Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multiyear Pattern Discovery and Policy Implications of Transit Bus involved Fatal Crashes

Tausif Islam Chowdhury

sgp98@txstate.edu

Texas State University

Sharif Ahmed Rafat

Texas State University

Subasish Das

Texas State University

Research Article

Keywords: Transit City Bus, Spatial Distribution, Cluster Correspondence Analysis, Vtree

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10028625/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Data Science for Transportation on October 15th, 2026. See the published version at https://doi.org/10.1007/s42421-026-00180-6.

Multiyear Pattern Discovery and Policy Implications of Transit Bus involved Fatal Crashes

Tausif Islam Chowdhury*, Sharif Ahmed Rafat and Subasish Das

Ingram School of Engineering, Texas State University, 601, University Drive, San Marcos, 78666, Texas, United States.

*Corresponding author(s). E-mail(s): sgp98@txstate.edu;
Contributing authors: hbx29@txstate.edu, subasish@txstate.edu;

Abstract

Fatal transit bus crashes remain a significant safety concern in urban transportation systems, with buses operating in complex, high-risk environments where interactions with pedestrians, cyclists, and other vehicles are frequent. This study seeks to identify the key factors contributing to these fatal incidents using Cluster Correspondence Analysis (CCA), applied to data from the Fatality Analysis Reporting System (FARS). By analyzing crash patterns over multiple years, the study uncovered distinct crash characteristics related to intersection-related conflicts, high-speed roadways, non-motorist involvement, and inadequate traffic control. The findings highlighted the crucial role of roadway design, including the presence of unprotected medians and the lack of traffic control devices, in increasing crash severity. In response to these patterns, the study proposes a series of policy recommendations aimed at improving road infrastructure, enhancing traffic management systems, and promoting safer driving behaviors. These policy measures are designed to reduce the risk of fatal crashes involving transit buses and to ensure safer streets for all road users. The insights provided in this study offered a comprehensive foundation for future transit safety strategies and policy development.

Keywords: Transit City Bus, Spatial Distribution; Cluster Correspondence Analysis, Vtree

1 Introduction

Transit city buses, although constituting a relatively small proportion of the overall vehicle fleet, are involved in a disproportionately high number of severe collisions due to their constant operation in dense, high-risk urban environments. These vehicles routinely navigate complex traffic conditions, interact with vulnerable road users, and make frequent stops near intersections and pedestrian crossings, factors that collectively heighten their exposure to crash risks. From 2008 to 2023, over 8,200 bus-to-person collisions were reported in the U.S., resulting in nearly 600 fatalities and over 8,000 injuries, with the majority occurring in urban settings (FTA 2023, 2021). Nationally, an average of 60,000 bus crashes occur annually, with about 14,000 involving injuries (Coletti 2025), and in 2024 alone, 13,452 bus collisions led to 171 fatalities, with a 51% injury rate among the incidents. A comprehensive safety analysis by MITRE revealed that the fatality rate in city bus crashes rose by approximately 50%, from 3.0 to 4.5 deaths per 100 million vehicle revenue miles between 2018–2019 and 2020–2023, with driver fatigue and extended shifts cited as key contributing factors (Colavito and Blassic 2025). A study observed that pedestrian-involved bus crashes near intersections often result in more severe injuries, largely due to visibility limitations and conflicts during turning movements (Kaplan and Prato 2012). Another study noted that crash risks increase when passengers board or alight in areas without designated safety infrastructure (Zheng et al. 2024). The longer shift durations and dense urban routes are associated with higher crash likelihood, especially during mid-day and evening hours (Hutter et al. 2019).

A wide array of analytical methodologies has been utilized in the literature to examine the risk factors associated with city bus collisions. Traditional approaches such as Poisson and negative binomial regression models have been frequently applied to model crash frequency and severity based on roadway design, environmental conditions, and transit operating characteristics (Strathman et al. 2010; Zheng et al. 2024). Studies found that road geometry, intersection proximity, and signal density significantly influenced crash likelihood, particularly during peak hours. To address spatial and temporal heterogeneity in urban environments, one study adopted hierarchical and mixed-effects modeling, revealing how route-specific factors like traffic volume, stop density, and pedestrian presence contributed to collision patterns (Mardi Safitri et al. 2020). Alongside these statistical methods, researchers have increasingly applied machine learning techniques for their ability to uncover nonlinear interactions and hidden structures within complex datasets. Random forests and gradient boosting have been used to identify key driver-related predictors of crash involvement, including shift duration, route familiarity, and age (Feng et al. 2016). Similarly, another study demonstrated that support vector machines (SVMs) could effectively classify injury severity outcomes with high accuracy using multi-source data including weather, driver workload, and operational timing (Zheng et al. 2024). Video analytics and sensor integration have further advanced the precision of behavior-based crash prediction models; for instance, D'Souza and Maheswari used computer vision techniques to detect driver distraction and correlate visual cues to near-miss events and safety violations (D'Souza and Maheswari 2012). Moreover, recent studies have emphasized

the use of hybrid models that integrate statistical and machine learning methods to improve generalizability and interpretability (Li et al. 2024; Zhu et al. 2023).

2 Literature Review

2.1 Comparative Safety and Emerging Data-Driven Approaches in Transit Bus Systems

Public transportation, particularly transit buses, is widely recognized for its safety advantages over private vehicle travel. One study found that car occupants experience 3.7 times higher injury risk than bus passengers, while pedestrians and cyclists face 4.1 to 5.3 times higher injury risks from car travel compared to bus routes (Morency et al. 2018). Such evidence underscores the modal safety benefits of transit systems, particularly within urban environments. Complementing this, another study demonstrated that public transport not only reduces accident exposure through congestion mitigation and improved traffic discipline but also lowers emissions per passenger-kilometer, highlighting its multifaceted public health benefits (Henezi and Winkler 2023). However, safety is not solely determined by empirical crash rates. One study emphasized that users often prioritize convenience, cost, and travel time over safety when choosing travel modes (Ibrahim et al. 2024). Nevertheless, latent preferences for safety emerge in long-term modal shift intentions, suggesting that safety awareness campaigns and policy interventions may shift user behavior over time. Public perception plays a particularly critical role in developing countries. Another study on Indonesian paratransit systems, concluded that perceived and actual safety are equally influential in shaping public transit usage, particularly when regulatory oversight is limited (Joewono and Kubota 2006).

2.2 Multifactorial and Systemic Influences on Transit Bus Crash Risk

Multiple studies have emphasized that bus crashes are rarely due to isolated factors. One study identified high crash concentrations near intersections and bus stops (Kaplan and Prato 2012), where interactions with vulnerable road users (VRUs) and turning vehicles increase complexity. Another study associates multitasking, passenger interaction, and schedule pressure with increased crash risk (D'Souza and Maheswari 2012). Singh and Kathuria (2024) further showed that nonlinear interactions among age, road curvature, and traffic volume significantly influence crash severity. Beyond individual and environmental variables, systemic and institutional factors also shape crash risk. A study in Charlotte, NC, employed generalized linear and ridge regression models to show how gender, income, and neighborhood affect perceived transit safety and wait times (Hosseini et al. 2025). A combination of latent class clustering and survival analysis revealed that lane invasions and rainy weather disproportionately impacted intercity transit crashes (Goh et al. 2014). Other studies have addressed operator health, showing that cardiovascular risks stemming from irregular schedules can result in sudden incidents (Lu and Liang 2023), while institutional policies in Norway were found to indirectly affect crash risk via pressure-induced behavioral changes

(Nævestad et al. 2022). Finally, the adoption of collision avoidance warning systems (CAWS) and automated emergency braking technologies has shown promising reductions in crash risk and severity, albeit with limited current implementation (Song et al. 2021; Yin et al. 2024).

2.3 Applications for Dimension Reduction Methods in Traffic Safety

Cluster Correspondence Analysis (CCA) and Multiple Correspondence Analysis (MCA) have been widely utilized in traffic safety research to uncover latent patterns in crash data. These methods offer a powerful framework for exploring categorical variables and identifying co-occurring risk factors in complex crash environments. For instance, a study by Das et al. (2022) analyzed seven years of motorcycle crash data in Louisiana, applying correspondence analysis to identify multiple clusters of contributing factors. Similarly, Dzinyela et al. (2025) extended this approach to other motorcycle-related crash scenarios, reinforcing the method's versatility. Further expanding the scope, a study investigated fatal crashes involving mopeds and seated motor scooters using CCA, identifying significant factors influencing injury severity (Das et al. 2023). In the pedestrian safety domain, Hossain et al. employed MCA to analyze 722 fatal and severe pedestrian crashes occurring under dark, no-streetlight conditions in Louisiana (Hossain et al. 2023). Their analysis revealed high-risk crash patterns involving specific combinations of pedestrian, driver, and roadway characteristics, such as impaired riders crashing on unlit rural highways or rear-end collisions on local roads with traffic violations.

3 Methodology

3.1 Data Description

This section includes the description of the data used in this study. Table 1 presents the distribution of fatal transit city bus crashes reported in the FARS database across four two-year periods from 2016 to 2023. Seasonal trends shifted over time, with summer-related fatal crashes peaking in 2020–2021 (30.4%), while fall emerged as the most common season in 2022–2023, accounting for 37.2% of all cases. Weekday crashes remained predominant throughout, though a gradual increase in weekend fatalities was observed, rising from 23.4% in 2016–2017 to 28.4% in 2022–2023. Night non-peak hours became increasingly common, representing 38.8% of fatal crashes in the latest period compared to 25.8% in 2016–2017. Temporal shifts were also observed in roadway functional classifications. Fatal crashes on local streets declined from 43.0% to 27.2% over the study period, while crashes on U.S. highways and interstates showed increases in the later years. Urban areas consistently dominated the crash setting, increasing from 85.6% to 94.2%. Among harmful event types, motor vehicle in-transport remained the most frequent across all years, while curb-related crashes rose noticeably in the final period, reaching 6.84%. Non-motorist involvement remained substantial throughout the study period, ranging from 19.7% to 26.2%.

Collision types remained relatively stable, with angle crashes comprising the largest share across all time periods. Front-to-rear collisions increased in 2022–2023, accounting for 27.4% of fatal crashes, while front-to-front collisions declined over time. Four-way intersections became increasingly represented in fatal crashes, rising from 26.9% in 2016–2017 to 40.0% in 2022–2023. Lighting conditions also changed, with crashes occurring under dark but lighted environments remaining frequent, while the share of daylight-related crashes declined to 50.1% in the most recent period. Weather patterns also evolved over time. Fatal crashes during rainy conditions increased in both 2020–2021 (13.1%) and 2022–2023 (11.2%), while clear weather remained the predominant condition across all periods. Driver-related contributing factors such as careless driving, improper lane usage, and failure to yield were most frequently reported between 2016 and 2019, but declined sharply in the later years. The configuration of roadways also shifted, with a growing proportion of fatal crashes occurring on undivided two-way roads with unprotected medians, reaching 30.8% in 2022–2023. The presence of traffic control signals at crash locations increased steadily, while crashes at sites with no controls showed a declining trend. Male drivers represented the majority of fatal crash involvements in all years, with their proportion peaking at 66.6% in 2020–2021. Drivers aged 45–64 consistently accounted for the largest age group, while a slight rise in fatal crashes involving drivers aged 24 or younger was observed in the final period.

Table 1: Crash Attribute Distributions Categorized by Year.

Variables	**2016–2017 N=849**	**2018–2019 N=667**	**2020–2021 N=602**	**2022–2023 N=775**
Season (Ssn)				
Fall	259 (30.5%)	190 (28.5%)	143 (23.8%)	288 (37.2%)
Spring	201 (23.7%)	176 (26.4%)	128 (21.3%)	134 (17.3%)
Summer	169 (19.9%)	117 (17.5%)	183 (30.4%)	201 (25.9%)
Winter	220 (25.9%)	184 (27.6%)	148 (24.6%)	152 (19.6%)
Day of Week (DOW)				
Weekday	650 (76.6%)	524 (78.6%)	451 (74.9%)	555 (71.6%)
Weekend	199 (23.4%)	143 (21.4%)	151 (25.1%)	220 (28.4%)
Time of Day (Hr)				
Day Non-peak	277 (32.6%)	229 (34.3%)	171 (28.4%)	191 (24.6%)
Evening Peak	184 (21.7%)	161 (24.1%)	157 (26.1%)	147 (19.0%)
Morning Peak	169 (19.9%)	96 (14.4%)	110 (18.3%)	136 (17.5%)
Night Non-Peak	219 (25.8%)	181 (27.1%)	164 (27.2%)	301 (38.8%)
Route Type (Fnc)				
County Road	37 (4.36%)	38 (5.70%)	44 (7.31%)	49 (6.32%)
Interstate	68 (8.01%)	76 (11.4%)	45 (7.48%)	87 (11.2%)
Local Street	365 (43.0%)	259 (38.8%)	259 (43.0%)	211 (27.2%)

Continued on next page

Table 1 continued from previous page

Variables	2016–2017 N=849	2018–2019 N=667	2020–2021 N=602	2022–2023 N=775
State Highway	234 (27.6%)	146 (21.9%)	146 (24.3%)	163 (21.0%)
U.S. Highway	123 (14.5%)	74 (11.1%)	66 (11.0%)	120 (15.5%)
Other	22 (2.59%)	74 (11.1%)	42 (6.98%)	145 (18.7%)
Area Type (Loc)				
Rural	120 (14.1%)	87 (13.0%)	71 (11.8%)	41 (5.29%)
Urban	727 (85.6%)	576 (86.4%)	529 (87.9%)	730 (94.2%)
Unknown	2 (0.24%)	4 (0.60%)	2 (0.33%)	4 (0.52%)
Harmful Event (HAE)				
Curb	2 (0.24%)	0 (0.00%)	3 (0.50%)	53 (6.84%)
Fixed Object	7 (0.82%)	49 (7.35%)	23 (3.82%)	14 (1.81%)
Motor Vehicle In-Transport	598 (70.4%)	466 (69.9%)	413 (68.6%)	530 (68.4%)
Non-Motorist	167 (19.7%)	148 (22.2%)	158 (26.2%)	162 (20.9%)
Rollover/Overturn	42 (4.95%)	2 (0.30%)	3 (0.50%)	8 (1.03%)
Others	33 (3.89%)	2 (0.30%)	2 (0.33%)	8 (1.03%)
Collision Type (Col)				
Angle	229 (27.0%)	231 (34.6%)	174 (28.9%)	225 (29.0%)
Front-to-Front	147 (17.3%)	61 (9.15%)	118 (19.6%)	64 (8.26%)
Front-to-Rear	147 (17.3%)	158 (23.7%)	99 (16.4%)	212 (27.4%)
Not a Collision with MVIT	251 (29.6%)	201 (30.1%)	189 (31.4%)	245 (31.6%)
Sideswipe	75 (8.83%)	14 (2.10%)	22 (3.65%)	21 (2.71%)
Other	0 (0.00%)	2 (0.30%)	0 (0.00%)	8 (1.03%)
Intersection Type (Int)				
Four-Way Intersection	228 (26.9%)	230 (34.5%)	221 (36.7%)	310 (40.0%)
Not Reported	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (0.26%)
Segment	505 (59.5%)	323 (48.4%)	320 (53.2%)	386 (49.8%)
T-Intersection	114 (13.4%)	104 (15.6%)	46 (7.64%)	65 (8.39%)
Others	2 (0.24%)	10 (1.50%)	15 (2.49%)	12 (1.55%)
Lighting Condition (Lgh)				
Dark - Lighted	234 (27.6%)	199 (29.8%)	194 (32.2%)	224 (28.9%)
Dark - Not Lighted	69 (8.13%)	62 (9.30%)	15 (2.49%)	82 (10.6%)
Daylight	527 (62.1%)	373 (55.9%)	359 (59.6%)	388 (50.1%)
Others	19 (2.24%)	33 (4.95%)	34 (5.65%)	81 (10.5%)
Weather Condition (Wth)				
Clear	602 (70.9%)	523 (78.4%)	446 (74.1%)	578 (74.6%)
Cloudy	126 (14.8%)	41 (6.15%)	32 (5.32%)	86 (11.1%)
Inclement weather	35 (4.12%)	16 (2.40%)	8 (1.33%)	9 (1.16%)

