Investigation of Contributing Factors to Traffic Crash Severity in Southeast Texas Using Multiple Correspondence Analysis

Guanlong Li[1†], Yueqing Li[1], Yalong Li[2†], Brian Craig[3] and Xing Wu[4]

[1] *Department of Industrial and Systems Engineering, Lamar University, Beaumont, Texas, U.S., 77710,* gli@lamar.edu, yli6@lamar.edu
[2] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, U.S., 30332,* yli3397@gatech.edu
[3] *College of Engineering, Lamar University, Beaumont, Texas, U.S., 77710,* brian.craig@lamar.edu
[4] *Department of Civil and Environmental Engineering, Lamar University, Beaumont, Texas, U.S., 77710,* xwu1@lamar.edu
† These authors contributed equally to this work.

Corresponding author: Dr. Yueqing Li; 4400 MLK Blvd., PO Box 10009, Beaumont, Texas 77710; yli6@lamar.edu; +1 (409) 880-8804

Key Findings

- Contributing factors like weather condition, lighting condition, crash time, speed limit, road class, surface condition and driving risk factors are significantly associated with crash severity.
- Multiple correspondence analysis (MCA) has been successfully applied to detect patterns and identify groups of contributing factors and combinatorial influence on the severity of traffic crashes.
- Driving in adverse climate conditions such as rain and extreme weather on the wet road surface is susceptible to traffic crashes with severe injuries or even fatality.
- Driving in the dark during non-rush time is more likely to cause serious traffic accidents even with the presence of street lights.
- Young male drivers are more prone to experience severe traffic crashes when driving used vehicles in high speed under the influence of drug use and driving mistakes.

Abstract

Driving is the essential means of travel in Southeast Texas, a highly urbanized and populous area that serves as an economic powerhouse of the whole state. However, driving in Southeast Texas is subject to many risks as this region features a typical humid subtropical climate with long hot summers and short mild winters. Local drivers would encounter intense precipitation, heavy fog, strong sunlight, standing water, slick road surface, and even frequent extreme weather such as tropical storms, hurricanes and flood during their year-around travels. Meanwhile, research has revealed that the fatality rate per 100 million vehicle miles driven in urban Texas became considerably higher than national average since 2010, and no conclusive study has elucidated the association between Southeast Texas crash severity and potential contributing factors. This study used multiple correspondence analysis (MCA) to examine a group of contributing factors on how their combinatorial influences determine crash severity by creating combination clouds on a factor map. Results revealed numerous significant combinatorial effects. For example, driving in rain and extreme weather on a wet road surface has a higher chance in causing crashes that incur severe or deadly injuries. Besides, other contributing factors involving risky behavioral factors, road designs, and vehicle factors were well discussed. The research outcomes could inspire local traffic administration to take more effective countermeasures to systematically mitigate road crash severity.

Keywords

Contributing factors, Crash severity, Southeast Texas, MCA, Categorical Variable, Combination Clouds.

Received: 13/08/2021; **Received in revised form:** 26/09/2021; **Accepted:** 27/09/2021;

Available on-Line: 11/08/2021 https://doi.org/10.33492/JRS-D-21-00051

Introduction

Traffic crash is widely considered as one of the leading causes of accidental human death around the world. World Health Organization (WHO) (WHO, 2018) estimates that more than 1.35 million people lose their lives every year as a result of traffic crash, and 20-50 million more victims suffer a variety of traffic crash-related injuries. Without effective countermeasures, traffic crash is anticipated to become the seventh leading cause of human death by 2030 (WHO, 2017). Within the year of 2018, 22,697 passenger vehicle occupants died, and an estimated 2.43 million people were injured from motor vehicle crashes on US roads (National Center for Statistics and Analysis, 2020). Apart from life losses and health damages, road crashes cost the U.S. $230.6 billion annually, or an average of $820 per person every year (Association of Safe International Road Travel, 2019).

Southeast Texas geographically covers Greater Houston, and Beaumont-Port Arthur metropolitan areas. The economy of Southeast Texas is composed primarily by industries relating to energy, petrochemicals, fishing, aerospace, agriculture, and tourism. Particularly, with a population of 7,066,141 residents by July 2019 and Gross Domestic Product (GDP) at 478.8 billion in 2018, Houston-The Woodlands-Sugar Land Metropolitan Statistical Area (MSA) makes it one of the largest and most economically vibrant metropolitans in the US (Greater Houston Partnership, 2020).

The general climate of Southeast Texas is subtropical, warm, moisture with heavy precipitation. Yearly, winds from the Gulf of Mexico mitigate the heat of summer and the cold of winter (Ning & Abdollahi, 2003). Southeast Texas averages more than 55 inches of rain annually and in some parts, rainfall may exceed 60 inches (Lyons, 1990). The two prominent precipitation peaks in Southeast Texas occur in between May and June and during September (Nielson-Gammon, 2011). Extreme weather such as tropical storms and hurricanes pass through the region periodically in summer and fall and would bring destructive gales, storm surges, tornados and floods on local communities (Beaman, 2019). Historically, disastrous hurricanes wreaked havoc on Southeast Texas, among which Galveston Hurricane in 1900 and Hurricane Harvey in 2017 are the most devastating.

Literature Review

Rainfall's effect on vehicle crash has been a global research focus for decades. Previous studies claimed a greater risk of road crash in the presence of rain (Andrey, Mills, Leahy, & Suggett, 2003; Qiu & Nixon, 2008). From a macroscopic view, a statistically significant linear trend exists between the number of traffic crashes and the amount of rainfall (Sherretz & Farhar, 1978). From a microscopic view examining the rainfall's impact on driving performance, people found that drivers in heavy rainfall were 3.8 times more likely to show a higher standard deviation of lane position than in clear weather (Ghasemzadeh & Ahmed, 2017). Further when investigating the crash severity, researchers found that heavy rain, deep water, and roads with a long drainage length are more likely to be associated with aggravated accident severity (Lee, Chae, Yoon, & Yang, 2018). Specifically, rain and warmer air temperatures were discovered to be linked to more serious crash injuries in single-vehicle truck crashes (Naik, Tung, Zhao, & Khattak, 2016). Nonetheless, recent study showed that wet weather, along with other factors such as male and young age tends to decrease driver injury seriousness (Li et al., 2019).

Fog is prevalent in humid regions, and it has a significant impact on driving behavior and the overall traffic safety. It was found that compared to clear day crashes, fog-related crashes tend to result in more severe injuries and involve more vehicles (Abdel-Aty, Ekram, Huang, & Choi, 2011). Foggy weather contributes to a higher odds of vehicle collisions as it deteriorates driver's vision range to less than 100 m (Tu, Li, Sun, & Dai, 2014). Poor visibility significantly increases driver's reaction time, cognitive and physiological demand, thereby impairing their skills to quickly respond to critical traffic events (Harb, Radwan, & Yan, 2007; Harb, Radwan, Yan, & Abdel-Aty, 2007). Moreover, even speed reductions are commonly implemented during fog condition by drivers, it was found to be insufficient to compensate for the crash risk (Brooks et al., 2011; Mueller & Trick, 2012).

