses during long-distance travel during daytime could also be beneficial.

Although user-related precautions are dominant in reducing sun glare-related crashes, there is potential in the consideration of the safe system approach. Instead of the traditional perspective of "improving human behavior," agencies can consider safe and sustainable design to eliminate human errors. For example, road geometry redirection or consideration of sun glare impact on roadways while building new roads can be considered as alternatives. Advanced warning messages with the capability of providing wireless information about traffic control devices can alert drivers in case there is any problem from sun glare (*28*). In future, full self-driving mode in automated vehicles could also be instrumental in avoiding human error.

The current study is not without limitations. One limitation is that the data selection is primarily based on police-reported crash narrative documents. Missing information related to "sun glare" in the crash narrative will prohibit the inclusion of the crash to be considered as a sun glare-related crash. Additionally, this study is limited to crash and roadway-level variables only. Note that placement of the visor makes a lot of difference in reducing sun glare crashes. Thus, inclusion of vehicle type could provide information on the position of the vehicle visor. The current study did not include "vehicle type" because of disproportional missing value issues. Future studies should examine the inclusion of vehicle and driver-level variables to provide intuitive insights on overall sun glare-related crashes.

Author Contributions

The authors confirm the contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das; analysis and interpretation of results: Subasish Das; draft manuscript preparation: Subasish Das, Xiaoduan Sun, Bahar Dadashova, M. Ashifur Rahman, and Ming Sun. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

The authors like to thank three anonymous reviewers for their excellent suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Subasish Das https://orcid.org/0000-0002-1671-2753
Xiaoduan Sun https://orcid.org/0000-0001-7282-1340
Bahar Dadashova https://orcid.org/0000-0002-4592-9118
M. Ashifur Rahman https://orcid.org/0000-0001-6940-1599

References

1. Churchill, A. M., Y. Tripodis, and D. J. Lovell. Sun Glare Impacts on Freeway Congestion: Geometric Model and Empirical Analysis. *Journal of Transportation Engineering*, Vol. 138, No. 10, 2012, pp. 1196–1204.
2. Hagita, K., and K. Mori. The Effect of Sun Glare on Traffic Accidents in Chiba Prefecture, Japan. *Asian Transport Studies*, Vol. 3, No. 2, 2014, pp. 205–219.
3. Staver, J. *Sun Glare Accidents*. Staver Law Group PC. https://www.chicagolawyer.com/sun-glare-accidents/. Accessed July 6, 2017.
4. Schreuder, D. A. *Side Lights and Low-Beam Headlights in Built-up Areas*. Institute for Road Safety Research, Voorburg, The Netherlands, 1968.
5. McGwin, G., V. Chapman, and C. Owsley. Visual Risk Factors for Driving Difficulty among Older Drivers. *Accident Analysis & Prevention*, Vol. 32, No. 6, 2000, pp. 735–774.
6. Mitra, S. Investigating Impact of Sun Glare on Transportation Safety. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
7. Mitra, S. Sun Glare and Road Safety: An Empirical Investigation of Intersection Crashes. *Safety Science*, Vol. 70, 2014, pp. 246–254.
8. Hagita, K., and K. Mori. Traffic Accident Analysis of Sun Glare and Twilight Shortly Before and after Sunset in Chiba Prefecture, Japan. *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 11, 2015, pp. 1999–2014.
9. Li, X., B. Y. Cai, W. Qiu, J. Zhao, and C. Ratti. A Novel Method for Predicting and Mapping the Occurrence of Sun Glare Using Google Street View. *Transportation Research Part C: Emerging Technologies*, Vol. 106, 2019, pp. 132–144.
10. Das, S., and X. Sun. Association Knowledge for Fatal Run-off-Road Crashes by Multiple Correspondence Analysis. *IATSS Research*, Vol. 39, No. 2, 2016, pp. 146–155.
11. Das, S. Ridesharing Services and Car-Seats: Technological Perceptions and Usage Patterns, 2020. https://arxiv.org/ftp/arxiv/papers/2011/2011.02277.pdf