Continued on next page

Table 1 continued from previous page

Variables	2016–2017 N=849	2018–2019 N=667	2020–2021 N=602	2022–2023 N=775
Rain	44 (5.18%)	49 (7.35%)	79 (13.1%)	87 (11.2%)
Unknown	42 (4.95%)	38 (5.70%)	37 (6.15%)	15 (1.94%)
Roadway configuration (RdT)				
1-way Traffic	10 (1.18%)	26 (3.90%)	20 (3.32%)	46 (5.94%)
2-Way div with med bar	135 (15.9%)	111 (16.6%)	59 (9.80%)	100 (12.9%)
2-Way div with unprtcd med bar	167 (19.7%)	110 (16.5%)	138 (22.9%)	239 (30.8%)
2-Way undiv	391 (46.1%)	249 (37.3%)	249 (41.4%)	287 (37.0%)
2-Way undiv with Continuous Left-Trn Lane (CLTL)	52 (6.12%)	61 (9.15%)	53 (8.80%)	22 (2.84%)
Not Reported	94 (11.1%)	110 (16.5%)	83 (13.8%)	81 (10.5%)
Traffic Control (TCD)				
Flashing Traffic Control Signal	3 (0.35%)	0 (0.00%)	0 (0.00%)	6 (0.77%)
No Controls	573 (67.5%)	363 (54.4%)	295 (49.0%)	339 (43.7%)
Not Reported	89 (10.5%)	83 (12.4%)	106 (17.6%)	145 (18.7%)
Other Signs	1 (0.12%)	1 (0.15%)	5 (0.83%)	2 (0.26%)
Stop Sign	27 (3.18%)	40 (6.00%)	16 (2.66%)	23 (2.97%)
Traffic control signal	151 (17.8%)	175 (26.2%)	173 (28.7%)	259 (33.4%)
Other	4 (0.47%)	5 (0.75%)	0 (0.00%)	1 (0.13%)
Gender (Gen)				
Female	316 (37.2%)	251 (37.6%)	194 (32.2%)	278 (35.9%)
Male	512 (60.3%)	409 (61.3%)	401 (66.6%)	478 (61.7%)
Not Reported	21 (2.47%)	7 (1.05%)	7 (1.16%)	19 (2.45%)
Age Group (Age)				
25–44 Years	258 (30.4%)	222 (33.3%)	200 (33.2%)	240 (31.0%)
45–64 Years	301 (35.5%)	267 (40.0%)	234 (38.9%)	267 (34.5%)
65 Years and Over	107 (12.6%)	49 (7.35%)	69 (11.5%)	87 (11.2%)
LTE 24 years	170 (20.0%)	121 (18.1%)	91 (15.1%)	154 (19.9%)
Not Reported	13 (1.53%)	8 (1.20%)	8 (1.33%)	27 (3.48%)

3.2 Spatial Distribution

Figure 1 presents the spatial distribution of fatal transit bus crashes in the United States across four two-year periods: 2016–2017, 2018–2019, 2020–2021, and 2022–2023. Each panel utilizes a hexagonal tile map to visualize the state-level frequency of fatal crashes, with a diverging color scale indicating crash volume, where lighter shades represent higher counts. The spatial patterns remain consistent in identifying several states as persistent high-frequency locations. California recorded the highest number of fatal transit bus crashes across all four periods, maintaining a yellow tone throughout, which indicates crash counts surpassing 20 or more in each time frame. Florida

and Texas also consistently reported elevated frequencies, particularly in 2020–2021 and 2022–2023, reinforcing their roles as high-risk states for fatal transit bus incidents. Pennsylvania, New York, and Illinois exhibited moderately high counts, with noticeable increases during the 2020–2021 period before stabilizing or slightly declining in 2022–2023.

Several states demonstrated emerging trends over time. Georgia and Ohio, for instance, displayed a gradual rise in fatal crash involvement, transitioning from darker to more neutral tones by the end of the study period. Western states such as Nevada, Utah, and Colorado also showed mild increases during the middle years, reflecting possible urban growth or increased transit service coverage. In contrast, large rural states including Montana, Wyoming, and the Dakotas maintained consistently low fatal crash counts, indicated by deep purple hues across all periods. The spatial visualization underscores the concentration of fatal transit bus crashes in high-population or transit-dense regions. The persistence of elevated crash frequencies in specific states suggests structural or operational challenges that may not be fully mitigated by existing safety measures.

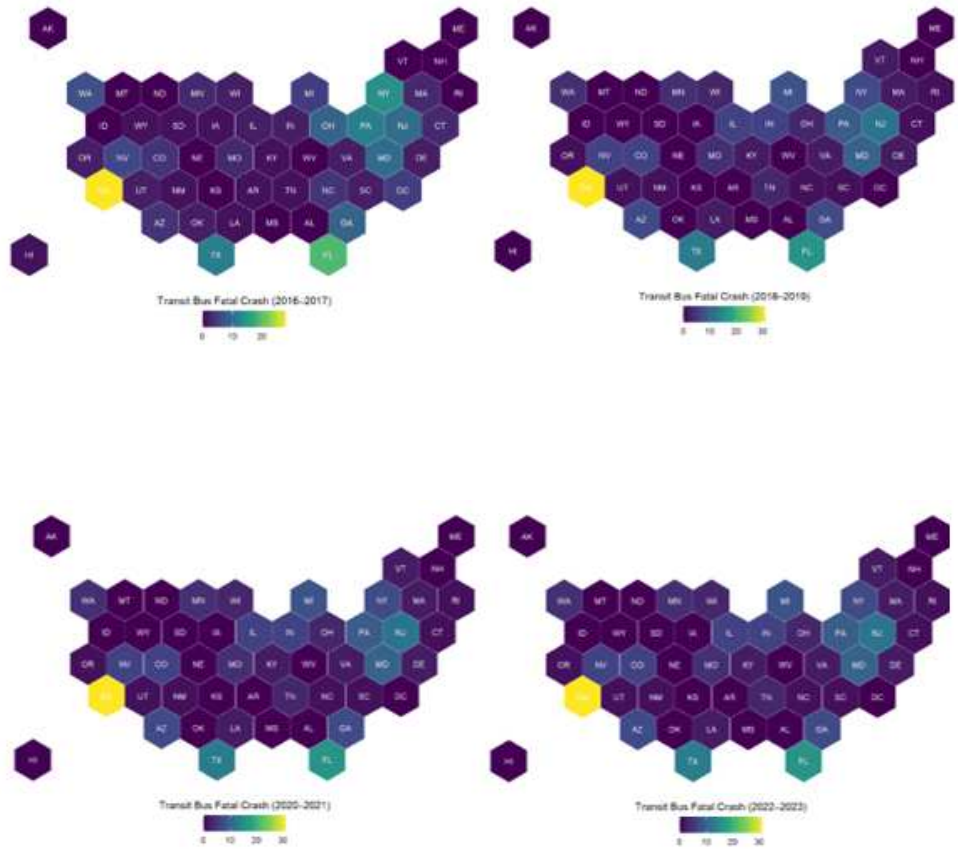


Fig. 1: Spatial Distribution of Fatal Transit Bus Crashes by State

3.3 Study Design

This study was conducted through three sequential phases. In the data preparation phase, fatal transit bus crash records obtained from the Fatality Analysis Reporting System (FARS) from 2016 to 2023 were compiled into a unified database. Relevant crash, roadway, environmental, temporal, vehicle, and driver-related variables were extracted for analysis. During the preprocessing phase, duplicate records were removed, missing or incomplete observations were addressed, and variables were screened for consistency and analytical relevance. Continuous variables were categorized into meaningful groups where appropriate to facilitate categorical analysis. In the analytical phase, Cluster Correspondence Analysis (CCA) was employed to identify latent crash typologies and uncover associations among crash characteristics. Finally, in the interpretation phase, the resulting clusters were visualized using bi-plots and variable-tree (vtree) diagrams, and the identified crash patterns were translated into targeted safety recommendations and policy implications aimed at reducing fatal transit bus crashes (Barrowman et al. 2022). Figure 2 represents the study framework followed in this study.