Lighting condition invokes a controversy within academia as to how it affects driving safety. On one side, night driving is subject to many risks such as impaired vision, fatigue and inattention (Keall, Frith, & Patterson, 2004; Clarke, Ward, Bartle, & Truman, 2006). A Hong Kong study found that speeding was more likely to happen at night without road lighting, and driving in daylight featured the lowest likelihood of severe crash (Zhang, Yau, & Chen, 2013). Besides, poor illumination aggravates crash damage as some scholars found the ratio of fatal crashes per 100 collisions spiked on roads without street lighting (Plainis, Murray, & Palllikaris, 2006). On the other hand, a study from Mexico did find that drivers face greater risk of highway traffic crash in daytime compared to in nighttime. (Hijar, Carrillo, Flores, Anaya, & Lopez, 2000).

Human factor is another area of interest in traffic safety research. A great amount of research has investigated the association between crash risk and gender. Generally, male drivers have been found more likely to experience traffic crashes than female drivers (Holubowycz & Kloeden, 1994; Hayakawa, Fischloff, & Fischbeck, 2000). Besides, numerous studies have examined the effect of age on traffic crash rate and seriousness, and found that novice (young) drivers are at greater risk of traffic crash (Massie, Campbell, & Williams, 1995; Hijar et al., 2000).

As for road design, research has demonstrated that geometric design of road, traffic sign design and position have significant impacts on drivers' behavior under both normal and emergency conditions (Jamson, Tate, & Jamson, 2005; Wang & Song, 2011; Hang, Yan, Ma, Duan, & Zhang, 2018). It was also discovered that vehicle characteristics such as size, weight, and safety devices impose great impact on the consequence of traffic crashes (Evans & Frick, 1992; Huang, Siddiqui, & Abdel-Aty, 2011).

Research Goal

According to National Highway Traffic Safety Administration (NHTSA) national statistics, traffic fatality rate per 100 million vehicle miles traveled (VMT) in Texas has a higher rate than national average since 2010 (NHTSA, 2021). In 2017 alone, 3,726 people perished, and 17,538 people sustained a serious injury from motor vehicle traffic crashes on Texas roads (Texas Department of Transportation, 2018). Although the Crash Records Information System (CRIS) created by Texas Department of Transportation (TxDOT) stores exhaustive records of crashes on Texas roads since 2010, little research has utilized such database to conduct systematic traffic safety research. This study aims to identify contributing factors to the severity of traffic crashes in Southeast Texas by analyzing CRIS data. The implications of this study would assist traffic administrators in understanding the combinatorial effects of contributing factors on crash severity, thus inspire them to propose effective approaches to mitigate these risks facing local drivers.

Methods

Data Treatment

First, traffic crash data from 2010-2017 in Southeast Texas was retrieved from the TxDOT's CRIS by selecting areas: Houston-Galveston Area Council (HGAC) + Southeast Texas Regional Planning Commission (SETRPC). Then, data cleansing was conducted to remove those records with invalid or insufficient description on contributing factors. As a result, 46,063 records of crashes were retained for further analysis.

Contributing Factors

This study attempts to investigate twelve contributing factors to crash severity from four aspects including 1. Environment factor: weather condition, lighting condition, crash time, day of week and surface condition; 2. Road design factor: speed limit of road, road class; 3. Human factor: driver age, driver gender, risk factor; 4. Vehicle factor: vehicle age and vehicle body style, as shown in Table 1.

Table 1 Contributing factors

Environment factor	Weather condition, Lighting condition, Crash time, Day of week, Surface Condition
Road design factor	Speed limit of road, Road class
Human factor	Driver age, Driver gender, Risk factor
Vehicle factor	Vehicle age, Vehicle body style

All the variables within the contributing factors are transformed into categorical variables for the sake of following statistical analysis.

- Weather conditions are divided into five categories: Clear, Rain, Fog, Cloudy, and Extreme. Severe crosswinds, snow, sleet/hail, blowing sand are included within extreme weather group due to their infrequent occurrence in Southeast Texas.
- Light conditions are divided into five categories: Daylight, Dawn, Dusk, Dark Lighted, Dark Not Lighted.
- Crash time: twenty-four hours are sorted into four periods: Morning rush, Afternoon rush, Non rush daytime, Non rush nighttime.
- Day of week: days from Monday to Friday are grouped as Weekday while Saturday and Sunday are combined as Weekend.
- Surface conditions are categorized into two groups: Dry and Wet.
- Speed limits fall into three groups: Low speed limit (0-30 miles/hour), Medium speed limit (30-50 miles/hour), and High speed limit (50-80 miles/hour).
- Road Class is divided into three categories: Farm to Market, US & State Highways, Interstate.
- Driver age has four categories: 16-30, 31-45, 46-60, 60+.
- Driver gender is divided as Male and Female.
- Risk factors include fourteen groups, according to CRIS, Texas DoT. They are: 1. Cellphone Use 2. Distraction 3. Driver Inattention 4. Driving Errors 5. Drug Driving 6. Drunk Driving 7. Emergency 8. Failure in driving 9. Fatigue 10. Inability 11. Invalid Driver 12. Risky Driving 13. Taking Medication 14. Unsafe vehicle Condition.
- Vehicle body style contains five categories: Big Vehicle, Motorcycle, Pickup, Sedan and SUV.
- Vehicle age is divided as New, Used, Old.

Definitions of the relevant terms are provided in an appendix at the end of this research article.

Chi-squared Test

Firstly, an approach of nonparametric statistical analysis called chi-squared test for independence is performed to examine whether there exists statistically significant association between multiple contributing factors and the seriousness of crash injury in 3 levels: minor injury, severe injury and fatality. For each contributing factor, the Pearson statistic is calculated by summing up the variabilities between the actual observed frequency (O) in different crash severities and expected frequency (E) corresponding to that type of severity at a given categorical level of contributing factor, shown as Equation (1):

$$\mathcal{X}^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

The summary of chi-squared test results with significance level set to 5% is presented in Table 2.

Table 2. Chi-squared test examining associations between contributing factors and crash severity

Contributing Factor	Crash Severity		
	*	DOF	P-value
Crash time	950.25	6	< 0.001
Day of week	209.09	2	< 0.001
Light Condition	927.11	8	< 0.001
Road Class	73.559	4	< 0.001
Speed Limit	278.13	4	< 0.001
Surface Condition	38.143	2	< 0.001
Weather Condition	61.032	8	< 0.001
Vehicle Body Style	924.35	8	< 0.001
Vehicle Age	39.765	4	< 0.001
Driver Age	11.236	6	0.08136
Driver Gender	246.23	2	< 0.001
Risk Factor	961.79	26	< 0.001

Note: * Significant at 5% level.

Multiple Correspondence Analysis (MCA)

In addition to the chi-squared test for examining the associations between each contributing factor and the levels of crash severity from a quantitative perspective, we propose the use of multiple correspondence analysis (MCA) as a type of geometric data analysis technique (Le Roux & Rouanet, 2004) to implement unsupervised (machine) learning for clustering and identifying traffic contributing factors with similar frequency of coincidence from the large, complex multivariate dataset of crash records. Graphical illustrations of these clusters in the form of "combination clouds" (Das & Sun, 2015). are developed on the dimensionality-reduced factor map, allowing us to recognize the distribution pattern of variable groupings on a lucid 2-dimensional space and study the combinatorial effects of clustered variable categories on the crash severity. Other advantage from utilizing MCA approach involves that it does not require any pre-assumption of underlying relationships between responses and predictor variables before the analysis of data (Das & Sun, 2016).

In this research, the application of MCA primarily focuses on the clustering and identification of significant contributing factors responsible for traffic crashes where drivers get severely injuried or even killed. Therefore, only traffic crash records with crash severity specified as "Severe Injury" and "Killed" are considered in the MCA study, consequently the total number of selected crash records reduces to 11,650. Table 3 enlists a summary of all the categorical variables in contributing factors that participate in the MCA with corresponding levels of category, frequency and percentage specified, the statistical significance of involved factors has been examined in the Chi-squared tests. To implement the MCA computation, an open source statistical software R Version 4.0.2 is used with the aid of FactoMineR package for data analysis and factoextra package for data visualization (Husson & Pagès, 2011).

Historically developed in Benzécri's treatise on data analysis in 1973 (Benzécri, 1973; Beaudouin, 2016), multiple correspondence analysis is regarded as an extension of the simple correspondence analysis (CA) which allows the user to analyze the pattern of relationships between multiple dependent nominal variables with a large amount of data. It can be also seen as analogous to principal component analysis (PCA) where the variables to be analyzed are categorical instead of quantitative (Abdi & Valentin, 2007; Abdi & Williams, 2010). Similar as PCA, MCA also employs a dimension-reducing technique to extract the most important information from a given data set and produces a low-dimensional representation of the data while containing maximum variation (Abdi & Valentin, 2007; James, Witten, Hastie, & Tibshirani, 2013). For years, Bourdieu (Lebaron, 2009; Duval, 2018) has contributed significantly to the popularization of CA and MCA

Table 3. Summary of contributing factors

Categorical Variables	Category Level	Frequency	Percentage (%)
Driver Age	16-30	5514	47.33
	31-45	3117	26.755
	46-60	1998	17.15
	60+	1021	8.764
Driver Gender	Female	4007	34.395
	Male	7643	65.605
Vehicle Age	New	6476	55.588
	Used	4811	41.296
	Old	363	3.116
Crash Time	Afternoon Rush	1831	15.717
	Morning Rush	1285	11.03
	Non Rush Daytime	4101	35.202
	Non Rush Nighttime	4433	38.052
Day of Week	Weekday	7549	64.798
	Weekend	4101	35.202
Light Condition	Dark, Lighted	3043	26.12
	Dark, Not Lighted	2137	18.343
	Dawn	174	1.494
	Daylight	6130	52.618
	Dusk	166	1.425
Weather Condition	Clear	8513	73.073
	Cloudy	2080	17.854
	Extreme	9	0.077
	Fog	127	1.09
	Rain	921	7.906
Road Class	Farm to Market	3435	29.485
	Interstate	3016	25.888
	US & State Highways	5199	44.627
Speed Limit	High Speed Limit	6283	53.931
	Medium Speed Limit	13	0.112
	Low Speed Limit	5354	45.957
Surface Condition	Dry	10266	88.12
	Wet	1384	11.88
Vehicle Body Style	Big Vehicle	946	8.12
	Motorcycle	997	8.558
	Pickup	2696	23.142
	Sedan	4855	41.674
	SUV	2156	18.506
Risk Factor	Cellphone Use	21	0.18
	Distraction	62	0.532
	Driver Inattention	491	4.215
	Driving Mistake	535	4.592
	Drug Driving	132	1.133
	Drunk Driving	835	7.167
	Emergency	62	0.532
	Failure in Driving	6261	53.742
	Fatigue	177	1.519
	Inability	86	0.738
	Invalid Driver	121	1.039
	Risky Driving	2837	24.352
	Taking Medication	5	0.043
	Unsafe Vehicle Condition	25	0.215

applications in French-language scientific communities. In spite of a certain rarity of MCA-related research published in English-language publications that promote hypothetico-deductive approaches (Beaudouin, 2016), multiple correspondence analysis has recently found extensive applications in academic fields in social science including economics (Parchomenko, Nelen, Gillabel, & Rechberger, 2019), education (Costa, Santos, Cunha, Cotter, & Sousa, 2013; Kalayci & Basaran, 2014), psychology (Rodriguez-Sabate, Morales, Sanchez, & Rodriguez, 2017), public policy (Esmaelian, Tavana, Di Caprio, & Ansari,2017), and archaeology, etc. (Macheridis & Magnell, 2020) In particular, the use of MCA in transportation research receives increasing attention in recent years: Das and Sun studied vehicle-pedestrian crashes and fatal run-off-road crashes by using MCA approach (Das & Sun, 2016). Other researchers applied the same methodology into various traffic crash scenarios through different angles of research. (Factor, Yair, & Mahalel, 2010; Mitchell, Senserrick, Bambach, & Mattos, 2015; Jalayer & Zhou, 2016; Jalayer, Pour-Rouholamin, & Zhou, 2018). In addition, Chauvina expanded the application scope of MCA to maritime accidents analysis (Chauvin, Lardjane, Morel, Clostermann, & Langard, 2013). To our knowledge, no previous study using multiple correspondence analysis has been performed to recognize the associated contributing factors in severe traffic crashes from Texas area, where the regional traffic fatality rate is considerably high compared with the national average (NHTSA, 2021) and underlying causes need to be ascertained.

The theoretical foundation of multiple correspondence analysis (MCA) is intricate and has been well elucidated in previous publications (Benzécri, 1973; Roux & Rouanet, 2010). The core component in MCA is an indicator matrix (also called complete disjunctive table) in which the columns of table refer to the categories of qualitative variables corresponding to various contributing factors in crash analysis while the rows represent each individual crash record (Greenacre, 1993; Greenacre & Blasius, 2006). The point clouds of individuals and categories (Le Roux & Rouanet, 2004) are built through the calculation of inter-individual and inter-category distances, distance of points to the origin and total inertia of point clouds based on the components in the complete disjunctive table. The relevant mathematical description is accessible from the online tutorial of Husson's textbook and previous MCA-related publications (Das & Sun, 2015; Das & Sun, 2016); thus, it will not be detailed in this research. Table 4 provides a summary of relevant parameters in the indicator matrix and equations for creating point cloud of individuals and categories, respectively.

Results

From Table 2, it is shown that the majority of the selected contributing factors have large and P-values lower than 0.05, which suggests they are significantly associated with the crash severity levels. One exception comes from the factor of driver age with a P-value larger than the specified significance level. While this implies the current classification of driver's age as "16-30", "31-45", "46-60", "60+" is not sensitive to the variance in crash severity, we do find other way of grouping driver's age with fewer bins can yield a P-value less than 0.5 and larger . For the purpose of doing comprehensive cluster analysis on contributing factors in the following research, the current categorization of the factor "Driver Age" is preserved for further discussion.

In Figure 1, a panoramic 2-dimensional MCA factor map is presented in which a point cloud of all variable categories and associations among categories can be explicitly visualized based on the closeness between category points on the map. The factor plot shows the distribution of coordinates of all the variable categories on an orthogonal coordinate system constituted by two principal dimensions: Dim1 and Dim2. Like Principal Component Analysis (PCA), MCA assumes the dimension with largest variance is perceived as the most principal direction which has the maximum eigenvalue. Shown in Table 5, the eigenvalue, percentage of variance of first 10 dimensions are listed in a descending order with corresponding cumulative percentage of variance. It is observed that the first two principal dimensions only account for 10% variation of the original data. This reveals the heterogeneous and complex nature of the contributing factors involved in traffic crash dataset where there are 12 categorical variables and 11,650 individual data points, leading to high level of variability and uncorrelatedness.

The pattern of the point cloud of category can be interpreted in three aspects: first, the distance between any variable categories reflects a measure of their correlations, combination clouds can be created when some variable categories are relatively close (Das & Sun, 2016). Second, negatively correlated variable categories are located on the opposite sides of the origin of the factor plot, i.e., the coordinates of "Morning Rush" and "Weekend" (see dashed box), alluding that the occurrence of morning rush is not likely to happen on the weekend which is line with common sense. Third, the distance between category points and the origin reflects the quality of the variable category in a 2-dimensional orthogonal coordinate. From Figure 1, it is clear to see category points of "Daylight", "Dark, Lighted", "Dark, Not Lighted" and "Non Rush Nighttime", "Non Rush Daytime" are spreading out over the 1st principal dimension which indicates the categorical levels in the contributing factors "Crash Time" and "Light Condition" can be distinctly classified along the Dim 1. Similarly, the weather condition of "Rain", "clear" and surface condition of "Wet", "Dry" can be easily characterized by using Dim 2. As a result, these particular variable categories are better represented among other categories on the current factor map.

Table 4. Description of relevant parameters and equations for point clouds of individuals and categories

Parameters	Description
G_I	Center of gravity of the point cloud of individuals
G_J	Center of gravity of the point cloud of categories
I	Total number of the individuals i
$1/I$	The weight of an individual
J	Total number of the qualitative categorical variables j
K	Total number of categories k in all variables
K_j	The number of categories in the given variable j
N_I	Total inertia of the point cloud with I individuals
N_J	Total inertia of the point cloud with J categorical variables
O_{R^I}	The origin in the space R^I, $G_I = O_{R^I}$
O_{R^J}	The origin in the space R^J, $G_J = O_{R^J}$
p_k	The proportion of individuals in category k
v_{ij}	Category of j-th variable possessed by the i-th individual
y_{ik}	= 1 if the i-th individual is in k-th category of the j-th variable (for each p_k); = 0 otherwise

Point cloud	Distance between a pair of points in the cloud	Distance between points and origin	Total inertia
Individuals	$d^2_{(i,i')} = \frac{1}{J}\sum_{k=1}^{K}\frac{1}{p_k}(y_{ik} - y_{i'k})^2$	$d^2_{(i,G_I)} = \frac{1}{J}\sum_{k=1}^{K}\frac{y_{ij}}{p_k} - 1$	$N_I = \frac{K}{J} - 1$
Categories	$d^2_{(k,k')} = \frac{p_k + p_{k'} - 2p_{kk'}}{p_{kk'}}$	$d^2_{(k,G_J)} = \frac{1}{p_k} - 1$	$N_J = \frac{K}{J} - 1$

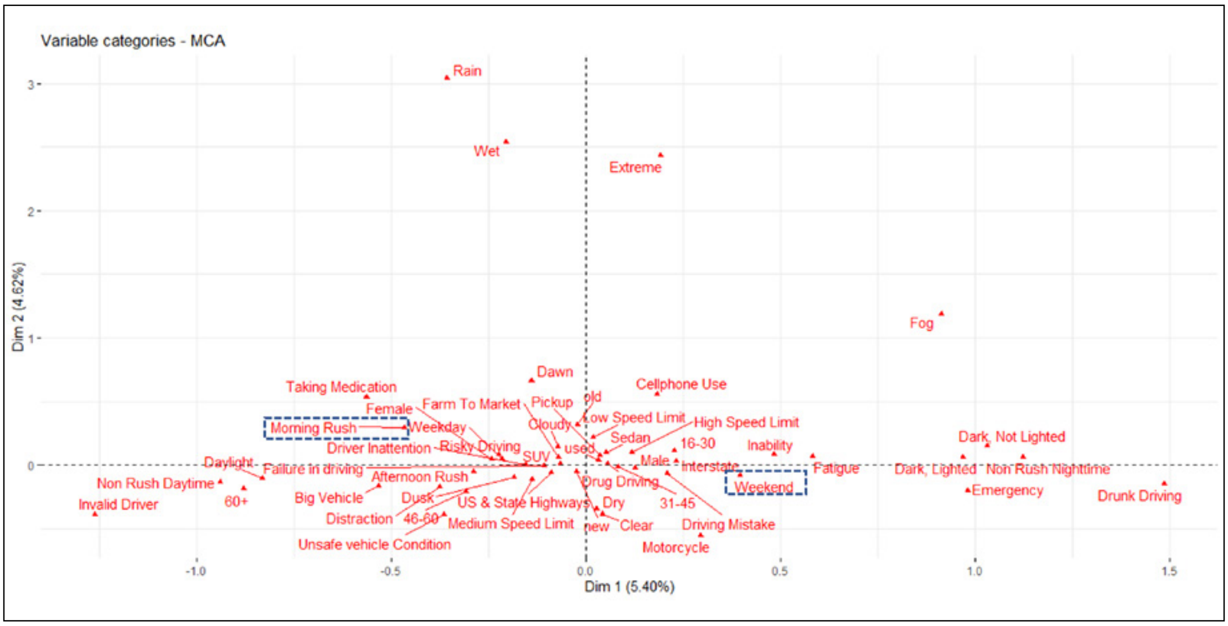

Figure 1. MCA factor map for variable categories

Table 5. Eigenvalues and percentages of variance of the first 10 dimensions

	Eigenvalue	Percentage of Variance	Cumulative Percentage of Variance
Dim 1	0.1801	5.4033	5.4033
Dim 2	0.1541	4.6224	10.0257
Dim 3	0.1223	3.6703	13.6960
Dim 4	0.1141	3.4230	17.1190
Dim 5	0.1065	3.1943	20.3133
Dim 6	0.1029	3.0879	23.4012
Dim 7	0.1003	3.0103	26.4115
Dim 8	0.0959	2.8756	29.2871
Dim 9	0.0933	2.7979	32.0851
Dim 10	0.0907	2.7208	34.8059

The quality of a variable category can be quantified by its contribution (in %) to the definition of principal dimensions. The larger the percentage value of a category to a given dimension, the more it can explain the variability in the dataset along that dimension. Shown in the Figure 2(a), 2(b), two bar plots demonstrate the most contributed variable categories to Dim 1 and Dim 2, respectively, only top 10 categories are displayed in each plot. It is evident to see contributing factors of "Crash Time" and "Light Condition" are the dominant categorical variables in the 1st principal dimension while "Weather Condition" and "Surface condition" account for the most variances in the 2nd principal dimension.

Discussion

Four combination clouds are created on the factor map shown collectively in Figure 3. In each cloud, several points of variable category are clustered together based on their relative proximity and interestingness. In the combination cloud 1, three types of variable categories are grouped into the cloud: "Rain", "Wet" and "Extreme", which indicates the occurrences of traffic crashes leading to severe injuries and fatality are significantly correlated with adverse weather conditions like rain and extreme climate events. The wet and slippery road surface, as the byproduct of rainy and humid climate is very likely to cause the serious traffic crashes. These findings accord with the conclusions drew from previous studies (Sherretz & Farhar, 1978; Andrey, Mills, Leahy, & Suggett, 2003). The second combination cloud encompasses variable categories of "Dark, Not Lighted", "Dark, Lighted", "Non Rush Nighttime" and other categories like "Fatigue", "Emergency" and "Drunk Driving". This combination can be explained by the fact that driving in complete darkness during non-rush time at night can be risky to cause severe or even fatal traffic crashes even in the presence of street lights. The inclusion of other two risk factors implies that drunk driving can be dangerous and should be strictly prohibited by law while driving in an emergency condition is also prone to severe traffic crashes partly due to the prevalence of incompetence and inexperience of drivers in safely handling emergency driving situations.

Combination cloud 3 and 4 are relatively closer to the origin of the coordinate, implying the included variable categories are less-represented by the current two principal dimensions. However, a careful examination of these two combination clouds still yields some meaningful information about the potential contributing factors linked to severe traffic crashes. In combination cloud 3, it is obvious to see male drivers in their young and early middle adulthood (ages between 16 to 45), driving used sedans

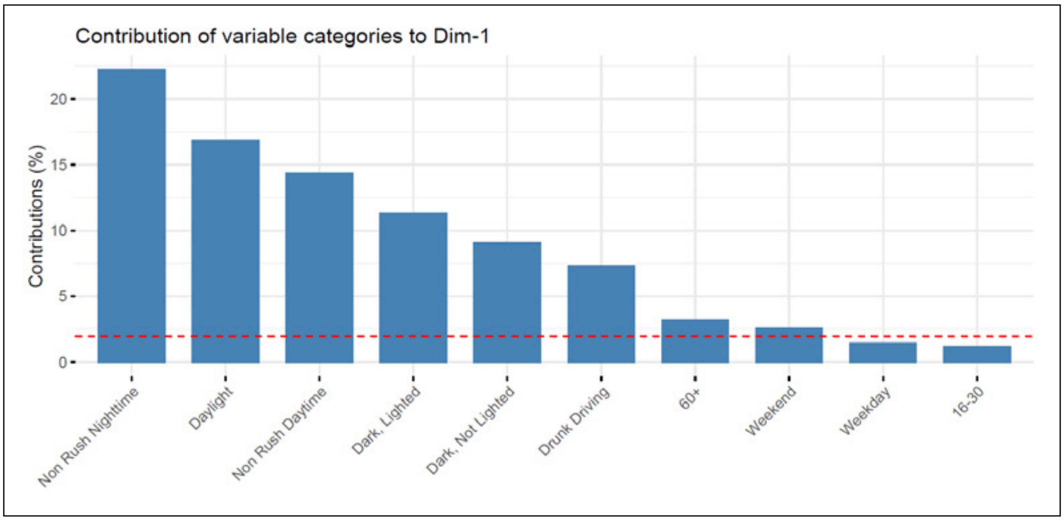

Figure 2(a). of variable categories to the 1st principal dimension

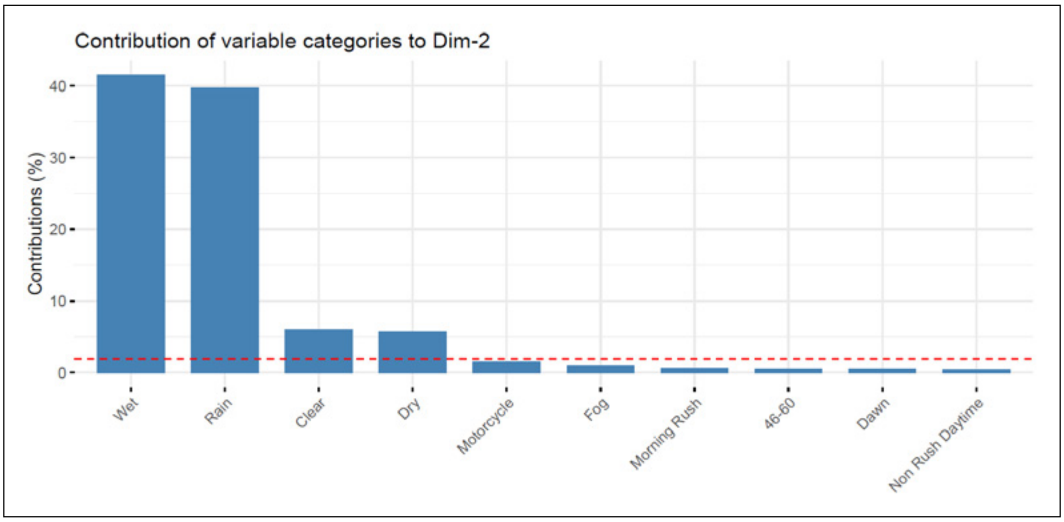

Figure 2(b). Contribution of variable categories to the 2nd principal dimension

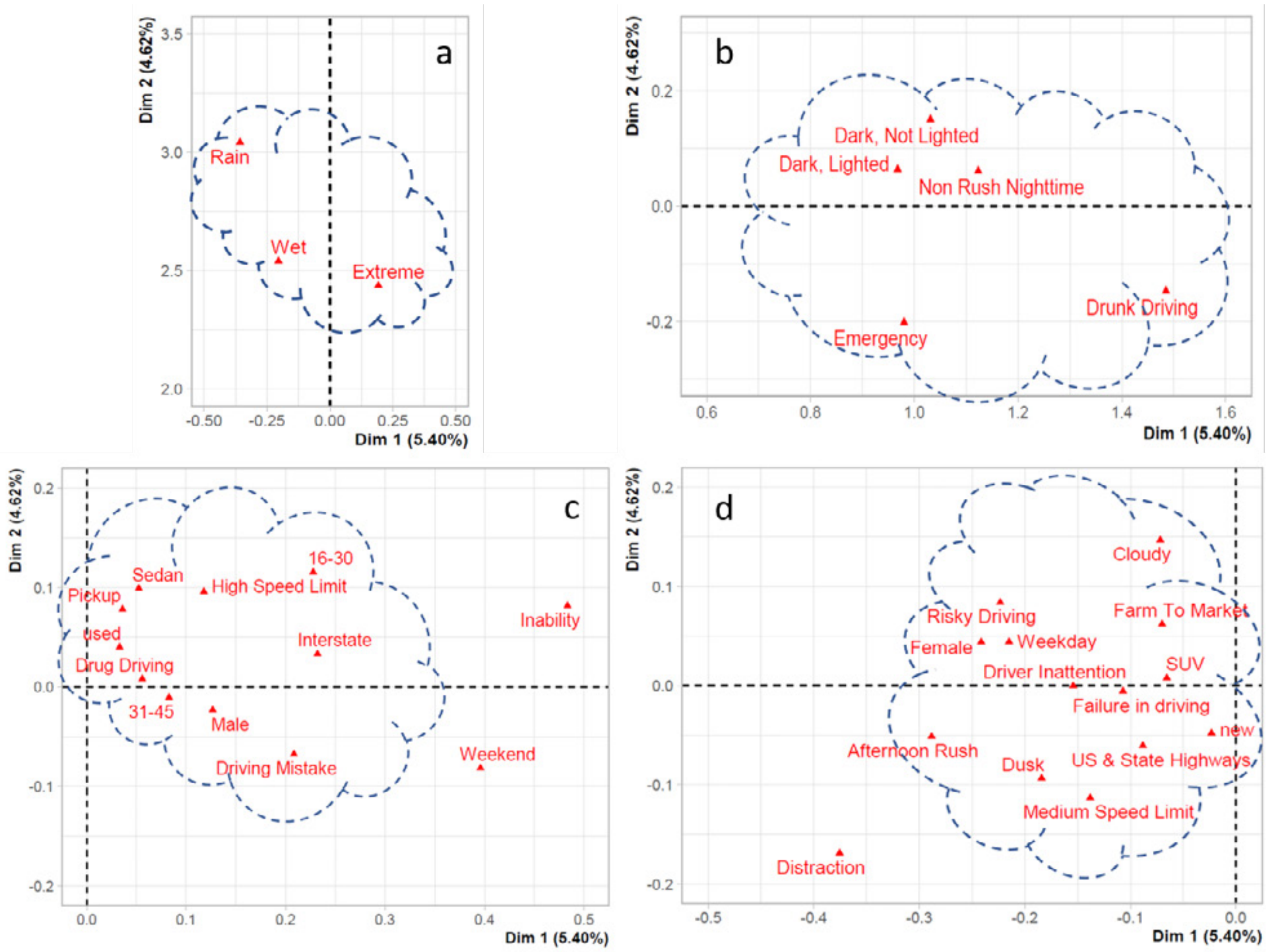

Figure 3 (a-d) Combination cloud 1; Combination cloud 2; Combination cloud 3; Combination cloud 4

or pickups on interstate roads in high speed are highly associated with traffic crashes resulted in serious injuries and fatality. In addition, risky behavioral factors like drug driving and driving mistakes are usually concurrent with the traffic accident scenarios described above, the underlying association is also confirmed by previous studies where evidences of link between drug consumption and motor vehicle crashes with high morbidity were provided (Sewell, Poling, & Sofuoglu, 2009; Romano & Voas, 2011). Therefore, stricter drug-related traffic laws need to be enacted to curb drugged driving among young drivers in the future. Combination cloud 4 relates some different variable categories such as "Driver Inattention", "Failure in driving", "Risky Driving", "Female", "Afternoon rush" and "Dusk", etc. This combination indicates insufficient attention to driving details and improper driving habits are likely to cause severe crashes among female drivers during the afternoon rush hour in weekdays. Moreover, driving on Farm-to-market road and State Highways in the rural area where street lights are usually sparsely distributed in a cloudy weather or at dusk is more likely to cause higher traffic crash severity and fatality, this might be due to the insufficient lighting in the abovementioned scenarios that reduce the drivers' visibility and perception to the ambient environment. This analysis is endorsed by other research (Jägerbrand & Sjöbergh, 2016) which confirms the relationship between road lighting and traffic safety because of light condition's impact on visual performance during driving.

In sum, a multivariate statistical method of multiple correspondence analysis (MCA) has been applied to identify the associated contributing factors that contribute to traffic crashes resulting in severe injuries and fatality to the drivers. The use of combination clouds gives explicit graphical display of multiple clusters of variable categories, from which the impact and combinatorial effect of various factors can be easily interpreted in different traffic crash scenarios. Despite the achievements from applying MCA approach in this crash factor analysis, a limitation should be pointed out that because of the highly uncorrelated structure of the traffic crash dataset that contains a dozen of categorical variables and a large volume of individual crash data points, only 10% of the total variance is retained by the selected two principal dimensions. This may lead to an underrepresentation of some contributing factors based on the current 2-dimensional factor plot. Therefore, further analysis on the MCA factor map constituted by 3rd and 4th principal dimensions might be needed to ensure the significance of other combinations of variables is examined.

Conclusions

In this study, traffic crash analysis based on the historical crash data from Southeast Texas area has been performed to identify the significant contributing factors that affect the severity of traffic crashes. Pearson's chi-squared test reveals that factors like weather condition, lighting condition, crash time, speed limit, road class, surface condition, risk factor have statistically significant associations with different levels of crash severity. Moreover, multiple correspondence analysis (MCA) is implemented to identify groups of contributing factors and study their combinatorial influence on severe crash-induced injuries and fatality by creating a number of combination clouds on the factor map. Based on the relative closeness of variable categories on the 2-dimensional space, category points from multiple contributing factors are clustered together and form combination clouds that provide a collective graphical view of the potential traffic scenarios in which deadly crash can take place. Upon the analysis on the elements contained in these combination clouds, following indications can be achieved:

- Driving in adverse climate conditions like rain and extreme weather on the wet road surface has a higher chance to cause traffic crashes with severe injuries or even fatality.
- Driving in a complete dark environment during non-rush time regardless of the presence of street lights is more likely to induce serious traffic accidents mainly because of the poor light condition. Behaviors like drunk driving and driving in emergency condition can impose more risks on the drivers and result in severely injured or fatal crashes.
- Male drivers in youth and early middle adulthood are more prone to traffic crashes ended up in being seriously injured and killed when they are driving used vehicles in high speed on the interstate road under the detrimental effects of drug use and resultant driving mistakes.
- Risk factors like driving inattention, risky driving and failure in driving are likely to the cause crashes with high severity among female drivers during the afternoon rush in the weekdays.
- Driving on Farm-to-market roads and State Highways at dusk or in the cloudy weather is subject to traffic crashes with higher crash severity and fatality, which can be explained by the insufficiency of ambient light condition that leads to reduced visibility and shorter reaction time of drivers to avoid the accident.

In spite of the limitation of MCA application in this research which might underrepresent the significance of other clusters of contributing factors due to the relatively low variance explained by the first two principal dimensions, the graphical representation of the clustered factors helpfully shed light on the traffic crashes causing severe injuries and death, where assorted contributing factors are playing a combinatorial role in the occurrence of the crash. Based on the results from this crash analysis using MCA method, the following improvements are advised: 1. Remediation of road infrastructure issues related to surface and light conditions on the roads. 2. Modification of current traffic codes and enactment of stricter law to control risky driving behaviors. 3. Delivering more targeted driving safety education to drivers of different ages and genders accordingly

Acknowledgements

The researchers express great gratitude to the abundant material and informative support from Lamar University and Georgia Institute of Technology.

This study was partially supported by the Natural Science Foundation (1726500). The findings and conclusions of this paper are those of the authors and do not necessarily represent the official position of NSF.

References

Abdel-Aty, M., Ekram, A. A., Huang, H., & Choi, K. (2011). A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis & Prevention, 43*(5), 1730-1737. https://doi.org/10.1016/j.aap.2011.04.003

Abdi, H., & Valentin, D. (2007) Multiple correspondence analysis. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks, California: SAGE Publications.

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics, 2*(4), 433-459. Hoboken, New Jersey: Wiley.

Andrey, J., Mills, B., Leahy, M., & Suggett, J. (2003). Weather as a chronic hazard for road transportation in Canadian cities. *Natural Hazards, 28*(2), 319-343. https://doi.org/10.1023/A:1022934225431

Association of Safe International Road Travel. (2019). *Annual United States Road Crash Statistics*. Retrieved from https://www.asirt.org/safe-travel/road-safety-facts/

Beaman, J. (2019). *Gulf Coast Hurricane Preparedness*. Mobile, Alabama: National Weather Service https://www.weather.gov/media/mob/pdf/GulfCoastHurricanePrep.pdf

Beaudouin, V. (2016). *Retour aux origines de la statistique textuelle: Benzécri et l'école française d'analyse des données.* 13es Journées internationales d'Analyse statistique des Données Textuelles, Nice, France. https://hal.archives-ouvertes.fr/hal-01376938

Benzécri, J. P. (1973). L'Analyse des Données. Vol. 2: *Analyse des Correspondances*. Paris, France: Dunod.

Brooks, J.O., Crisler, M.C., Klein, N., Goodenough, R., Beeco, R.W., Guirl, C., Tyler, P.J., Hilpert, A., Miller, Y., Grygier, J. & Burroughs, B., (2011). Speed choice and driving performance in simulated foggy conditions. *Accident Analysis & Prevention, 43*(3), 698-705. https://doi.org/10.1016/j.aap.2010.10.014

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J. P., & Langard, B. (2013). Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis & Prevention, 59*, 26-37. https://doi.org/10.1016/j.aap.2013.05.006

Clarke, D. D., Ward, P., Bartle, C., & Truman, W. (2006). Young driver accidents in the UK: The influence of age, experience, and time of day. *Accident Analysis & Prevention, 38*(5), 871-878. https://doi.org/10.1016/j.aap.2006.02.013

Costa, P. S., Santos, N. C., Cunha, P., Cotter, J., & Sousa, N. (2013). The Use of Multiple Correspondence Analysis to Explore Associations between Categories of Qualitative Variables in Healthy Ageing. *Journal of Aging Research, 2013,* 302163. https://doi.org/10.1155/2013/302163

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record, 2519*(1), 95-103. https://doi.org/10.3141/2519-11

Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research, 39*(2), 146-155. https://doi.org/10.1016/j.iatssr.2015.07.001

Duval, J. (2018). *Correspondence Analysis and Bourdieu's Approach to Statistics: Using Correspondence Analysis within Field Theory*, in Thomas Medvetz and Jeffrey J. Sallaz (ed.), *The Oxford Handbook of Pierre Bourdieu*, Oxford, United Kingdom: Oxford University Press.

Esmaelian, M., Tavana, M., Di Caprio, D., & Ansari, R. (2017). A multiple correspondence analysis model for evaluating technology foresight methods. *Technological Forecasting and Social Change, 125,* 188-205. https://doi.org/10.1016/j.techfore.2017.07.022

Evans, L., & Frick, M. C. (1992). Car size or car mass: which has greater influence on fatality risk? *American Journal of Public Health, 82*(8), 1105-1112. https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.82.8.1105

Factor, R., Yair, G., & Mahalel, D. (2010). Who by accident? The social morphology of car accidents. *Risk Analysis: An International Journal, 30*(9), 1411-1423. https://doi.org/10.1111/j.1539-6924.2010.01423.x

Ghasemzadeh, A., & Ahmed, M. M. (2017). Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. *Transportation Research Record, 2663*(1), 99-108. https://doi.org/10.3141/2663-13

Greater Houston Partnership. (2020). *Houston Facts 2020*. Houston, TX: Greater Houston Partnership. https://www.houston.org/sites/default/files/2020-08/Houston%20Facts%202020_1.pdf

Greenacre, M. (1993). *Correspondence analysis in practice.* Cambridge, Massachusetts: Academic Press.

Greenacre, M., & Blasius, J. (2006). *Multiple correspondence analysis and related methods.* London, United Kingdom: Chapman and Hall/CRC.

Hang, J., Yan, X., Ma, L., Duan, K., & Zhang, Y. (2018). Exploring the effects of the location of the lane-end sign and traffic volume on multistage lane-changing behaviors in work zone areas: A driving simulator-based study. *Transportation Research Part F: Traffic Psychology and Behaviour, 58,* 980-993. https://doi.org/10.1016/j.trf.2018.07.024

Harb, R., Radwan, E., & Yan, X. (2007). Larger size vehicles (LSVs) contribution to red light running, based on a driving simulator experiment. *Transportation Research Part F: Traffic Psychology and Behaviour, 10*(3), 229-241. https://doi.org/10.1016/j.trf.2006.10.005

Harb, R., Radwan, E., Yan, X., & Abdel-Aty, M. (2007). Contribution of light truck vehicles to rear-end collisions. *Accident Analysis & Prevention, 35*(5), 1026-1036. https://doi.org/10.1016/j.aap.2007.01.007

Hayakawa, H., Fischbeck, P. S., & Fischhoff, B. (2000). Traffic accident statistics and risk perceptions in Japan and the United States. *Accident Analysis & Prevention, 32*(6), 827-835. https://doi.org/10.1016/S0001-4575(00)00007-5

Hijar, M., Carrillo, C., Flores, M., Anaya, R., & Lopez, V. (2000). Risk factors in highway traffic accidents: a case control study. *Accident Analysis & Prevention, 32*(5), 703-709. https://doi.org/10.1016/S0001-4575(99)00116-5

Holubowycz, O. T., Kloeden, C. N., & McLean, A. J. (1994). Age, sex, and blood alcohol concentration of killed and injured drivers, riders, and passengers. *Accident Analysis & Prevention, 26*(4), 483-492. https://doi.org/10.1016/0001-4575(94)00064-S

Huang, H., Siddiqui, C., & Abdel-Aty, M. (2011). Indexing crash worthiness and crash aggressivity by vehicle type. *Accident Analysis & Prevention, 43*(4), 1364-1370. https://doi.org/10.1016/j.aap.2011.02.010

Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory multivariate analysis by example using R* (Vol. 15). Boca Raton, Florida: CRC press.

Jägerbrand, A. K., & Sjöbergh, J. (2016). Effects of weather conditions, light conditions, and road lighting on vehicle speed. *SpringerPlus, 5*(1), 1-17. DOI 10.1186/s40064-016-2124-6

Jalayer, M., & Zhou, H. (2016). A multiple correspondence analysis of at-fault motorcycle-involved crashes in Alabama. *Journal of Advanced Transportation, 50*(8), 2089-2099. DOI: 10.1002/atr.1447

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.

Jamson, S. L., Tate, F. N., & Jamson, A. H. (2005). Evaluating the effects of bilingual traffic signs on driver performance and safety. *Ergonomics, 48*(15), 1734-1748. https://doi.org/10.1080/00140130500142191

Kalayci, N., & Basaran, M. A. (2014). A combined approach using multiple correspondence analysis and log-linear models for student perception in quality in higher education. *Procedia Economics and Finance, 17*, 55-62. https://doi.org/10.1016/S2212-5671(14)00878-8

Keall, M. D., Frith, W. J., & Patterson, T. L. (2004). The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand. *Accident Analysis & Prevention, 36*(1), 49-61. https://doi.org/10.1016/S0001-4575(02)00114-8

Lebaron, F. (2009). How Bourdieu “quantified” Bourdieu: The geometric modelling of data. In *Quantifying Theory: Pierre Bourdieu* (pp. 11-29). Berlin, Germany: Springer, Dordrecht.

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

Le Roux, B., & Rouanet, H. (2004). *Geometric data analysis: from correspondence analysis to structured data analysis*. Paris, France: Springer Science & Business Media.

Le Roux, B., & Rouanet, H. (2010). *Multiple Correspondence Analysis*. Thousand Oaks, California: SAGE Publications, Inc.

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z. S., Prevedouros, P. D., & Ma, D. T. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident Analysis & Prevention, 124*, 219-229. https://doi.org/10.1016/j.aap.2018.12.020

Lyons, S. W. (1990). Spatial and temporal variability of monthly precipitation in Texas. *Monthly Weather Review, 118*(12), 2634-2648.

Macheridis, S., & Magnell, O. (2020). Disentangling taphonomic histories at Old Uppsala, a Late Iron Age central place in Sweden, using Multiple Correspondence Analysis (MCA). *Journal of Archaeological Science: Reports, 33,* 102536. https://doi.org/10.1016/j.jasrep.2020.102536

Massie, D. L., Campbell, K. L., & Williams, A. F. (1995). Traffic accident involvement rates by driver age and gender. *Accident Analysis & Prevention, 27*(1), 73-87. https://doi.org/10.1016/0001-4575(94)00050-V

Mitchell, R. J., Senserrick, T., Bambach, M. R., & Mattos, G. (2015). Comparison of novice and full-licensed driver common crash types in New South Wales, Australia, 2001-2011. *Accident Analysis & Prevention, 81,* 204-210. https://doi.org/10.1016/j.aap.2015.04.039

Mueller, A. S., & Trick, L. M. (2012). Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance. *Accident Analysis & Prevention, 48,* 472-479. https://doi.org/10.1016/j.aap.2012.03.003

Naik, B., Tung, L. W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. *Journal of Safety Research, 58,* 57-65. https://doi.org/10.1016/j.jsr.2016.06.005

National Center for Statistics and Analysis. (2020). *Passenger vehicles: 2018 data.* Washington, D.C.: National Highway Traffic Safety Administration. (Traffic Safety Facts. Report No. DOT HS 812 962). https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812962.

National Highway Traffic Safety Administration. (2021). *Traffic Safety Facts Annual Report*. Retrieved from https://cdan.dot.gov/tsftables/National%20Statistics.pdf

Nielsen-Gammon, J. W. (2011). The changing climate of Texas. In *The impact of global warming on Texas* (Schmandt, J., G. North, and J. Clarkson, Eds.), University of Texas Press, College Station, TX, 2011, pp. 39-68.

Ning, Z. H., & Abdollahi, K. (2003). Gulf Coast regional climate. *Gulf Coast Region: Findings of the Gulf Coast Regional Assessment. Gulf Coast Climate Change Assessment Council*, 79-82.

Parchomenko, A., Nelen, D., Gillabel, J., & Rechberger, H. (2019). Measuring the circular economy-A Multiple Correspondence Analysis of 63 metrics. *Journal of Cleaner Production, 210*, 200-216. https://doi.org/10.1016/j.jclepro.2018.10.357

Plainis, S., Murray, I. J., & Pallikaris, I. G. (2006). Road traffic casualties: understanding the night-time death toll. *Injury Prevention, 12*(2), 125-138. doi: 10.1136/ip.2005.011056

Qiu, L., & Nixon, W. A. (2008). Effects of adverse weather on traffic crashes: systematic review and meta-analysis. *Transportation Research Record, 2055*(1), 139-146. https://doi.org/10.3141/2055-16

Rodriguez-Sabate, C., Morales, I., Sanchez, A., & Rodriguez, M. (2017). The multiple correspondence analysis method and brain functional connectivity: its application to the study of the non-linear relationships of motor cortex and basal ganglia. *Frontiers in Neuroscience, 11,* 345. https://doi.org/10.3389/fnins.2017.00345

Romano, E., & Voas, R. B. (2011). Drug and alcohol involvement in four types of fatal crashes. *Journal of Studies on Alcohol and Drugs, 72*(4), 567-576. https://doi.org/10.15288/jsad.2011.72.567Sewell, R. A., Poling, J., & Sofuoglu, M. (2009). The effect of cannabis compared with alcohol on driving. *American Journal on Addictions, 18*(3), 185-193. doi: 10.1080/10550490902786934

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

Texas Department of Transportation. (2018). *Texas Motor Vehicle Traffic Crash Facts Calendar Year 2017*. Retrieved from https://ftp.txdot.gov/pub/txdot-info/trf/crash_statistics/2017/01.pdf

Tu, H., Li, Z., Sun, Y., & Dai, X. (2014). Driving behavior in case of haze: a high fidelity driving simulator study. In *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems* (pp. 2408-2421).

Wang, J. H., & Song, M. (2011). Assessing drivers' tailgating behavior and the effect of advisory signs in mitigating tailgating. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Lake Tahoe, California. DOI: 10.17077/drivingassessment.1450

World Health Organization. (2017*). Road Safety: Basic Facts*. Geneva, Switzerland: World Health Organization.https://www.who.int/violence_injury_prevention/publications/road_traffic/Media_brief_all_factsheets_web_rev_nov_2017.pdf?ua=1.

World Health Organization. (2018). *Global status report on road safety 2018: summary*. Geneva, Switzerland: World Health Organization. (WHO/NMH/NVI/18.20). License: CC BY-NC-SA 3.0 IGO). https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf?ua=1.

Zhang, G., Yau, K. K., & Chen, G. (2013). Risk factors associated with traffic violations and accident severity in China. *Accident Analysis & Prevention, 59*, 18-25. https://doi.org/10.1016/jaap.2013.05.004

Appendix

Definitions of traffic safety related terms

Term	**Definition**
Driving mistake	The term driving mistake is synonym for driving error which can be classified into four categories according to 1. recognition errors (inadequate surveillance, internal distraction, and external distraction), 2. decision errors (speeding, illegal maneuver, aggressive driving) 3. performance errors (overcompensation, poor directional control) 4. critical non-performance errors (fatigue, sleeping, physical impairment) [6,8] [6] K. Rumar, The basic driving error: late detection, Ergonomics 33 (1990) 1281-1290. [8] J. Treat, A study of pre-crash factors involved in traffic accidents, HSRI Res. Rev. 10 (1980) 1-35.
Failure in driving	Failure in driving is a general term that encompasses any types of human error that take place during driving (e.g. fail to control speed, fail to pass to left safely). Details are listed in https://cris.dot.state.tx.us/public/Query/app/query-results/list
Emergency	In this study emergency refers to any sudden, abnormal circumstance that affects safe driving and calls for immediate action to cope with, e.g. animals crossing the road.
Big vehicle	Big vehicle generally refers to any large and heavy vehicles weighing more than 4.5t, such as truck, bus, fire truck.
New vehicle	In this research, we group the vehicles with 'Vehicle Age' from 0 to 9 years as 'New vehicle', 'Vehicle Age' means the age of a vehicle computed by totaling the number of the years in between and including both the calendar year and the model year.
Used vehicle	'Vehicle Age' from 10 to 19 years as 'Used vehicle'
Old vehicle	'Vehicle Age' from 20 to 27 (maximum year in the dataset) as 'Old vehicle'.
Crash time	Crash time is divided into 24 groups: 00:00-00:59 as "0"; 01:00-01:59 as "1"; 02:00- 02:59 as "2"; 03:00-03:59 as "3"; 04:00-04:59 as "4"; 05:00-05:59 as "5"; 06:00-06:59 as "6"; 07:00-07:59 as "7"; 08:00-08:59 as "8"; 09:00-09:59 as "9"; 10:00-10:59 as "10"; 11:00-11:59 as "11"; 12:00-12:59 as "12"; 13:00-13:59 as "13"; 14:00-14:59 as "14"; 15:00-15:59 as "15"; 16:00-16:59 as "16"; 17:00-17:59 as "17"; 18:00-18:59 as "18"; 19:00-19:59 as "19"; 20:00-20:59 as "20" 21:00-21:59 as "21"; 22:00-22:59 as "22"; 23:00-23:59 as "23". Next, numbers of 6,7,8 are grouped as "Morning Rush", numbers of 17,18,19 are grouped as "Afternoon Rush", numbers of 20, 21, 22, 23, 0, 1, 2, 3, 4, 5 are grouped as "Non Rush Nighttime", numbers of 9, 10, 11, 12, 13, 14, 15, 16 are "Non rush daytime"