12. Das, S., and A. Dutta. Extremely Serious Crashes on Urban Roadway Networks: Patterns and Trends. *IATSS Research*, Vol. 44, No. 3, 2020, pp. 248–252.
13. Das, S., K. Jha, K. Fitzpatrick, M. Brewer, and T. H. Shimu. Pattern Identification from Older Bicyclist Fatal Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673(6): 638–649.
14. Jalayer, M., H. Zhou, and S. Das. Exploratory Analysis of Run-Off-Road Crash Patterns. In *Data Analytics for Smart Cities* (A. Alavi, and W. G. Buttlar, eds.) CRC Press, Boca Raton, FL, 2018.
15. Das, S. Identifying Key Patterns in Motorcycle Crashes: Findings from Taxicab Correspondence Analysis. *Transportmetrica A: Transport Science*, Vol. 17, No. 4, 2021, pp. 593–614.
16. Ali, F., D. Dissanayake, M. Bell, and M. Farrow. Investigating Car Users' Attitudes to Climate Change Using Multiple Correspondence Analysis. *Journal of Transport Geography*, Vol. 72, 2018, pp. 237–247.
17. Das, S., S. Ashraf, L. Tran, and A. Dutta. Pedestrians Under Influence (PUI) Crashes: Patterns from Correspondence Regression Analysis. *Journal of Safety Research*, Vol. 75, 2020, pp. 14–23.
18. Baireddy, R., H. Zhou, and M. Jalayer. Multiple Correspondence Analysis of Pedestrian Crashes in Rural Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672(38): 116–127.
19. Das, S., A. Dutta, and K. Fitzpatrick. Technological Perception on Autonomous Vehicles: Perspectives of the Non-Motorists. *Technology Analysis & Strategic Management*, Vol. 32, No. 11, 2020, pp. 1335–1352.
20. Wen, C.-H., and W.-Y. Chen. Using Multiple Correspondence Cluster Analysis to Map the Competitive Position of Airlines. *Journal of Air Transport Management*, Vol. 17, No. 5, 2011, pp. 302–304.
21. Das, S., L.N. Tran, and M. Theel. Understanding Patterns in Marijuana Impaired Traffic Crashes. *Journal of Substance Use*, Vol. 26, No. 1, 2020, pp. 21–29.
22. Das, S., X. Kong, S. Lavrenz, L. Wu, and M Jalayer. Fatal Crashes at Highway Rail Grade Crossings: A US based Study. *International Journal of Transportation Science and Technology*, 2021. https://doi.org/10.1016/j.ijtst.2021.03.002
23. Kong, X., S. Das, H. Zhou, and Y. Zhang. Characterizing Phone Usage While Driving: Safety Impact From Road and Operational Perspectives Using Factor Analysis. *Accident Analysis & Prevention*, Vol. 152, 2021, p. 106012.
24. Hosseini, P., M. Jalayer, and S. Das. A Multiple Correspondence Approach to Identify Contributing Factors Related to Work Zone Crashes. Presented at 100th Annual Meeting of the Transportation Research Board, Washington, D.C., 2021 (virtual).
25. Das, S., and A. Dutta. Light Delivery Vehicles Crashes: Identifying Insights using Joint Dimension Reduction and Clustering. Presented at 100th Annual Meeting of the Transportation Research Board, Washington, D.C., 2021 (virtual).
26. Van de Velden, M., A. I. D' Enza, and F. Palumbo. Cluster Correspondence Analysis. *Psychometrika*, Vol. 82, No. 1, 2017, pp. 158–185.
27. Allard, J. *clustrd: Methods for Joint Dimension Reduction and Clustering*. 2019. https://cran.r-project.org/web/packages/TaxicabCA/TaxicabCA.pdf. Accessed August 1, 2020.
28. Li, Q., F. Qiao, and L. Yu. *Eliminating Sun Glare Disturbance at Signalized Intersections by a Vehicle to Infrastructure Wireless Communication*. Report No. TranLIVE/TSU/16/04. U.S. Department of Transportation, Washington D.C., 2016.