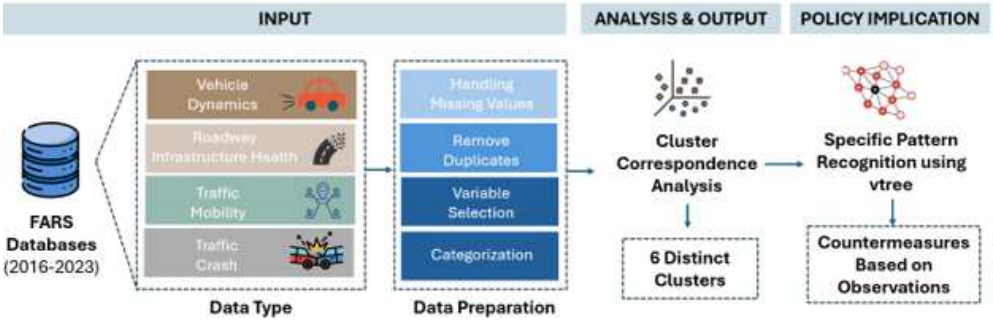


Fig. 2: Study Flowchart

3.4 Cluster Correspondence Analysis

The CCA method introduced by Markos et al. (2019) integrates dimension reduction and clustering specifically for categorical data, offering improved cluster convergence over earlier approaches. It relies on a binary super indicator matrix Z, structured as an $n \times Q$ matrix, where n represents the number of observations and Q is the total number of distinct categories from all categorical variables. Each observation is encoded by a one-hot representation, with a '1' placed at the category it belongs to and zeros elsewhere. The matrix Z is composed of segments $[Z_1, Z_2, \ldots, Z_q]$, where Z_j is an $n \times p_j$ indicator matrix for the j-th categorical variable, and q denotes the total number of such variables. Similar to factor loading matrices in continuous data analysis, a matrix B is used to assign numerical values to each category. B is constructed as:

$$B = [B_1^T, B_2^T, \ldots, B_q^T]^T,$$

where each B_j has dimensions $p_j \times d$ and contains the quantifications for categories of the j-th variable. The low-dimensional coordinates of the observations are stored in matrix Y, an $n \times d$ matrix, cluster memberships are recorded in Z_k, an $n \times K$ matrix, and the corresponding cluster centroids are held in matrix G with dimensions $K \times d$ (Markos et al. 2019).

To evaluate the relationships between clusters and categorical variables, a contingency matrix is constructed by cross tabulating the indicator matrix with the cluster membership matrix. This results in a matrix: $F = Z_k^T Z$ (Markos et al. 2019). By applying CCA to this contingency matrix, the optimal scaling values for both clusters and categories can be derived to maximize between-group variances. This process enhances the separability of clusters while improving the distribution of category quantifications within variables.

The CCA algorithm operates through an iterative process involving several key steps:

1. The cluster assignment matrix Z_k is initialized by randomly grouping each data point into one of the clusters.
2. The category quantification matrix B is derived based on the current cluster structure.
3. The coordinate matrix Y is calculated by averaging the centered scores, expressed as:
$$Y = \frac{1}{q}\left(I_n - \frac{1_n 1_n^{'}}{n}\right) ZB.$$
4. The cluster membership matrix Z_k is updated by applying K-means clustering to the updated coordinates Y.

These steps are repeated, with each iteration updating Z_k, until convergence is achieved. Convergence is defined by the stabilization of Z_k, Y, and G, indicating that a consistent clustering structure has been found (Markos et al. 2019). Once convergence is reached, the final matrices Z_k, G, and B represent optimal cluster memberships, centroids, and category quantifications, respectively. These can be used to construct a biplot for visualizing clusters alongside their associated categories. However, due to the potential mismatch between the number of clusters, often fewer, and the number of categories, standard Correspondence Analysis (CA) normalization may not guarantee an equal spatial distribution of row and column points.

To ensure consistency in visualization, a scaling adjustment is recommended. The cluster mean points are scaled by a constant γ, while category quantifications are scaled by its inverse. This maintains a uniform average squared deviation from the origin for both clusters and categories:

$$\gamma = \left(\frac{K}{Q} \cdot \frac{\operatorname{Tr} B^T B}{\operatorname{Tr} G^T G}\right)^{1/4}.$$

Using this scaling factor, the transformed matrices are defined as:

$$G_s = \gamma G \quad \text{and} \quad B_s = \frac{1}{\gamma} B,$$

ensuring consistent scaling and interpretability of the resulting biplots (Markos et al. 2019).

4 Results and analysis

Prior to conducting the CCA, the optimal number of clusters was determined using the elbow method. The model criterion was evaluated for cluster solutions ranging from $k = 2$ to $k = 10$, and the resulting values were plotted against the corresponding number of clusters, as illustrated in Figure 3. The curve exhibited a substantial improvement in the model criterion between two and six clusters, after which the rate of improvement became considerably smaller. This inflection point, commonly referred to as the "elbow," indicates that additional clusters beyond six provide only marginal gains in model performance while increasing solution complexity. Therefore, a six-cluster solution was selected as the optimal representation of the underlying crash typologies and was subsequently employed in the CCA analysis.

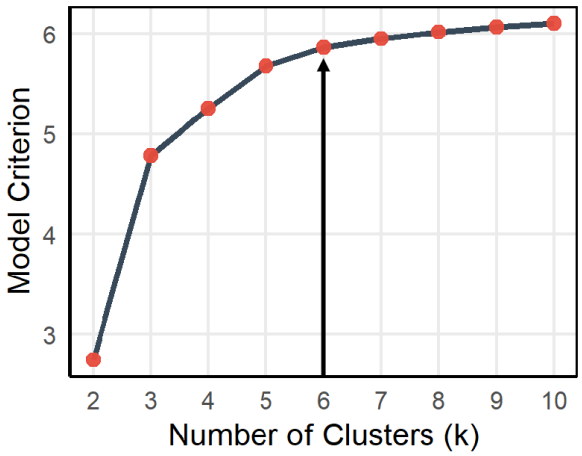


Fig. 3: Elbow method used to determine the optimal number of clusters for the transit bus fatal crash dataset.

This section includes the explanations on the findings from the CCA analysis in this study. It also includes a description of the policy level suggestions that are made from the analysis. The Bi-Plot in Figure 4 shows how each of the six clusters is shaped by different crash factors, helping identify where policy changes can reduce transit bus fatalities. Cluster 6 is clearly separated and linked to rollover crashes, improper lane

use, and bad weather on interstates. Cluster 5 is tied to rain and fixed-object crashes, pointing to a need for safer roadside design and drainage improvements. Cluster 4 shows strong links to curb-related crashes and missing traffic controls, highlighting the importance of curb zone upgrades and consistent infrastructure reporting.

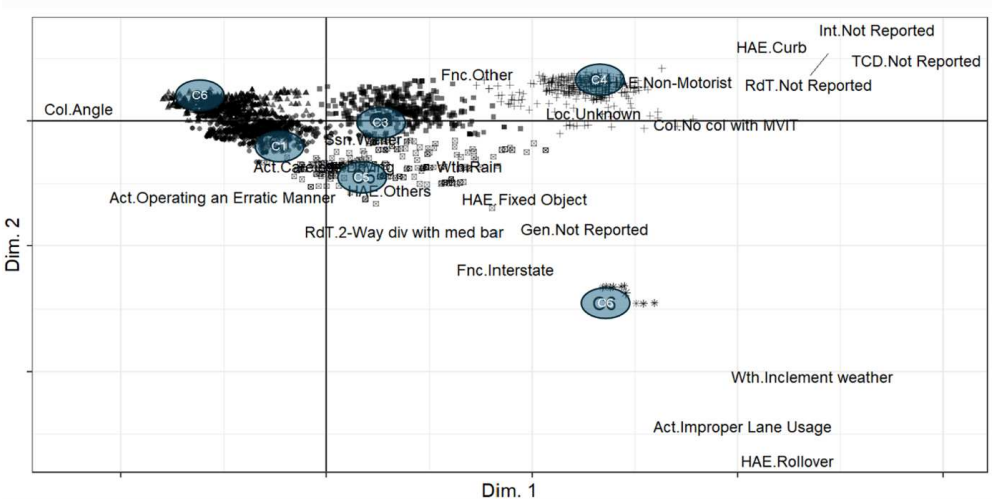


Fig. 4: Cluster Bi-Plot

4.1 Cluster Correspondence Analysis

4.1.1 Cluster 1 - High-Speed Roadway Conflicts with Limited Traffic Control

Cluster 1 in Figure 5, which accounts for 31.6% of fatal transit bus crashes, reflects scenarios where crashes often occur on highways and rural roads with limited or no traffic control. The representation of U.S. Highways and rural locations indicates that these fatal incidents frequently happen outside dense urban centers, where higher travel speeds and longer gaps between controlled intersections are common. In such settings, drivers and non-motorists may have fewer regulated points of interaction, which can increase the severity of any resulting crash. This cluster shows a strong association with front-to-front and front-to-rear collisions, both of which are typically high-impact crash types. These patterns may arise from high-speed travel on undivided or partially divided roads, specifically, two-way roads with either unprotected medians or center left-turn lanes (CLTL). Transit buses operating in such environments may have less protection from oncoming vehicles and may face risks when turning or merging. The frequent presence of locations without traffic control devices indicates that uncontrolled roadway environments are a common characteristic of this crash typology. Additionally, the segment-level crash locations (as opposed to intersection-related crashes) and low reporting of traffic signals suggest that many

of these events do not occur at regulated crossings, leaving more room for driver misjudgment. Improper or erratic lane changing, a behavior that appears positively in this cluster, may be especially dangerous on high-speed roads, where sudden maneuvers can result in severe collisions, particularly with large vehicles like transit buses. The involvement of sideswipe crashes also supports this, as they often result from unsafe lane changes or vehicles encroaching on bus lanes or shoulders. In contrast, variables negatively associated with this cluster, such as non-motorist involvement, four-way intersections, and stop signs, suggest that pedestrian-related and intersection-based crashes are less frequent here. Instead, fatal crashes in this group are more likely due to vehicular movements at higher speeds in loosely controlled environments.



Fig. 5: High-Speed Roadway Conflicts with Limited Traffic Control (Cluster 1)

4.1.2 Cluster 2 - Urban Intersection Violations and Angle Collisions

Cluster 2 in Figure 6, representing 30.3% of fatal transit bus crashes, highlights crash patterns that are primarily shaped by intersection-related conflicts within urban settings, particularly at signalized and stop-controlled intersections. This cluster is strongly associated with four-way intersections and angle collisions, which are commonly seen in urban traffic networks where multiple road users, including pedestrians, cyclists, and vehicles, converge. The presence of traffic control signals and stop signs, both of which appear prominently in this cluster, points to environments where right-of-way rules and signal compliance are critical for crash prevention. Failure to obey traffic rules emerged as a prominent characteristic of this cluster, suggesting that regulatory noncompliance may be an important contributing factor in these crash scenarios.. This could refer to either the bus driver or other involved parties disregarding signals, signs, or lane markings, especially at intersections, leading to crashes. Angle collisions, which are often severe due to the side impact, are typically a result of such

violations, particularly when drivers attempt to cross or turn at intersections during red lights or without yielding. Another important aspect of this cluster is the location context. Unlike the previous cluster, this group is more associated with local streets and one-way or undivided two-way roads, where frequent intersections and varied road users are more common. The positive association with T-intersections also points to urban geometries that can complicate visibility and judgment, especially for large vehicles like transit buses attempting left or right turns. Although rural areas, rainy weather, and segments without reported traffic control devices are negatively associated with this cluster, their absence reinforces that these crashes are more typical of regulated, built-up environments rather than uncontrolled or open-road settings. The negative association with non-motorist involvement suggests these crashes are more vehicle-to-vehicle in nature rather than involving pedestrians or cyclists.

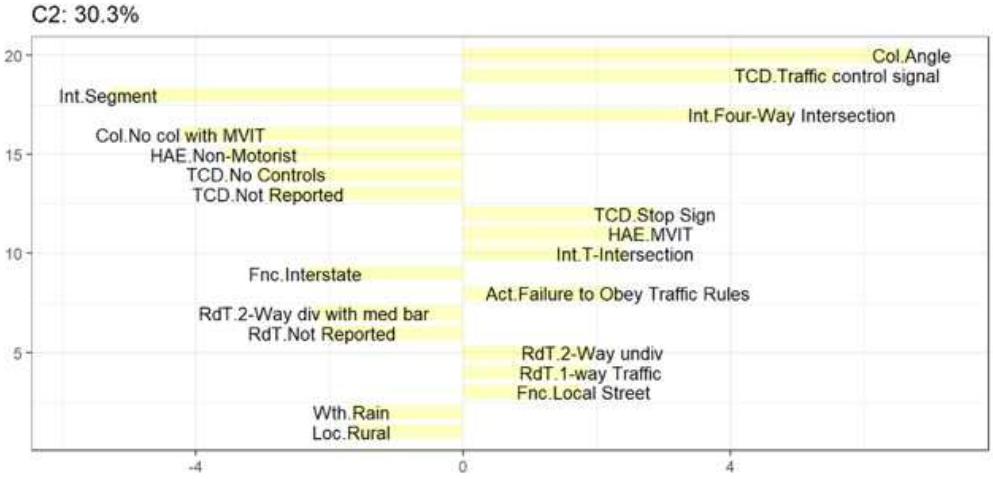


Fig. 6: Urban Intersection Violations and Angle Collisions (Cluster 2)

4.1.3 Cluster 3 - Non-Motorist Involvement in Low-Speed Rural Settings

Figure 7 represents Cluster 3, which includes 13.8% of the fatal transit bus crashes, is defined by the strong presence of non-motorist involvement, particularly pedestrians or bicyclists, as the primary harmed party. This group of crashes is characterized by incidents where the transit bus was not in collision with another motor vehicle in transport, reinforcing the idea that these fatal events likely involved vulnerable road users. The positive association with the harmful event category "non-motorist" and the collision type "no collision with MVIT" further supports this. These crashes are more likely to occur in rural areas, where sidewalks, pedestrian crossings, and visibility-enhancing infrastructure may be limited. Such settings often place pedestrians in shared road space with large vehicles like buses, increasing the risk of fatal outcomes in the event of any conflict. The appearance of "curb" and "other" as harmful events also suggests

that pedestrians may be struck during boarding, alighting, or while walking near the edge of the roadway. Transit buses may be maneuvering in areas without marked bus stops or designated pedestrian zones, making such situations especially risky. Additionally, this cluster shows a negative relationship with poor weather, meaning many of these incidents happen in clear or moderate conditions, further suggesting visibility or infrastructure, not weather, may be the underlying concern. This cluster highlights a serious safety concern related to vulnerable road users in less regulated environments.

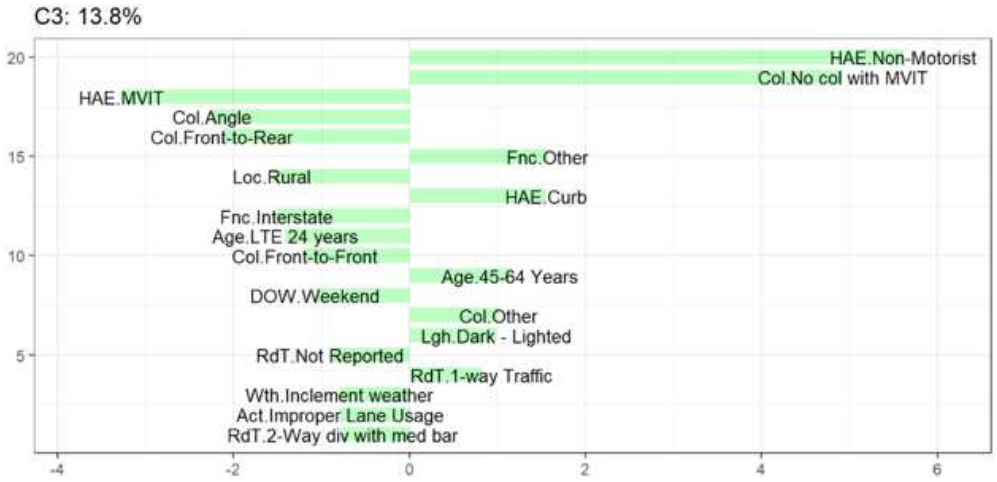


Fig. 7: Non-Motorist Involvement in Low-Speed Rural Settings (Cluster 3)

4.1.4 Cluster 4 - Data Gaps and Non-Motorist Fatalities in Uncontrolled Settings

Cluster 4 in Figure 8, covering 12.7% of fatal transit bus crashes, is shaped by a notable concentration of incomplete reporting and non-motorist fatalities, often in environments where traffic control and roadway details are not documented or absent. The most prominent feature of this cluster is the representation of "Not Reported" under both traffic control devices (TCD) and road type (RdT), which highlights a concerning limitation in crash documentation. These data gaps could reflect less formal roadway settings or limitations in post-crash investigation procedures, but they also make it more difficult to assess risk patterns and design preventive measures. A major theme in this cluster is the involvement of non-motorists, such as pedestrians or cyclists, who ended up in fatality. The prominence of curb-related events and non-motorist involvement may reflect conflicts occurring near curbside activity areas, although the exact operational context cannot be determined from the available data. The involvement of "curb" as a harmful event supports this interpretation. Elderly individuals (age 65 and over) are more commonly involved in this group, which is significant given the increased physical vulnerability of older pedestrians. This cluster draws attention

to the intersection of data limitations, missing infrastructure, and the exposure of non-motorists, especially in poorly monitored or loosely regulated environments.

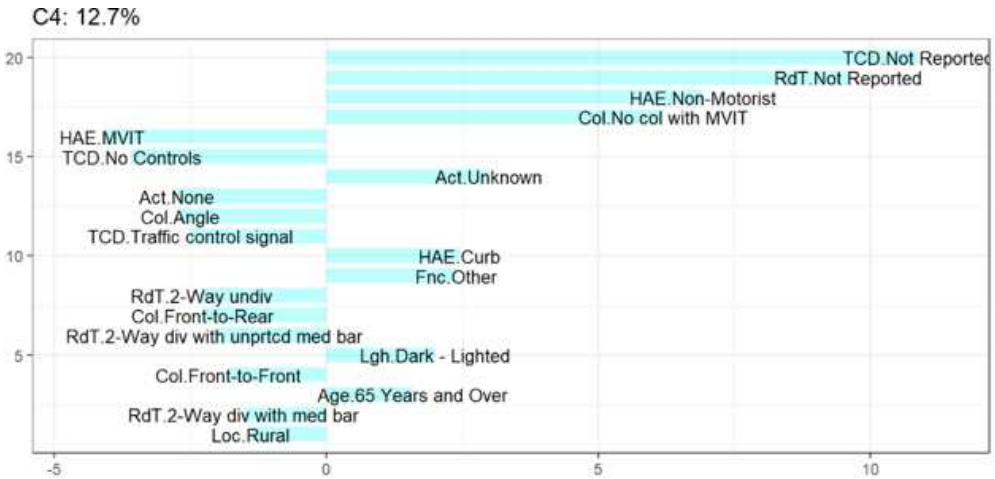


Fig. 8: Data Gaps and Non-Motorist Fatalities in Uncontrolled Settings (Cluster 4)

4.1.5 Cluster 5 - Interstate and Fixed-Object Collisions Under Poor Conditions

Cluster 5 in Figure 9 accounts for 10.5% of fatal transit bus crashes and is primarily shaped by incidents occurring on high-speed divided highways, particularly interstates and two-way roads with medians. The dominant presence of fixed-object crashes, including harmful events involving stationary roadside features, suggests that these incidents often involve loss of vehicle control or roadway departure. This may occur when transit buses leave their lanes due to slippery surfaces, reduced visibility, or sudden evasive maneuvers. Environmental factors play a significant role in this cluster. Fatal crashes here are more likely to happen during rain, in dark areas without lighting, or under generally inclement weather. These conditions reduce visibility, increase stopping distances, and provide challenging driving situations, especially for large vehicles like buses, which have longer braking distances and less maneuverability. Winter as a season is also positively associated, reinforcing the role of adverse weather in increasing crash risk. Driver-related factors are also evident in this group, particularly through the presence of careless driving and improper lane usage. On interstates, even small mistakes in lane discipline can have severe consequences, especially during bad weather or low visibility. Additionally, these crashes occur more often on road segments (rather than intersections), where speeds are typically higher and recovery from mistakes is more difficult. Variables that are negatively represented in this cluster include local streets, four-way intersections, and traffic signals, indicating that these crashes are not typical urban intersection-type collisions. Instead, they reflect

highway-specific risks such as drifting into medians or off-road objects, especially when drivers fail to adjust for environmental conditions.

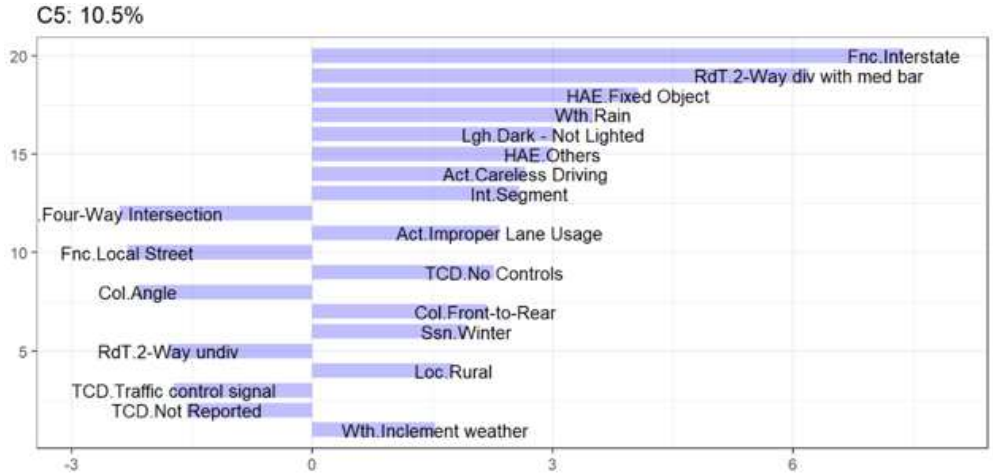


Fig. 9: Interstate and Fixed-Object Collisions Under Poor Conditions (Cluster 5)

4.1.6 Cluster 6 - Rollover Crashes on High-Speed Roads During Harsh Conditions

Cluster 6 in Figure 10, though representing only 1.2% of the total fatal transit bus crashes, reveals a distinct crash scenario involving rollover events on high-speed divided roads, particularly interstates, under poor weather conditions. The most prominent factor in this cluster is the harmful event "rollover" which often leads to severe or fatal outcomes, especially when involving large vehicles like transit buses. Rollover crashes are closely associated with unstable maneuvers, often triggered by improper lane usage, which is another key factor in this cluster. When transit buses veer off their lane at high speeds, especially on slippery or curved road segments, the risk of rollover increases significantly. The frequent appearance of inclement weather and the winter season further suggests that these crashes tend to occur in harsh driving conditions, where traction is limited and vehicle stability is harder to maintain. This cluster also includes divided two-way roads with medians and interstates, both of which support higher travel speeds. The combination of speed, large vehicle mass, and sudden steering or braking in response to environmental or driver-related issues can easily result in rollover events. Interestingly, most of these crashes are not associated with a collision involving another moving vehicle, as indicated by the presence of "no collision with MVIT". This suggests that the transit bus may be the only vehicle involved, with the crash severity stemming entirely from the bus's own instability or control loss. The crashes occur more during non-peak hours of the day, possibly when traffic volumes are lower, but weather-related risks are higher (e.g., early morning frost or snow). This cluster highlights a rare but serious crash pattern involving rollovers

during poor weather on high-speed roads. It calls attention to the importance of safe driving practices during adverse conditions, the need for stability control systems on transit buses, and consideration of how road geometry and weather together influence crash risk.

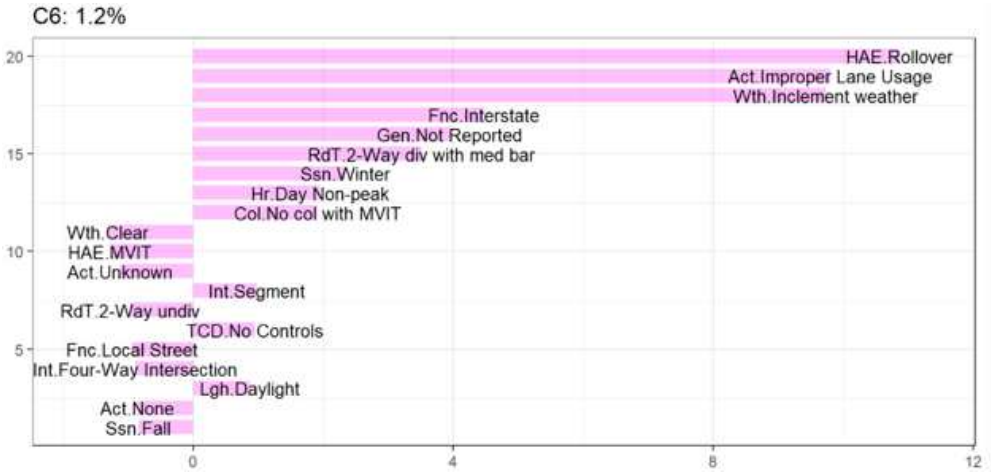


Fig. 10: Rollover Crashes on High-Speed Roads During Harsh Conditions (Cluster 6)

4.2 Policy Implications

This section includes explanations and discussions on providing countermeasures to reduce fatalities by making development through policy level changes. Figure 11, the variable-tree diagram provides a clear visual summary of the dominant crash patterns across different clusters of fatal transit bus crashes, focusing on three critical factors: collision type, intersection type, and roadway configuration. The clusters on the left side of the diagram account for the majority of cases. One of the largest clusters, comprising 31% of the cases, is characterized by crashes at four-way intersections located on roads classified as "Other", indicating a potential lack of standardized roadway designations. This suggests a policy gap in roadway documentation and design consistency, calling for improved infrastructure classification and formal design standards. Another major cluster, involving 26% of the cases, shows a strong association with front-to-rear collisions at four-way intersections. This pattern points to issues such as inadequate signal timing, sudden stops, and limited visibility, which can be addressed through traffic signal upgrades, better signage, and dedicated bus lanes at intersections. A third significant cluster, accounting for 25% of the cases, involves other types of collisions on two-way undivided roads with continuous left-turn lanes (CLTL), also at four-way intersections. These roads lack physical separation between opposing lanes and are risky for large vehicles like buses. This reinforces the need for converting such roadways into divided lanes with raised medians or protected zones, particularly in corridors with high transit activity. Clusters in the middle-right of the diagram,

although smaller in frequency, highlight recurring issues such as undefined roadway types and poorly documented intersection characteristics. These patterns indicate the need for standardized crash scene reporting and consistent deployment of traffic control devices, especially on less formal road segments. Smaller clusters on the far right, while representing fewer cases, still reveal consistent risk factors such as front-to-rear collisions on undivided roads and crashes in non-standard intersection settings. These findings support broader policy recommendations made in the manuscript, including the need to replace continuous left-turn lanes with safer alternatives, implement collision avoidance systems in transit buses, and improve roadway lighting and design clarity on undivided roads.

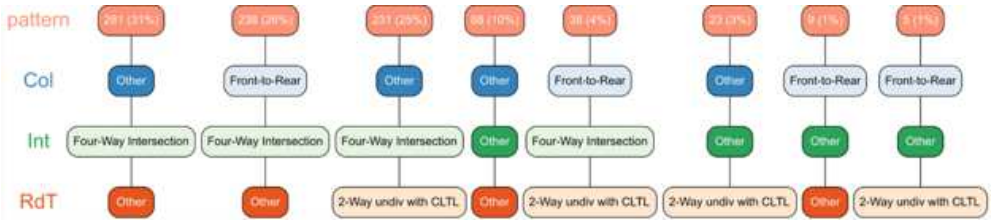


Fig. 11: Findings from Clusters

Some other findings from the cluster analysis (see Figure 12) points to a clear policy issue: a large proportion of fatal transit bus crashes in this group occur on two-way divided roads with unprotected median barriers, in areas with no traffic control devices, and frequently involve curb-related events. This combination suggests a gap in basic roadway safety infrastructure and regulation. At the policy level, one of the first steps should be to prioritize upgrades on these road segments by introducing consistent and effective traffic control measures. Roads with no signals, signs, or markings leave too much room for uncertainty, particularly when large vehicles like buses are operating in the area. Implementing stop signs, flashing beacons, or full traffic signals, depending on traffic volume and pedestrian activity, can help organize movement and reduce risk. Another key policy solution is to review the use of unprotected median barriers. These types of roadways often give the appearance of divided roads but do not offer the safety benefits of true physical separation. Policies should support replacing unprotected medians with raised or buffered medians, especially where buses regularly operate or where pedestrian activity is present. This would reduce the risk of head-on or side-swipe crashes and make curbside maneuvers safer.

Based on the findings, the Safe System Approach (SSA) Roadway Design Hierarchy can provide a critical framework for guiding targeted safety improvements in fatal transit bus crash contexts. At *Tier 1: Remove Severe Conflicts*, several clusters (e.g., high-speed rural roads, two-way undivided segments, and intersections with continuous left-turn lanes) highlight the need to eliminate or redesign dangerous configurations. Countermeasures such as replacing CLTLs with raised medians, creating dedicated turn bays, and adding protected intersections can significantly reduce opportunities for high-impact front-to-front or front-to-rear collisions. *Tier 2: Reduce Vehicle Speeds* is

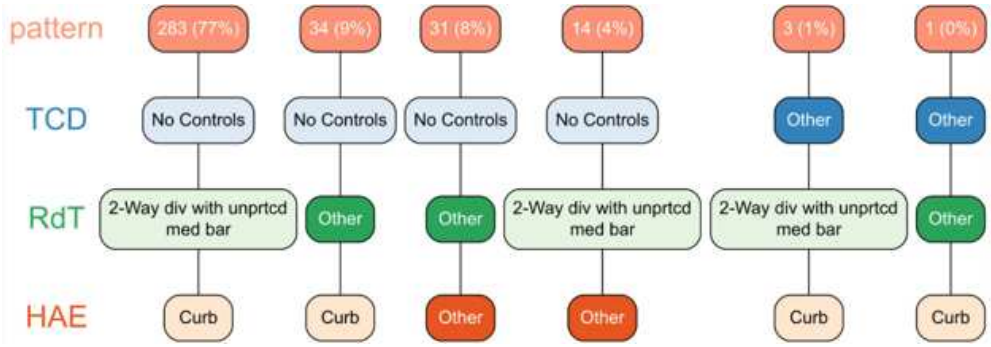


Fig. 12: Findings from clusters on TCD

especially relevant in rural and suburban segments where limited traffic control allows excessive speed. Installing dynamic speed management systems, narrowing lanes, and using signage or enforcement tools like speed cameras in identified crash hotspots can lower impact severity. Clusters involving non-motorist fatalities (e.g., pedestrian impacts near curbs or during bus boarding) point to the importance of *Tier 3: Manage Conflicts in Time*, where the introduction of clearly marked crosswalks, curb extensions, and pedestrian-priority signals can improve visibility and reaction time for all users.

For *Tier 4: Increase Attentiveness and Awareness*, crashes associated with incomplete reporting, unlit roads, and driver errors under poor weather conditions reveal the need for both behavioral and technological solutions. These include bus-mounted collision avoidance systems, improved roadway lighting, in-vehicle stability controls, and targeted training for transit drivers on high-risk conditions such as winter weather or poorly controlled intersections. Finally, data gaps in crash reports highlight the need for improved crash documentation and infrastructure inventory. This aligns with SSA's system-level principles, where better data supports proactive risk identification. Policies mandating standard reporting formats, enhanced mapping of roadway features, and integration of real-time environmental data will support more robust interventions. By systematically aligning countermeasures across SSA design tiers, cities can create layered defenses that account for human error, vehicle limitations, and environmental complexity.

5 Conclusions

Transit buses, while essential to urban transportation, are often involved in severe crashes due to their operation in dense, high-risk urban environments, where they interact with a variety of road users and navigate complex intersections. These environments expose buses to a range of risks, including collisions with other vehicles, pedestrians, and cyclists. The challenges are compounded by the nature of urban streets, which often lack proper infrastructure and traffic control systems to protect the safety of transit buses and other road users. This study aimed to address the

growing concern of fatal transit bus crashes by utilizing advanced analytical methods such as CCA to identify the key contributing factors and patterns that lead to such crashes. By focusing on fatal crashes reported in the FARS database, this study sought to uncover hidden crash typologies and provide insights for enhancing transit safety through policy-level interventions. The goal was to not only understand the crash characteristics but also offer practical solutions that can be applied in real-world settings to prevent future fatalities.

The major findings of this study revealed that fatal transit bus crashes are heavily influenced by several key factors, including intersection-related conflicts, high-speed roadway environments, non-motorist involvement, and inadequate traffic controls. Specifically, the analysis identified that many fatal crashes occur at intersections, particularly those with poor visibility or inadequate traffic management. Additionally, crashes on high-speed roads, where buses are often at risk of colliding with other vehicles due to limited protective barriers or lane separations, were highlighted as a significant safety concern. Another important factor was the involvement of non-motorists, such as pedestrians or cyclists, especially in areas where there are no designated safety zones or pedestrian infrastructure. The study emphasized the importance of improving road design, such as redesigning intersections, upgrading traffic control systems, and providing dedicated bus lanes in high-risk areas. Moreover, the findings called attention to the need for policies that address driver behavior. These policies should focus on educating both bus operators and general road users about safe driving practices around buses. Through this analysis, the study focused on practical, policy-level countermeasures, like infrastructure upgrades, better road safety features, and stricter enforcement of traffic rules, which can significantly reduce fatalities associated with transit bus crashes. These measures not only aim to enhance the safety of bus passengers but also improve safety for pedestrians, cyclists, and other road users who share the roads with transit buses.

Despite its comprehensive approach, this study does have some limitations. The reliance on the FARS database, while invaluable, means that the data is inherently dependent on the accuracy and consistency of reporting across jurisdictions, which can vary and impact the completeness of the analysis. Additionally, the study primarily focused on categorical crash data, which may overlook more complex and dynamic aspects of crash causation that could offer additional insights. For instance, temporal variations, such as the exact time of day, seasonality, or even specific operational patterns of buses, were not explored in detail. Future research could address these limitations by incorporating real-time behavioral data from transit vehicles, such as GPS and driver behavior logs, to capture more subtle, dynamic interactions between transit buses, traffic, and road users. Moreover, integrating more granular environmental data, such as traffic flow, road conditions, and weather patterns, could provide a better understanding of how these factors contribute to crash risks. Further investigation into the effectiveness of specific countermeasures, such as the impact of collision avoidance systems, adaptive traffic management, or the role of emerging technologies like automated braking, could provide valuable insights into how these innovations can reduce fatalities and improve transit safety in the future.

6 Funding Source

The authors disclosed no financial support for the research, authorship, and/or publication of this article.

7 Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Chowdhury. Tausif Islam, Das. Subasish; data collection: Das. Subasish; analysis and interpretation of results: Chowdhury. Tausif Islam; draft manuscript preparation: Rafat, Sharif Ahmed. All authors reviewed the results and approved the final version of the manuscript.

8 Declaration of Conflicting Interests

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

9 Acknowledgement

The authors would like to express their sincere gratitude to Rohit Chakrobarty and Ananya Ghosh Tusti for their valuable support and contributions to this study. Their insights, guidance, and assistance throughout the research process greatly enhanced the quality and depth of this work.

10 Data Availability Statement

The data used in this study were obtained from the FARS, maintained by the National Highway Traffic Safety Administration (NHTSA), a publicly accessible national database. The FARS data are freely available to the public through the NHTSA website (https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars). The specific subset of records compiled and processed for this analysis (transit/city bus involved fatal crashes, 2016–2023) can be made available by the corresponding author upon reasonable request.

11 Ethical Statement

This study utilized secondary, de-identified crash data obtained from the FARS, a publicly available national database maintained by the NHTSA. The dataset does not contain any personally identifiable information regarding individuals involved in the reported crashes, including drivers, passengers, pedestrians, or other road users. As the data are anonymized, publicly accessible, and were not collected by the authors through direct interaction with human subjects, this research did not require approval from an Institutional Review Board (IRB) or equivalent ethics committee. All analyses were conducted in accordance with the data use terms specified by NHTSA for FARS data.

References

Barrowman N, Gatscha S, Momoli F (2022) vtree: Display Information About Nested Subsets of a Data Frame. Tech. rep.

Colavito T, Blassic L (2025) Bus Accident Statistics – 2025 Edition. Tech. rep., URL https://www.truckinfo.net/research/bus-accident-statistics

Coletti (2025) Bus Accident Statistics in the United States [2025]. URL https://www.paulsoncoletti.com/bus-accident-statistics/, accessed: 2026-06-11

Das S, Mousavi S, Shirinzad M (2022) Pattern recognition in speeding related motorcycle crashes. Journal of Transportation Safety & Security 14:1121–1138. https://doi.org/10.1080/19439962.2021.1877228

Das S, Hossain M, Rahman M, et al (2023) Understanding patterns of moped and seated motor scooter (50 cc or less) involved fatal crashes using cluster correspondence analysis. Transportmetrica A: Transport Science 19. https://doi.org/10.1080/23249935.2022.2029613

D'Souza K, Maheswari S (2012) Multivariate Statistical Analysis of Public Transit Bus Driver Distraction. Journal of Public Transportation 15:1–23. https://doi.org/10.5038/2375-0901.15.3.1

Dzinyela R, Dadashova B, Westfall G, et al (2025) Analysis of motorcyclists crash severity using cluster correspondence and hierarchical binary logit models. Multimodal Transportation 4:100197. https://doi.org/10.1016/j.multra.2025.100197

Feng S, Li Z, Ci Y, et al (2016) Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. Accident Analysis & Prevention 86:29–39. https://doi.org/10.1016/j.aap.2015.09.025

FTA (2021) Bus Safety Data Report: Bus Transit Safety Data 2008–2018. Technical report, Federal Transit Administration, URL https://www.transit.dot.gov/regulations-and-guidance/safety/bus-safety-data-report, accessed: 2026-06-11

FTA (2023) Bus-to-Person Collisions Safety Advisory 23-1. URL https://www.transit.dot.gov/regulations-and-programs/safety/bus-person-collisions-safety-advisory-23-1, accessed: 2026-06-11

Goh K, Currie G, Sarvi M, et al (2014) Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. Accident Analysis & Prevention 66:20–26. https://doi.org/10.1016/j.aap.2013.12.022

Henezi D, Winkler A (2023) The Role of Public Transport in Transport Safety and Public Safety. EPSTEM 23:505–512. https://doi.org/10.55549/epstem.1374907

Hossain A, Sun X, Islam S, et al (2023) Identifying roadway departure crash patterns on rural two-lane highways under different lighting conditions: association knowledge using data mining approach. Tech. rep., https://doi.org/10.48550/ARXIV.2302.14754

Hosseini S, Ardabili B, Azarbayjani M, et al (2025) Demographic disparities, service efficiency, safety, and user satisfaction in public bus transit system: A survey-based case study in the city of Charlotte, NC. Transportation Research Interdisciplinary Perspectives 29:101296. https://doi.org/10.1016/j.trip.2024.101296

Hutter F, Kotthoff L, Vanschoren J (2019) Automated Machine Learning: Methods, Systems, Challenges. The Springer Series on Challenges in Machine Learning, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-05318-5

Ibrahim M, Logan D, Koppel S, et al (2024) The role of safety in modal choice and shift: A transport users' perspective in Australia. Journal of Transport & Health 38:101863. https://doi.org/10.1016/j.jth.2024.101863

Joewono T, Kubota H (2006) Safety and Security Improvement in Public Transportation Based on Public Perception in Developing Countries. IATSS Research 30:86–100. https://doi.org/10.1016/S0386-1112(14)60159-X

Kaplan S, Prato C (2012) Associating Crash Avoidance Maneuvers With Driver Attributes and Accident Characteristics: A Mixed Logit Model Approach. Traffic Injury Prevention 13:315–326. https://doi.org/10.1080/15389588.2011.654015

Li H, Yu S, Deng S, et al (2024) Identification of Risk Factors for Bus Operation Based on Bayesian Network. Applied Sciences 14:9602. https://doi.org/10.3390/app14209602

Lu CC, Liang JK (2023) Exploring factors that influence the cardiovascular health of bus drivers for the improvement of transit safety. International Journal of Occupational Safety and Ergonomics 29:1263–1272. https://doi.org/10.1080/10803548.2022.2120259

Mardi Safitri D, Surjandari I, Jachrizal Sumabrata R (2020) Assessing factors affecting safety violations of bus rapid transit drivers in the Greater Jakarta Area. Safety Science 125:104634. https://doi.org/10.1016/j.ssci.2020.104634

Markos A, D'Enza A, Van De Velden M (2019) Beyond Tandem Analysis: Joint Dimension Reduction and Clustering in R. Journal of Statistical Software 91. https://doi.org/10.18637/jss.v091.i10

Morency P, Strauss J, Pépin F, et al (2018) Traveling by Bus Instead of Car on Urban Major Roads: Safety Benefits for Vehicle Occupants, Pedestrians, and Cyclists. Journal of Urban Health 95:196–207. https://doi.org/10.1007/s11524-017-0222-6

Nævestad TO, Elvik R, Milch V, et al (2022) Traffic safety in bus transport: An analysis of Norway's largest transit authority's contract requirements to bus companies. Transportation Research Part F: Traffic Psychology and Behaviour 89:317–333. https://doi.org/10.1016/j.trf.2022.07.004

Singh H, Kathuria A (2024) Behind the wheel: Probing into personality, skills, and driving behavior's role in bus rapid transit crashes. Traffic Injury Prevention 25:604–611. https://doi.org/10.1080/15389588.2024.2322672

Song Y, Tse H, Lind E, et al (2021) