Uncovering Deep Structure of Determinants in Large Truck Fatal Crashes

Subasish Das, Ph.D.
(Corresponding Author)
Associate Transportation Researcher, A&M Transportation Institute
1111 RELLIS Parkway, Room 4414, Bryan, TX 77807
Email: s-das@tti.tamu.edu
ORCID ID: 0000-0002-1671-2753

Mouyid Islam, Ph.D.
Research Faculty, Center for Urban Transportation Research, University of South Florida
4202 E. Fowler Ave., CUT100, Tampa, FL 33620
Email: mouyid@usf.edu
ORCID: 0000-0001-9068-9187

Anandi Dutta, Ph.D.
Senior Lecturer, Computer Science and Engineering, Ohio State University
2015 Neil Ave, Columbus, OH 43210
Email: anandixd@tamu.edu

Tahmida Hossain Shimu
Traffic Designer, HDR
17111 Preston Road, Suite 300, Dallas, TX 75248 -1232
Email: Tahmida.Shimu@hdrinc.com

Submitted to: Transportation Research Record
Date: December 15, 2019

ABSTRACT

The number of fatalities and severe injuries in large truck-related crashes has significantly increased since 2009. According to the safety experts, the recent increase in large truck-related crashes can be explained by the significant growth in freight tonnage all over the U.S. over the past few years. This notable freight-haul growth has allowed continuous day-night movement of freight on roads and highways, exposing the trucks to a greater number of potential crashes or near-crash scenarios. There are many ongoing research efforts that aim to identify the different factors that influence large truck crashes; however, further research with innovative approaches is still needed to better understand the relationship between crash-related factors. In this study, the project team applied taxicab correspondence analysis (TCA), a data mining method known for dimension reduction, to large truck fatal crash data to investigate the complex interaction between multiple factors under a two-dimensional map. For this study, six years (2010-2015) of large truck fatal crash data from the Fatality Analysis Reporting System (FARS) were used. The study found five clusters of attributes that show patterns of association between different crash attributes such as two-lane undivided roadways, intersection types, posted speed limit, crash types, number of vehicles, driver impairment, and weather. The findings of this study will help the safety professionals, trucking industry, and policymakers to make decisions for safer road design, and improvement in truck driver training, and education.

Keywords: large truck crashes, fatal crashes, injury severity, taxicab correspondence analysis, FARS.

INTRODUCTION

Large trucks account for over 80 percent of freight tonnage. Large truck crashes interrupt traffic flow and cause tremendous economic loss; however, the number of deaths and the severity of injuries sustained in these crashes are certainly the most acute and dire effects. The number of fatal crashes involving large trucks or buses has increased by 29 percent since 2009 (*1*). According to the National Highway Traffic Safety Administration (NHTSA), there were 4,761 fatalities involving large trucks in the year 2017 (increased by 9 percent from 2016); this amount accounted for 8 percent of all fatal crashes in that year. Furthermore, the risk associated with large trucks in terms of exposure accounts for 9 percent of the overall vehicle miles traveled (VMT) and 4 percent among all registered vehicles in 2016. According to national crash statistics, large truck-related crashes have increased significantly in recent years. This has drawn the attention of safety researchers, but more focus should be directed toward the issue in order to improve overall truck safety and to mitigate the impacts of these crashes on the U.S. economy and human lives. A more robust statistical approach was needed to understand the mechanisms and factors of roadway crashes involving large trucks.

A large truck is defined as a "medium or heavy truck having a gross vehicle weight rating (GVWR) of more than 10,000 pounds, excluding busses and recreational vehicles (RVs)" (*1*). Due to their considerable weight and size, crashes involving large trucks are more likely to result in a fatality or severe injury *(2, 3*). Hence, a specialized investigation of crash causation with fatality and injury severity analysis is necessary to identify the mechanisms involved in these crashes and to provide effective countermeasures. There are many ongoing research efforts, both conventional and innovative, that aim at determining the factors that influence large truck crash occurrences to develop more effective safety treatments. However, the number of large truck-related crashes is still very high, as shown by recent crash statistics. Thus, there is a need for research efforts with additional resources and newer approaches and techniques.

Correspondence analysis (CA) is a multivariate statistical method that summarizes the essential aspects of a data set by projecting the multivariate data on two-dimensional maps. A sturdy-robust-resistant variant of CA is known as taxicab correspondence analysis (TCA). This new method can smoothly handle complex datasets and produce satisfactory and meaningful results in the presence of outliers. In recent years, attention has been increasingly directed toward determining the factors that significantly affect crash occurrences. This study used large truck-related crash data from 2010-2015 from the Fatality Analysis Reporting System (FARS). The application of TCA on this dataset is appropriate due to the method's suitability in addressing the research problem related to this data.

LITERATURE REVIEW

A few recent research studies have conducted innovative statistical analyses to investigate traffic crashes involving large trucks. One of the earliest research efforts in documenting and investigating crash causation and crash risk factors associated with large trucks was carried out jointly by the U.S. Department of Transportation's (DOT) Federal Motor Carrier Safety Administration (FMCSA) and NHTSA in the early 2000s (*4*). This study, named the Large Truck Crash Causation Study (LTCCS), used crash data from 17 states from 2001 to 2003 and aimed to identify key factors of large truck crashes. Although the database created under this project contains descriptive data like other national traffic safety databases (e.g., FARS), the LTCCS study also considered pre-crash factors leading to a crash occurrence such as driver fatigue, distraction, weather, and roadway conditions. Before the LTCCS database was created, crash count and severity analyses involving large trucks were conducted using conventional police report and hospital data (*5*).

Lemp *et al.* (*6*) conducted an extensive large truck-related crash severity analysis using standard and heteroskedastic ordered probit models with the LTCCS, General Estimates System (GES), and Vehicle Inventory and Use Survey (VIUS) data sets. The focus of this study was to analyze the effect of vehicle, driver/occupant, and environmental characteristics on the severity of injuries sustained during large truck-related crashes. In another study, Hickman *et al.* (*7*) compared the LTCCS data with Naturalistic Driving (ND) data and argued that this comparison is necessary in order to bridge the information gap between high severity crash occurrences in the LTCCS dataset with non-crash related vehicular conflicts. Koupaenejad (*8*) investigated the factors contributing to the severity of crashes occurring between passenger cars and large trucks by using multinomial logit (MNL) and ordered probit models.

The factors influencing crash severity in hazardous material (HAZMAT) carrying large truck-related crashes utilizing ordered probit models with random and set parameters were investigated by Uddin *et al. (9)*. In this study, the most substantial factors that influence the crash severity of HAZMAT truck crashes were found to be drivers' sex and age, time of day, lighting condition, terrain, and crash type. Additionally, Islam and Hernandez *(10)* included random parameter ordered probit models to approximate the possibility of five injury severity outcomes. In another study, Islam and Hernandez (11) utilized random parameter tobit regression examining the large truck-involved fatal crash rates (instead of frequencies) in per million truck-miles traveled and ton-miles of freight as continuous censored variables. This study quantified fatality rates with factors related to the crash mechanism, temporal and spatial characteristics, road and environmental attributes, vehicle configuration, drivers and passenger attributes, which were found to be statistically significant. Some unobserved effects were found as a result from the intricate interaction between driver, roadway, traffic, and environmental factors that affect the injury severity of crashes. Qin *et al. (12)* utilized partial proportional odds (PPO), MNL, and mixed logistic (ML) models in large truck-related crash severity analysis. Drivers' age and seatbelt laws were found to not be statistically significant factors in influencing the severity of large-truck related crashes. In addition, Zheng *et al. (13)* used the gradient boosting data mining technique to perform a crash severity analysis. Eleven variables were found which accounted for over eighty percent of the total severe crashes in the dataset acquired from FMCSA in Colorado and North Dakota from 2010-2016. In comparison to other roadway or traffic-related factors, Dissanayake *(14)* found that driver-related factors had more of an effect on crash injury occurrences endured during large truck-related crashes.

Besides these severity analyses, the current literature consists of studies that investigate count data to examine and estimate the total number of crash occurrences and crash frequency. Amarasingha and Dissanayake *(15, 16)* evaluated the association of geometric properties and traffic with large truck-related crashes using Poisson and Negative Binomial (NB) regression models for access-controlled highways, like freeways. Similarly, Dong *et al. (17)* performed a study in which the authors considered NB and zero-inflated NB (ZINB) models and identified seven factors that were statistically significant in causing truck-related crashes, majority of which were consistent with the earlier studies. Offei *et al. (18)* reviewed the association between crash rates involving large trucks and truck percentages using regression models.

Apart from the common modeling techniques, other statistical and innovative machine learning methods were used to examine the injury severity and crash frequency of large truck-related crashes. Some of these examples consist of the classification tree model *(19)*, hierarchical

Bayesian random intercept model *(20)*, skewed logistic model *(21)*, and a risk analysis model dependent on collision diagram *(22)*. Even though most of these studies concentrated on highway crashes, Qin *et al. (23)* analyzed the factors contributing to the severity and frequency of truck-related crashes at freeway diverging section with the use of NB and MNL models. They found different genetic elements associated with median/shoulder width, deceleration areas, numbers of lanes, curvature, grade, speed limit, truck percentage, and AADT to be key factors in large truck-related crashes at freeway sections. However, Taylor *et al. (25)* conducted a similar study but found that geometric attributes of roadways (e.g., number of lanes, shoulder and median widths) and specific crash types were not significant factors influencing the severity and frequency of freight truck-related crashes. Ullman and Iragavarapu (*26*) investigated fatal work-zone crashes involving large trucks using odds ratio and found that these crashes are overrepresented in the fatal work-zone crash statistics when analyzed in terms of the time of day and roadway functional classes.

Although most of the studies found in the literature investigated factors that contribute to large truck-related crashes, a few other studies only looked at the effect of specific factors such as lighting condition (*2*), time of day (*27*), geometric elements such as horizontal curvatures (*31*), roadway separation (*32)*, and crash types (e.g., rear-end crashes (*28*), rollover crashes (*29*), run-off-road crashes (*30*), angle crashes (*21*), and driver age group *(46)* on crash occurrence or injury severity. Trimble *et al.* (*28*) conducted a GES analysis on rear-end crashes involving large trucks in order to determine the causation of this type of crash and to improve rear-signaling under a national project by NHTSA. Table 1 represents the key variables used by previous studies.

Correspondence analysis has become more popular in the field of transportation safety research *(33-40).* In the presence of rarely occurring variable categories, TCA produces visuals that are clearer and more easily interpreted than those produced by multiple correspondence analysis (MCA). The present study aims to investigate the trends of key contributing factors for large truck crashes via the application of TCA to six years (2010-2015) of FARS data. This new method gives this study unique value because it can easily handle the complex nature of large truck-related crash data and provide intuitive and significant results.

Table 1 Variables Considered in Large Truck Safety Studies

Variable name		**Found in studies**
Driver/Occupant characteristics	Action or inaction by drivers (maneuver, braking, acceleration, deceleration)	DOT (*4*), Hickman *et al.* (*7*), Islam and Hernandez (*10*), Kotikalapudi and Dissanayake (*14*), Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Trimble *et al.* (*28*), Park and Pierce (*29*), Al-Bdairi *et al.* (*30*)
	Driver perception and decision errors	DOT (*4*), Hickman *et al.* (*7*), Kotikalapudi and Dissanayake (*14*), Trimble *et al.* (*28*), Park and Pierce (*29*),
	Performance errors (sleeping, illness, disability)	DOT (*4*), Hickman *et al.* (*7*), Kotikalapudi and Dissanayake (*14*)
	Fatigue	DOT (*4*), Hickman *et al.* (*7*), Al-Bdairi *et al.* (*30*)
	Speeding	DOT (*4*), Hickman *et al.* (*7*), Kotikalapudi and Dissanayake (*14*), Dong *et al.* (*17*), Eustace *et al.* (*19*), Qin *et al.* (*23*), Trimble *et al.* (*28*), Islam and Hernandez (11)
	Driver age	Charbotel *et al.* (*5*), Koupaenejad (*8*), Uddin *et al.* (*9*), Zheng *et al.* (*13*), Islam and Hernandez (*10,11*), Eustace *et al.* (*19*), Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Taylor *et al.* (*25*), Islam and Seckin *(44)*

	Gender	Charbotel *et al.* (*5*), Koupaenejad (*8*), Uddin *et al.* (*9*), Islam and Hernandez (*10*), Islam and Hernandez (*10*), Eustace *et al.* (*19*), Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Taylor *et al.* (*25*)
	Driver licensing	Zheng *et al.* (*13*), Al-Bdairi *et al.* (*30*), Islam and Hernandez (12)
	Number of occupants	Lemp *et al.* (*6*), Chen *et al.* (*20*), Islam and Hernandez (12)
	Tailgating	Kotikalapudi and Dissanayake (*14*), Trimble *et al.* (*28*)
	Laws and features (seat belt law, DUI law, airbags)	Charbotel *et al.* (*5*),), Islam and Hernandez (*10*), Eustace *et al.* (*19*), Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Taylor *et al.* (*25*)
Vehicle characteristics	Vehicle types	Charbotel *et al.* (*5*), Uddin *et al.* (*9*), Balakrishnan *et al.* (*21*), Qin *et al.* (*23*), Taylor *et al.* (*25*), Yang *et al.* (*32*)
	Malfunction in braking system (brake failure, loss of control)	DOT (*4*), Hickman *et al.* (*7*), Kotikalapudi and Dissanayake (*14*), Trimble *et al.* (*28*), Al-Bdairi *et al.* (*30*)
	Vehicle design elements (front and rear overhang, width, weight, length, GVWR, trailing unit)	Lemp *et al.* (*6*), Koupaenejad (*8*), Zheng *et al.* (*13*), Kotikalapudi and Dissanayake (*14*), Park and Pierce (*29*), Islam and Hernandez *(12)*
	Number of trucks or other vehicles (truck percentages, AADT, vehicles involved in the crash)	Lemp *et al.* (*6*), Islam and Hernandez (*10*), Offei *et al.* (*18*), Chen *et al.* (*20*), Qin *et al.* (*23*), Wang *et al.* (*24*), Taylor *et al.* (*25*), Al-Bdairi *et al.* (*30*), Yang *et al.* (*32*)
Environmental, Roadway and Crash characteristics	Roadway condition (classification, terrain, visibility of markings, surface condition)	DOT (*4*), Charbotel *et al.* (*5*), Lemp *et al.* (*6*), Hickman *et al.* (*7*), Uddin *et al.* (*9*), Offei *et al.* (*18*), Eustace *et al.* (*19*), Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Qin *et al.* (*23*), Taylor *et al.* (*25*), Ullman and Iragavarapu (*26*), Park and Pierce (*29*), Al-Bdairi *et al.* (*30*), Fitzsimmons *et al.* (*31*), Yang *et al.* (*32*)
	Adverse weather effect	DOT (*4*), Lemp *et al.* (*6*), Hickman *et al.* (*7*), Kotikalapudi and Dissanayake (*14*), Offei *et al.* (*18*), Eustace *et al.* (*19*), Taylor *et al.* (*25*), Fitzsimmons *et al.* (*31*), Yang *et al.* (*32),* Islam and Hernandez *(12)*
	Interruptions in traffic flow (intersection, previous crash, work zone, peak hour congestion)	DOT (*4*), Hickman *et al.* (*7*), Amarasingha and Dissanayake (*15*), Dong *et al.* (*17*), Eustace *et al.* (*19*), Ullman and Iragavarapu (*26*), Park and Pierce (*29*)
	Roadway design elements (curvature, grade, width, median)	Islam and Hernandez (*10*), Amarasingha and Dissanayake (*15*), Dong *et al.* (*17*), Eustace *et al.* (*19*),Chen *et al.* (*20*), Qin *et al.* (*23*), Wang *et al.* (*24*), Taylor *et al.* (*25*), Park and Pierce (*29*), Al-Bdairi *et al.* (*30*), Fitzsimmons *et al.* (*31*), Islam and Hernandez *(11, 12)*
	Lighting condition	Al-Bdairi *et al.* (*2*), Koupaenejad (*8*), Uddin *et al.* (*9*), Islam and Hernandez (*10*), Eustace *et al.* (*19*), Chen *et al.* (*20*), Taylor *et al.* (*25*), Ullman and Iragavarapu (*26*), Pahukula (*27*), Fitzsimmons *et al.* (*31*), Yang *et al.* (*32*)
	Divided/undivided	Koupaenejad (*8*), Qin *et al.* (*23*), Taylor *et al.* (*25*), Al-Bdairi *et al.* (*30*), Yang *et al.* (*32*)
	Posted Speed Limit	Eustace *et al.* (*19*), Balakrishnan *et al.* (*21*), Qin *et al.* (*23*), Taylor *et al.* (*25*), Trimble *et al.* (*28*), Fitzsimmons *et al.* (*31*)
	Time of day	Charbotel *et al.* (*5*), Uddin *et al.* (*9*), Islam and Hernandez (*10*,*11*,*12*), Zheng *et al.* (*13*), Eustace *et al.* (*19*), Balakrishnan *et al.* (*21*), Ullman and Iragavarapu (*26*), Pahukula (*27*), Fitzsimmons *et al.* (*31*), Yang *et al.* (*32*)
	Crash type (head on, t-collision, angle and rear-end crashes, roll over)	Charbotel *et al.* (*5*), Hickman *et al.* (*7*), Koupaenejad (*8*), Uddin *et al.* (*9*), Islam and Hernandez (*10*,*11*), Zheng *et al.* (*13*), Eustace *et al.* (*19*),Chen *et al.* (*20*), Balakrishnan *et al.* (*21*), Taylor *et al.* (*25*),

		Al-Bdairi *et al.* (*30*), Fitzsimmons *et al.* (*31*), Yang *et al.* (*32*), Islam and Hernandez *(11, 12)*
	Miscellaneous (Truck company attributes)	Zheng *et al.* (*13*)

METHODOLOGY

Taxicab Correspondence Analysis (TCA)

In a series of studies *(41-43)*, Choulakian introduced the method Taxicab Correspondence Analysis (TCA). Based on Choulakian's theory, the following is a brief description of TCA:

Unlike correspondence analysis (CA), which is based on Euclidean distance, TCA uses the Manhattan, City Block, or Taxicab distance. Let, $X, Y, and\ \boldsymbol{v}$ be such that $X = (x_1, x_2, \ldots\ldots, x_n)$ and $Y = (y_1, y_2, \ldots\ldots, y_n)$ are the two components of a vector $\boldsymbol{v} = (v_1, v_2, \ldots\ldots, v_n)$ in a 2-D space. From these definitions, the following distances can be calculated *(40):*

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

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

Both CA and TCA are based on the singular value decomposition (SVD). This concept includes the decomposition of a real matrix A to $M\Lambda^{1/2}N'$, where Λ is the diagonal matrix of the real non-negative eigenvalues of AA', M is the orthogonal matrix of the corresponding eigenvectors, and N is the matrix of eigenvectors of $A'A$ (with constraints $M'M = I\ and\ N'N = I$). The SVD theory relates to the reconstruction formula of a k-rank matrix, written as:

$$a_{ij} = \sum_{i=1}^{k} \sqrt{\lambda_\alpha}\, m_{i\alpha} n_{i\alpha} \quad (3)$$

The SVD solution is developed via a recursive optimization process in the TCA framework. To solve the equivalent optimization problem, one must locate the first vectors m_1 and n_1 which are principal components of A.

$$max\ \|Am\|_2 \text{ subject to } \|m\|_2 = 1$$
$$max\ \|A'n\|_2 \text{ subject to } \|n\|_2 = 1$$

The solution provides the largest singular value of A.

$$\lambda_1 = \max_m \frac{\|Am\|_2}{\|m\|_2} = \max_n \frac{\|A'n\|_2}{\|n\|_2} = \max_{m,n} \frac{n'Am}{\|m\|_2\|n\|_2} \quad (4)$$

The reconstruction can be written as:

$$A = \sum_{i=1}^{k} \lambda_\alpha n_\alpha m'_\alpha \text{ [where: } \sum_\alpha \lambda_\alpha^2 = Tr(A'A) \quad (5)$$

The TCA framework also includes a distance matrix known as L_∞ matrix. The aim is to minimize the rank of T (and consequently of N) without displacing relevant information. To formalize the solution, an appropriate reduced rank matrix $\hat{T}$ is considered that best approximates T in the sense of the weighted least squares, that minimizes the residuals $\boldsymbol{R}$ that can be expressed as *(41):*

$$\boldsymbol{R} = n \sum_{i=1}^{r} \sum_{j=1}^{l} \frac{(t_{ij} - \hat{t}_{ij})^2}{t_{i.}t_{.j}} = n\, trace\left(D_l^{-1/2}(T - \hat{T})' D_r^{-1}(T - \hat{T}) D_l^{-1/2}\right) \tag{6}$$

TCA is defined as the Taxicab SVD of the data table $D = T - rl'$, considering the profiles of the table, $R = D_r^{-1}D$ for the rows and $L = D_l^{-1}D$ for each of the columns. The solution is recursive at each step by considering the residuals from the previous factors that differentiates it from CA. The reconstruction formula can be formulated as such:

$$T = p_r p_c' + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} B_\alpha C_\alpha' \tag{7}$$

After the final transformation has been conducted, it can be written as:

$$n_{ij} = n r_i l_j (1 + \sum_{\alpha=2}^{k} \frac{1}{\lambda_\alpha} b_{i\alpha} c_{j\alpha}) \tag{8}$$

DATA DESCRIPTION

A large truck is any medium or heavy truck, excluding buses and motor homes, with a gross vehicle weight rating (GVWR) greater than 10,000 pounds. Table 2 displays the description of the vehicles that are considered as large trucks in the FARS database *(47).*

Table 2 Large Truck Codes in the FARS

Code	**Description**
60	Step Van
61	Single-Unit Straight Truck (10,000 lbs < GVWR <= 19,500 lbs)
62	Single-Unit Straight Truck (19,500 lbs < GVWR <= 26,000 lbs)
63	Single-Unit Straight Truck (GVWR > 26,000 lbs)
64	Single-Unit Straight Truck
66	Truck/Tractor (Cab Only, or with Any Number of Trailing Units: Any Weight)
67	Medium/Heavy Pickup (GVWR > 10,000 lbs)

GVWR=Gross Vehicle Weight Rating

In this study, six years (2010-2015) of large truck fatal crash data were obtained from FARS. The crash data file, vehicle data file, and person data file are selected for this study. Preliminary data exploration was conducted at the beginning to examine the significant factors that may contribute to crash occurrence. After the preliminary analysis, this study excluded irrelevantly, and other redundant variables from the raw data before applying TCA. Figure 1 shows the flow chart of the data preparation task.

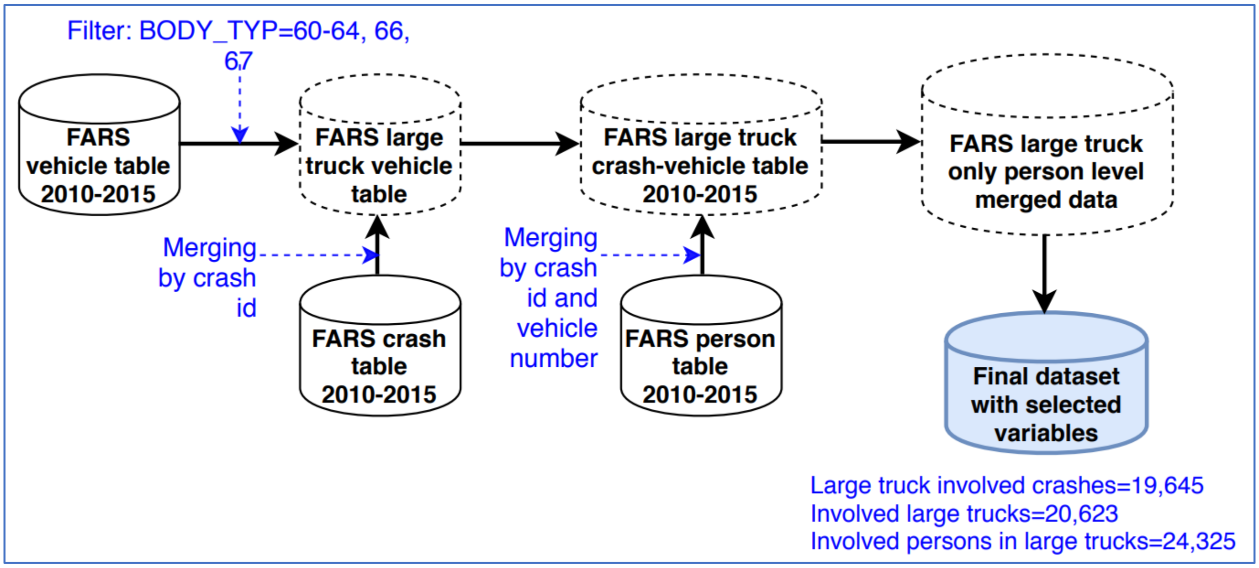


Figure 1 Flowchart of the data preparation.

There are 14 key variables in the final dataset; Table 3 displays the proportional distribution of these variables. Roadway functional class accounts for the roadway classification for each crash occurrence. From the percentage distribution, over 50 percent of the fatal crashes occurred under rural environment, which agrees with the study conducted by Chen *et al.* (*20*). They investigated the key factors affecting large truck-related crashes; they found significant statistical evidence that rural areas are more crash-prone for large trucks. They also found that, based on intersection type, approximately 75 percent of all crashes occurred on roadway segments. Furthermore, the proportions of attributes in roadway alignment show that approximately 83 percent of all roadway crashes occur on straight roadway segments. These large representations of crash statistics on roadway segments are in line with the findings by Dong *et al.* (*17*) where the authors investigated the effect of geometric design features and found a strong association of longer straight segment lengths to crash occurrences. The proportion distribution of several of the variables are also in line with other studies such as lighting condition (*2*), roadway gradient (*17*), posted speed limit (*19*), time of day (*27*), weather condition (*25*), number of vehicles involved (*6*), and types of collision (*9, 21*). Three person-level variables (previous accident record of the drivers, driver's past license suspension record, and impaired driving) are associated with the driving patterns of the large truck drivers associated in fatal crashes. However, association with prior crashes has high proportions compared to the other two traits.

Table 3 Descriptive Statistics of Key Variables

Attributes	Perc.	Attributes	Perc.
ROAD_FNC (FC)*		**VSPD_LIM (Posted Speed Limit or PSL)**	
Rural Interstate	29.7	0-25 mph	2.0
Rural Principal Arterial	10.0	30-40 mph	10.3
Rural Minor Arterial	8.4	40-60 mph	49.3
Rural Collector	4.1	60-70 mph	29.5
Rural Unknown	0.2	> 70 mph	8.4
Urban Interstate	12.6	Unknown	0.6

Urban Principal Arterial	8.7	**HOUR (Hour)**	
Urban Minor Arterial	3.9	12am- 6am	20.3
Urban Collector	1.3	7am-12pm	31.5
Urban Local	3.1	1pm- 6pm	31.6
Urban Unknown	0.4	7pm- 12am	14.0
Unknown	17.6	Unknown	2.5
TYP_INT (Int.)		**WEATHER (Weather)**	
Segment	74.9	Clear	70.7
4-way Intersection	17.2	Rain	7.4
T-Intersection	7.1	Others	21.9
Others	0.9	**VE_TOTAL (Inv.)**	
VTRAFWAY (Trafficway)		Single	18.6
Two-Way Undiv.	49.1	Two	60.6
Two-Way Div. Barier	22.0	Multi	20.8
Two-Way Div. Unprotected	22.6	**MAN_COLL (Coll.)**	
Others	6.3	Not with Other Veh.	25.8
VALIGN (Align.)		Angle	30.2
Straight	82.8	Front-to-Rear	20.9
Curve Left	8.2	Front-to-Front	14.4
Curve Right	6.1	Sideswipe (Oppo.)	3.8
Others	3.0	Sideswipe (Same)	3.4
LGT_COND (Light)		Others	1.5
Daylight	62.1	**PREV_ACC (Prev_Crash)**	
Dark -Not Lighted	24.0	No	76.1
Dark -Lighted	9.3	Yes	23.9
Others	4.7	**PREV_SUS (Prev_Sus)**	
VPROFILE (V_Profile)		**No**	88.4
Level	69.9	Yes	11.6
Grade, Unknown Slope	13.2	**DR_DRINK (Dr_Alc)**	
Downhill	5.6	No	97.6
Uphill	4.6	Yes	2.5
Others	6.7		

** Variable code used in FARS (variable code used in this study)*

RESULTS AND FINDINGS

The TCA method compiles the key components of a complex dataset by mapping the multivariate data on two-dimensional or three-dimensional displays. Ultimately, the goal is to produce a thorough biplot. The readability and interpretability of the map is highly important in most cases. The biplot displays the key attributes of the selected variables identified by their labels; for example, the selected variable could be roadway type, and rural two-lane roadways with a barrier would be displayed and identified by their labels. In the presence of a large set of attributes, the biplot can become cluttered and the labels might not be easy to read.

An important characteristic of TCA and CA is that columns (or rows) with identical profiles (conditional probabilities) possess the same factor scores. An important benefit of TCA compared to CA is that it directly acts on the correspondence matrix P without calculating a dissimilarity (or similarity) measure between the rows or columns; thus, it remains closer to the original data. Furthermore, TCA performs better in the presence of missing data.

The percentage of variance explained by both axes is around 52 percent (axis 1 explains 28.16 percent of the variance, and axis 2 explains 23.75 percent of variance). The locations of the variable categories indicate their association patterns. Figure 2 shows the complete TCA plot. As the plot is very cluttered with the presence of all attributes on the same plot, four separate plots are recreated (see Figure 3) for better visualization and interpretation. The solid parabolic shapes represent a cluster with a distinct association pattern. The clusters with both obvious and trivial associations are shown by dotted parabolic shapes.

Cluster 1a (Urban collector or minor arterial, Intersection= T-intersection or 4-way intersection, Posted speed limit= 30-40 mph)

This cluster is visible in the first quadrant of the TCA plot (see Figure 2). While several studies have invested research efforts in identifying factors corresponding to large truck-related crashes in a rural environment, only a few have investigated crashes under urban environment. Charbotel *et al.* (*5*) explored the differences of crash scenarios in both rural and urban areas in large truck-related crashes and found that the injury severity in truck-involved crashes has been increasing significantly, especially in urban collectors or arterials. Pahukula *et al.* (*27*) found a close association between factors of intersection and posted speed limit as contributing to crashes occurring in urban areas. Cluster 1b (shown in a dotted parabola in Figure 3b) in quadrant 1 shows a cluster of several attributes with trivial associations.

Cluster 2a (Lighting=Dark lighted or dark not lighted, Roadway= Two-way divided unprotected, Collision=Front to rear)

The first dominant cluster identified in the second quadrant indicates that the dark, not lighted condition in the two-way unprotected-divided roadway is prone to rear-end crashes involving large trucks (as shown in Figure 2). This cluster is also in the close neighborhood of interstate and high posted speed limit (60 mph and above). The association of factors discerned in this cluster is in line with multiple studies found in the literature. Al-Bdairi *et al.* (*2*) found that speeding while driving in darker conditions can significantly increase the chances of more severe crashes on rural interstate highways. Trimble *et al.* (*28*) investigated the factors contributing to rear-end crashes involving large trucks and concluded that speeding including acceleration and deceleration in anticipation of drivers' attempted avoidance maneuvers is a significant factor of crashes. In a comprehensive study, Koupaenejad (*8*) found an association between no divided medians and younger male drivers leading to severe crashes on interstate highways.

Cluster 3a (Impaired driver, Single truck or sideswipe same direction, and Single or multiple vehicle involvement)

The attributes in this cluster, as shown in Figure 2, illustrates close agreement with the findings by Chen *et al.* (*20*) where the authors showed that driver impairment in terms of drug or alcohol consumption, and a number of vehicles involved in the crash are significant factors in severe injury or fatal crashes. The findings by Al-Bdairi *et al.* (*2*) suggest that single-vehicle crashes that involve running off-road under a dark condition with a fixed roadside object can lead to fatality in large truck-related crashes. Uddin and Huynh (*9*) findings show contradiction to these studies where the authors claim that actions corresponding to single-vehicle crashes involving large trucks such as sideswipe, run-off-road, and hitting stationary object are associated with decreased probability of injury severity. Additionally, the likelihood of fatal or severe injury crashes increases for single-vehicle involvement and decreases if there are multiple vehicles involved in the crash (*9*, *30*).

Cluster 3b (Segment, Weather= Rain or others, Uphill, Previous crash conducted by the truck driver=Yes and unknown functional class)

The association is Cluster 3b, as shown in Figure 3c, suggests that adverse weather conditions, such as raining and dust on an uphill gradient significantly increase the probability of fatal crashes involving large trucks, specifically with drivers having past crash experience. The relationship deciphered in this cluster can be justified by the findings of Dong *et al.* (*17*), where the authors found that an uphill or downhill gradient significantly increases the chance of a severe crash under adverse weather condition. The inference of these geometric and weather conditions with that of the driver's previous crash experience is explainable by the plausible assumption that the drivers with past crash experience may have issues with safe driving behavior. Hence, this factor is well expected to be a significant one in contributing to a fatal crash.

Cluster 4a (Functional class= Rural principal or minor arterial, collector & urban local, Roadway= Two-way undivided, Collision= Sideswipe opposite, low-posted speed limit, and Alignment=Others)

As illustrated in Figure 2, the attributes grouped under this cluster agree with the findings listed in Cluster 1a and Cluster 1b, which identified two sets of significant factors related to large truck fatal crashes in rural and urban areas. Although a greater number of factors are found in one single cloud in this cluster, a distinction can be drawn as to which ones are related to the land use context (rural vs. urban). This is supported by a study conducted by Islam and Hernandez *(10).* For example, two-way undivided roadways are more likely to be found in the rural principal or minor arterials, whereas, the low-speed limit is likely to be associated with urban collector and local roadways. Cluster 4b is not referencing a strong cluster inference (shown in Figure 3d).

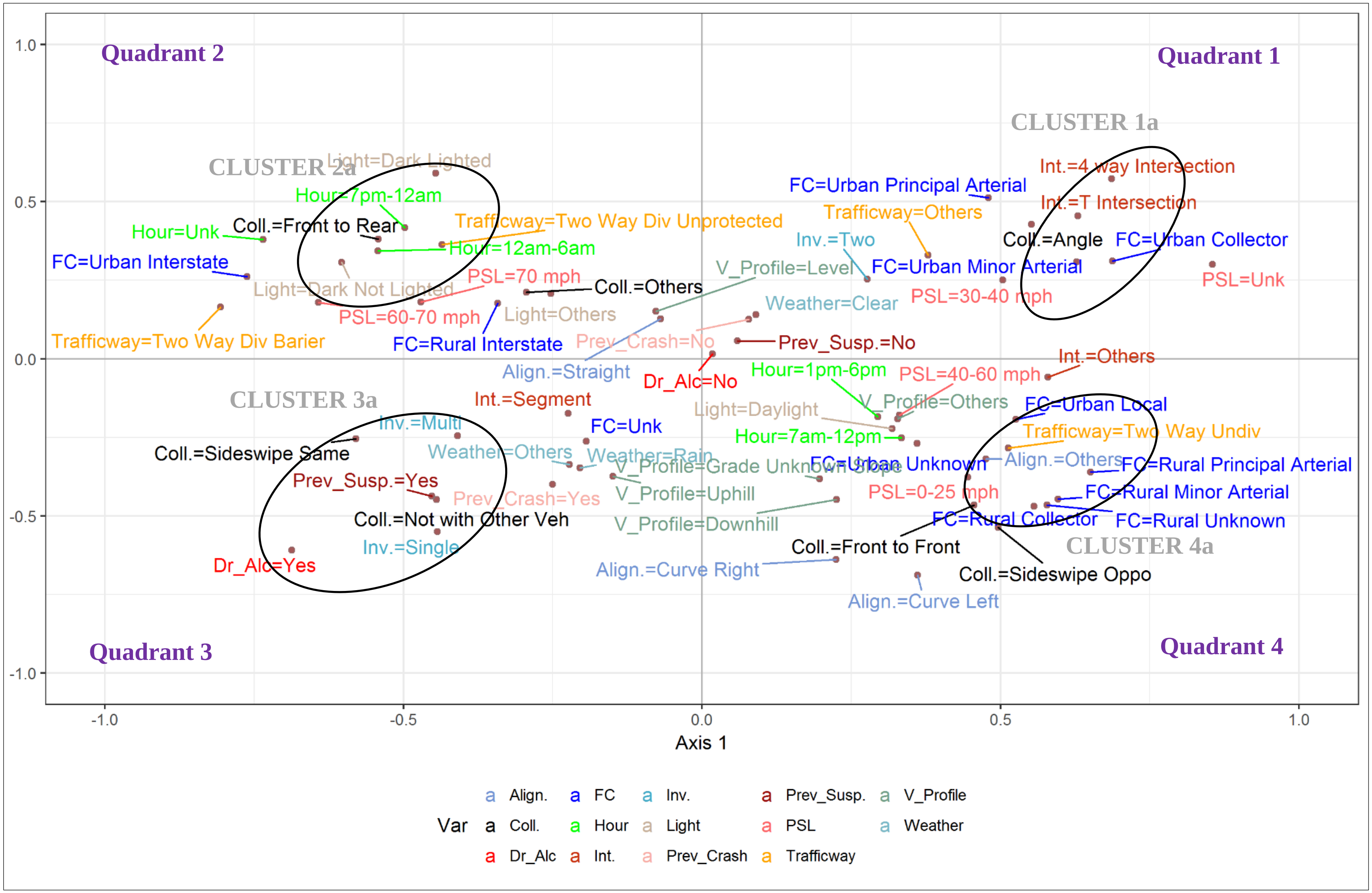
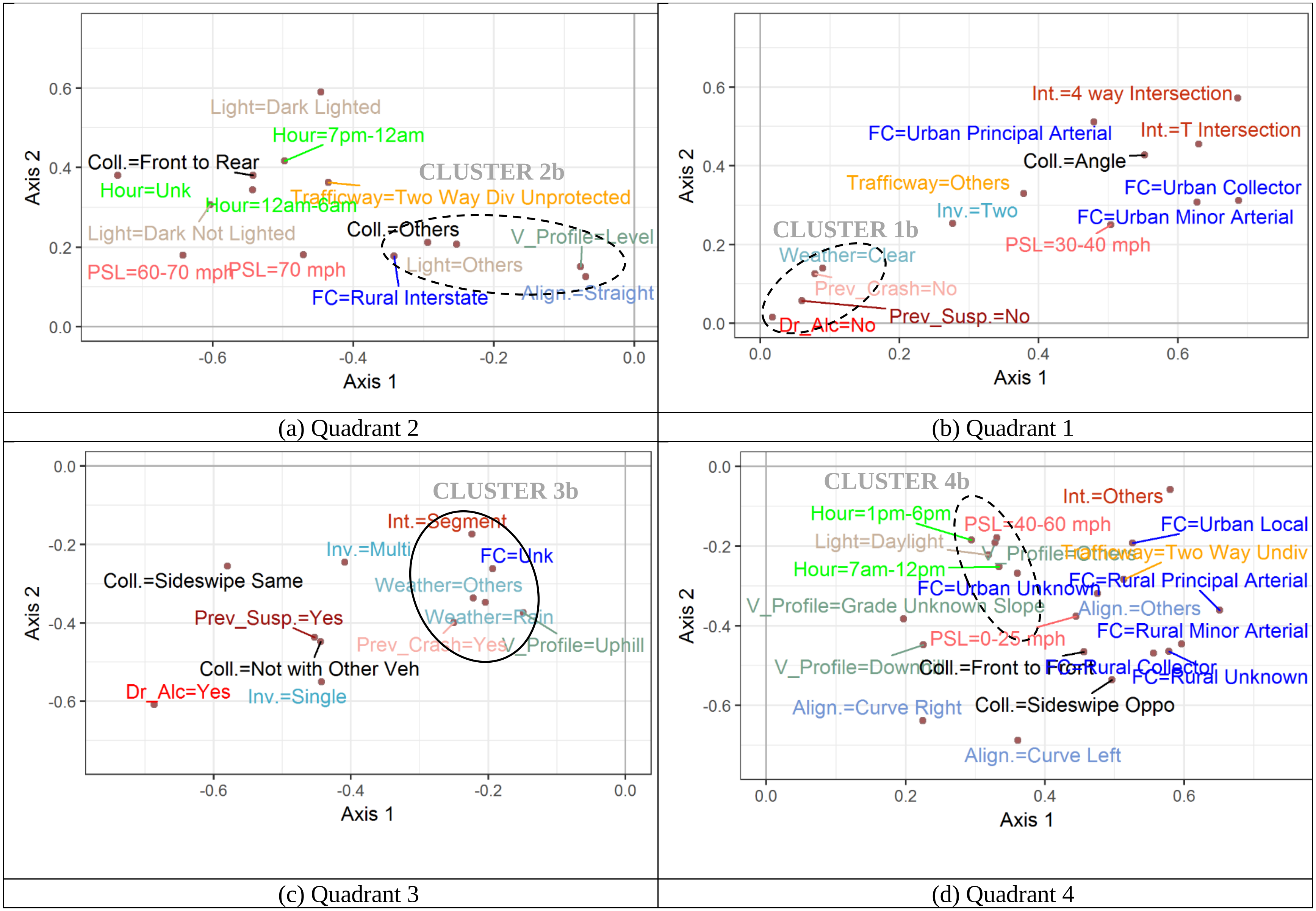


Figure 2 TCA plot with two axes and quadrants for distinct association pattern.



1 **Figure 3 TCA plot by quadrants for closer association pattern.**

The final dataset contains 26,275 individual data points representing the personal information of the large truck drivers and occupants (LTDO) involved in crashes. With the TCA method, LTDO locations can be divided into 69 distinct clusters. Figure 4 provides an illustration of the distribution of each LTDO (mapping the individuals in the biplot instead of mapping the variable attributes) on both axes. The figure shows 69 distinctive clusters if one clusters the points vertically based on axis 1 (each cluster represents each vertical line on the x-axis) and there are clusters located on both the positive and negative side of the x-axis. Points that are larger in size represent a greater count of LTDO with the same co-ordinates.

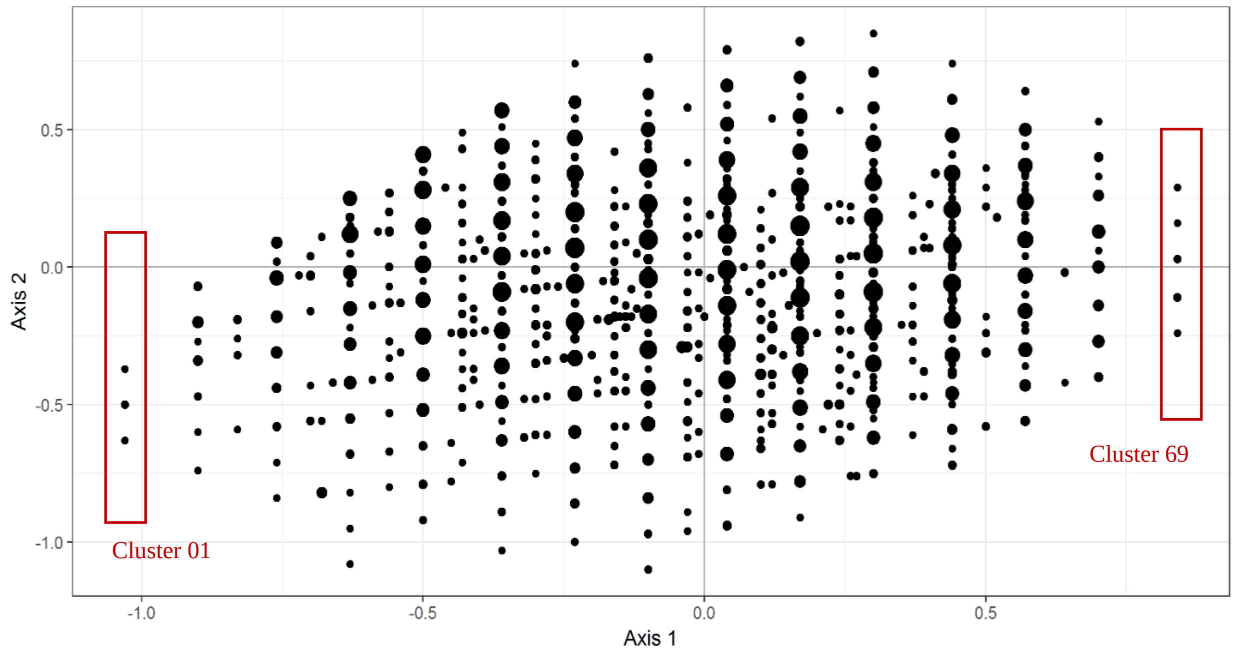


Figure 4 Clusters of large truck-involved people.

Four general crash prevalence conditions were considered for further analysis. These conditions are LTDO involvement in previous crashes (yes/no), LTDO with a record of previous suspension history (yes/no), intoxication (yes/no) of LTDO, and single truck crash (yes/no).

Table 4 lists the importance of the clusters by computing the log-odds ratio of the crash prevalence conditions with respect to the marginal distribution. The interpretation of LOR (X=x) is as follows:

- LOR (X=x) = 0.00 indicates that the proportion of category A in cluster x is equal to the proportion of category B in the sample. For example, the LOR value for LTDO in previous crashes (yes vs. no) is 0 for Cluster07. Cluster07 is associated with 29 LTDO.
- LOR (X=x) > 0.00 indicates that the proportion of category A in cluster x is greater than the proportion of category B in the sample. For example, the LOR value for LTDO in previous crashes (yes vs. no) is 1.77 for Cluster07. This cluster has 275 LTDO. The LOR value indicates that this cluster is positively associated with LTDO who have previous crash histories.
- LOR (X=x) < 0.00 indicates that the proportion of category A in cluster x is smaller than the proportion of category B in the sample. For example, the LOR value for LTDO in previous crashes is –0.59 for Cluster23. It indicates that this cluster is negatively associated with LTDO who have previous crash histories.

Some of the clusters have zero LOR values for all the crash prevalence scenarios. These clusters involved only 0.75 percent of all LTDOs.

Table 4 Log Odds Ratio of Four Crash Prevalence Scenarios

Row Labels	Count	Prev_Crash (Y_vs_N)	Prev_Suspen (Y_vs_N)	Alc (Y_vs_N)	SingleVeh (Y_vs_N)
Cluster01	7	2.07	3.82	3.97	3.27
Cluster02	72	2.67	2.72	1.61	1.42
Cluster03	16	2.26	3.97	1.74	2.58
Cluster04	275	1.77	1.82	1.51	1.59
Cluster05	3	0.00	0.00	0.00	0.00
Cluster06	30	0.47	1.34	2.30	2.49
Cluster07	29	0.00	0.00	5.03	3.64
Cluster08	5	0.00	0.00	0.00	0.00
Cluster09	764	1.14	1.09	0.82	1.01
Cluster10	6	0.00	2.03	0.00	0.00
Cluster11	5	0.00	0.00	0.00	0.00
Cluster12	72	0.47	1.69	0.13	1.93
Cluster13	9	0.00	2.72	0.00	0.00
Cluster14	1623	0.61	0.80	0.60	0.66
Cluster15	4	0.00	0.00	0.00	0.00
Cluster16	9	0.00	2.72	4.38	0.00
Cluster17	86	-0.17	0.70	1.41	1.43
Cluster18	18	0.47	0.42	2.08	2.17
Cluster19	8	1.16	2.03	0.00	1.48
Cluster20	11	0.98	0.00	0.00	1.29
Cluster21	2405	0.30	0.39	0.37	0.35
Cluster22	15	0.75	2.43	2.30	1.88
Cluster23	108	-0.59	-0.25	0.85	1.18
Cluster24	32	-0.31	-0.24	3.43	1.73
Cluster25	5	0.00	0.00	0.00	0.00
Cluster26	9	0.00	0.00	0.00	0.00
Cluster27	3006	0.15	0.20	0.16	0.16
Cluster28	4	0.00	0.00	0.00	0.00
Cluster29	12	0.00	0.00	2.59	0.00
Cluster30	5	0.00	0.00	0.00	0.00
Cluster31	18	0.00	0.00	0.00	0.00
Cluster32	106	-0.72	-0.03	1.80	1.59
Cluster33	5	0.00	0.00	0.00	0.00
Cluster34	36	0.06	0.42	1.29	2.17
Cluster35	8	1.16	0.00	0.00	1.48
Cluster36	12	0.00	0.00	0.00	0.00
Cluster37	3124	-0.05	-0.03	0.06	0.14
Cluster38	37	0.00	0.00	0.00	0.00
Cluster39	116	-0.41	0.20	1.32	1.55
Cluster40	27	-0.09	0.00	1.61	1.70
Cluster41	4	0.00	0.00	0.00	0.00
Cluster42	11	0.00	0.00	0.00	0.00
Cluster43	3186	0.05	-0.10	-0.61	0.02
Cluster44	4	0.00	0.00	0.00	0.00
Cluster45	3	0.00	0.00	0.00	0.00
Cluster46	134	-0.58	-0.12	-0.50	1.69

Row Labels	Count	Prev_Crash (Y_vs_N)	Prev_Suspen (Y_vs_N)	Alc (Y_vs_N)	SingleVeh (Y_vs_N)
Cluster47	89	0.68	-1.33	0.00	1.72
Cluster48	8	1.16	0.00	0.00	0.00
Cluster49	6	0.00	0.00	0.00	0.00
Cluster50	3328	-0.05	-0.36	-0.86	-0.38
Cluster51	4	0.00	0.00	0.00	0.00
Cluster52	3	0.00	0.00	0.00	0.00
Cluster53	19	0.00	0.00	0.00	2.51
Cluster54	104	-0.70	-0.21	-0.25	0.38
Cluster55	24	0.06	0.08	0.00	0.38
Cluster56	12	0.00	0.00	0.00	2.17
Cluster57	3220	-0.34	-0.79	-1.21	-1.05
Cluster58	3	0.00	0.00	0.00	0.00
Cluster59	34	-1.61	-0.74	1.67	-0.54
Cluster60	15	-0.23	0.00	0.00	1.07
Cluster61	8	0.00	0.00	0.00	0.00
Cluster62	10	0.00	0.00	0.00	0.00
Cluster63	2317	-0.77	-1.13	-1.86	-2.49
Cluster64	34	-1.61	0.00	0.00	-1.30
Cluster65	6	0.00	0.00	0.00	0.00
Cluster66	1244	-1.66	-2.04	-1.83	-3.56
Cluster67	8	0.00	0.00	0.00	0.00
Cluster68	279	-3.77	-2.49	-1.94	0.00
Cluster69	16	0.00	0.00	0.00	0.00

Note: Prev_Crash(Y_vs_N) indicates previous crash experiences in last 5 years??

Table 4 shows that 13 clusters have LOR values greater than zero for all prevalence groups. These clusters represent 8,339 LTDOs. Out of these 8,339 LTDOs, the people having these prevalence traits are over-represented. Future research is needed to explore the driver and occupant traits in these groups.

CONCLUSIONS

The U.S. economy benefits immensely from the effective movement of freight. An unprecedented peak in freight-hauling was recorded in 2015 due to an economic uprising following the recession from 2007-2009. In the U.S., the amount of freight transported on a daily basis averaged 49.3 million tons and was valued at nearly $53 billion in 2015 *(44. 45)*. In this study, TCA, a robust variant of CA, was applied to six years of fatal crashes obtained from FARS. This technique allows a powerful interpretation of the complex association of factors in multivariate events using two-dimensional maps. The method can handle complex datasets with many outliers, which fits the FARS dataset perfectly as it only screened for large truck crashes. The findings of this study are as follows:

- Urban intersections are the setting for a disproportionate number of large truck fatal crashes.
- There is a strong association between two-way roadways with an unprotected median and large truck fatal crashes.
- Two distinct clusters (impaired driver's involvement in single-vehicle crashes, and drivers with the past crash record being involved in inclement weather crashes) indicate human error associated patterns in large truck fatal crashes.

- Driving in non-daytime hours is associated with a high number of truck-involved crashes.
- Individual-level TCA analysis identified 69 distinct clouds based on four prevalence driving behaviors. A total of 13 clusters show LOR values greater than zero for all prevalence behavioral groups. These clusters represent 8,339 LTDOs. Out of these 8,339 LTDOs, the people that possess these prevalence traits are over-represented.

The CA approaches focus on the associations between the covariates of crashes rather than the associations between each covariate with the frequency (or odds) of crashes. However, TCA determines the strength of the association of a variable with the frequency of crashes, as shown in the LOR analysis. Because of this, TCA results are more easily interpretable than CA results.

The TCA method, currently being more applied to a large database such as FARS, helps to understand the crash patterns and also associate the contributing factors to fatal crashes. With the growing body of literature in large truck safety, this study explores TCA in uncovering the deeper relations of factors leading to fatalities. This study is not without limitations. Newer FARS data have been released in the recent years. The current study is limited to 2010-2016 FARS data. Additionally, this analysis is also limited to a broader group by defining it as large truck. There is a need for separating out large, medium, and very large trucks, which is not currently done in the current study. The current limitations can be improved in future studies.

DISCLAIMER

The contents of this paper reflect the views of the authors and not the official views or policies of the Louisiana Department of Transportation and Development (LADOTD).

ACKNOWLEDGMENT

The authors appreciate the assistance provided by the students on this manuscript preparation: Bita Maraghehpour and Magdalena Theel.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das, and Anandi Dutta; analysis and interpretation of results: Subasish Das, Mouyid Islam, Anandi Dutta, and Tahmida Hossain Shimu; draft manuscript preparation: Subasish Das, Mouyid Islam, Anandi Dutta, and Tahmida Hossain Shimu. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Large trucks: 2016 data. Publication DOT HS 812 497. NHTSA, U.S. Department of Transportation, Washington DC, 2016.
2. Al-Bdairi, N. S. S., S. Hernandez, and J. Anderson. Contributing factors to run-off-road crashes involving large trucks under lighted and dark conditions. *Journal of Transportation Engineering, Part A: Systems*, 2018, 144.
3. Knipling, R. R., J. S. Hickman, and G. Bergoffen. *Synthesis 1: Effective commercial truck and bus safety management techniques*. Transportation Research Board, Washington, DC, 2003.
4. Data.gov. (n.d.). Large Truck Crash Causation Data Extraction, 2006. Retrieved from: http://explore.data.gov/Transportation/Large truck-Crash-Causation-Study-LTCCS-/5zg4-dnsb.

5. Charbotel, B., J. L. Martin, B. Gadegbeku, and M. Chiron. Severity factors for truck drivers' injuries. *American Journal of Epidemiology*, 2003. 158: 753-759.
6. Lemp, J. D., K. M. Kockelman, and A. Unnikrishnan. Analysis of large truck crash severity using heteroskedastic ordered probit models. *Accident Analysis and Prevention*, 2011. 43: 370-380.
7. Hickman, J. S., R. J. Hanowski, and J. Bocanegra. A synthetic approach to compare the large truck crash causation study and naturalistic driving data. *Accident Analysis and Prevention,* 2018. 112: 11–14.
8. Koupaenejad, A. Statistical modeling and analysis of injury severity sustained by occupants of passenger vehicles involved in crashes with large trucks. *UNLV Theses, Dissertations, Professional Papers, and Capstones*, 2010, 815.
9. Uddin, M., and N. Huynh. Factors influencing injury severity of crashes involving HAZMAT trucks. *International Journal of Transportation Science and Technology*, 2018. 7: 1–9.
10. Islam, M. B., and S. Hernandez. Modeling injury outcomes of crashes involving heavy vehicles on Texas highways. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2388: 28-36.
11. Islam, M. B., and S. Hernandez. Large truck involved crashes: an exploratory injury severity analysis. *Journal of Transportation Engineering*, 2013. 139: 596-604.
12. Qin, X., K. Wang, and C. E. Cutler. Logistic regression models of the safety of large trucks. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2392: 1-10
13. Zheng, Z., P. Lu, and B. Lantz. Commercial truck crash injury severity analysis using gradient boosting data mining model. *Journal of Safety Research*, 2018. 65: 115–124.
14. Kotikalapudi S., and S. Dissanayake. Characteristics and contributory causes associated with large truck crashes and countermeasures for mitigation. Presentation at the 92nd Annual Meeting of the Transportation Research Board, Washington DC, 2013.
15. Amarasingha, N., and S. Dissanayake. Modeling frequency of truck crashes on limited-access highways. *Journal of the transportation research forum*, 2013. 52:123-139.
16. Dissanayake, S. and N. Amarasingha. Effects of geometric design features on truck crashes on limited-access highways. *Final Reports & Technical Briefs from Mid-America Transportation Center*, 2012, 66.
17. Dong, C., M. L. Burton, S. S. Nambisan, and S. Jian. Effects of roadway geometric design features on the occurrences of truck-related crashes. Presentation at the 95nd Annual Meeting of the Transportation Research Board, Washington DC, 2016.
18. Offei, E., R. K. Young and S. S. Wulff. Crash rates on interstate segments with very large truck percentages. Presentation at the 92nd Annual Meeting of the Transportation Research Board, Washington DC, 2013.
19. Eustace, D., T. Alqahtani and P. W. Hovey. Classification tree modelling of factors impacting severity of truck-related crashes in Ohio. Presentation at the 97nd Annual Meeting of the Transportation Research Board, Washington DC, 2018.
20. Chen, C., G. Zhang, Z. Tian, S.M. Bogus, and Y. Yang. Hierarchical Bayesian random intercept model-based cross-level interaction decomposition for truck driver injury severity investigations. *Accident Analysis and Prevention*, 2015. 85: 186–198.

21. Balakrishnan, S., S. Moridpour, and R. Tay. Analysis of injury severity in heavy vehicles angle crashes. Presentation at the 96nd Annual Meeting of the Transportation Research Board, Washington DC, 2017.
22. Zhao, Q., T. Goodman, M. Azimi, and Y. Qi. Roadway related truck crash risk analysis: Case studies in Texas. Presentation at the 97nd Annual Meeting of the Transportation Research Board, Washington DC, 2018.
23. Qin, X., M. A. Sultana, M. Chitturi, and D. A. Noyce. Developing a truck corridor crash severity index (CSI). *Transportation Research Record: Journal of the Transportation Research Board*, 2013, 2386.
24. Wang, Z., B. Cao, W. Deng, Z. Zhang, J. J. Lu, and H. Chen. Safety evaluation of truck-related crashes at freeway diverge areas. Presentation at the 90th Annual Meeting of the Transportation Research Board, Washington DC, 2011.
25. Taylor, S. G., B. J. Russo, and E. James. A comparative analysis of factors affecting the frequency and severity of freight-involved and non-freight crashes on a major freight corridor freeway. Presentation at the 97nd Annual Meeting of the Transportation Research Board, Washington DC, 2018.
26. Ullman, G. L., and V. Iragavarapu. An analysis of large truck overrepresentation in fatal work zone crashes. Presentation at the 96nd Annual Meeting of the Transportation Research Board, Washington DC, 2017.
27. Pahukula, J., S. Hernandez, and A. Unnikrishnan. A time of day analysis of crashes involving large trucks in urban areas. *Accident Analysis and Prevention*, 2015, 75: 155–163.
28. Trimble, T. E., A. M. Flintsch, W. A. Schaudt, D. S. Bowman, R. J. Hanowski, and C. Flanigan. Rear-end crashes involving heavy trucks: a general estimates system crash database analysis. Presentation at the 90th Annual Meeting of the Transportation Research Board, Washington DC, 2011.
29. Park, L., and D. Pierce. Mapping large truck rollovers: identification and mitigation through spatial data analysis. Presentation at the 92nd Annual Meeting of the Transportation Research Board, Washington DC, 2013.
30. Al-Bdairi, N. S. S., and S. Hernandez. An empirical analysis of run-off-road injury severity crashes involving large trucks. *Accident Analysis and Prevention*, 2017. 102: 93–100.
31. Fitzsimmons, E. J., T. E. Lindheimer, S. D. Schrock, and K. Gonterwitz. Evaluation of large truck crashes at horizontal curves on two-lane rural highways in Kansas. Presentation at the 92nd Annual Meeting of the Transportation Research Board, Washington DC, 2013.
32. Yang, H., K. Ozbay, and K. Xie. Impact of truck-auto separation on crash severity. Presentation at the 94th Annual Meeting of the Transportation Research Board, Washington DC, 2015.
33. Fontaine H. A Typological Analysis of Pedestrian Accidents. Presented at the 7th workshop of ICTCT, Paris, 26-27 October, 1995.
34. Factor R, Yair G., and Mahalel, D. Who by Accident? The Social Morphology of Car Accidents. *Risk Analysis*, 2010, 30.

35. Das, S., R. Avelar, K. Dixon, and X. Sun. Investigation on the Wrong Way Driving Crash Patterns Using Multiple Correspondence Analysis. *Accident Analysis & Prevention*, 2018. 111:43–55.
36. Das, S., and X. Sun. Exploring Clusters of Contributing Factors for Single-Vehicle Fatal Crashes Through Multiple Correspondence Analysis. *Proceedings in the 93rd Transportation Research Board Annual Meeting, Washington DC, 2014.*
37. Das, S., and X. Sun. Factor Association with Multiple Correspondence Analysis in Vehicle–Pedestrian Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2519: 95–103.
38. Jalayer, M., M. Pour-Rouholamin, and H. Zhou. Wrong-Way Driving Crashes: A Multiple Correspondence Approach to Identify Contributing Factors. *Traffic Injury Prevention*, 2018. 19: 35–41.
39. Das, S., L. Minjares-Kyle, R. Henk, and L. Wu, Understanding Crash Potential Associated with Teen Driving: Survey Analysis using Multivariate Graphical Method. Journal of Safety Research, 2019, 70: 213-222.
40. Das, S., B. Brimley, T. Lindheimer, and A. Pant. Safety Impacts of Reduced Visibility in Inclement Weather. Report No. ATLAS-2017-19, 2017.
41. Choulakian, V. Graph Partitioning by Correspondence Analysis and Taxicab Correspondence Analysis. *Journal of Classification*, 2013. 30: 397-427.
42. Choulakian, V. L1-norm projection pursuit principal component analysis. *Computational Statistics & Data Analysis*, 2006. 50:1441 – 1451.
43. Choulakian, V. Taxicab correspondence analysis. *Psychometrika*, 2006. 71: 1-13.
44. U.S. Department of Transportation, Federal Motor Career Safety Administration. *Large Truck and Bus Crash Facts*, 2016.
45. U.S. Department of Transportation, Bureau of Transportation Statistics. *Freight Facts and Figures*, 2017.
46. Islam, M., and Ozkul, S. Identifying Fatality Risk Factors for the Commercial Vehicle Driver Population. Transportation Research Record, 2019.
47. Fatality Analysis Reporting System (FARS) Analytical User's Manual, 1975-2015. https://www.nber.org/fars/ftp.nhtsa.dot.gov/fars/FARS-DOC/Analytical%20User%20Guide/USERGUIDE-2015.pdf Accessed: July 2019.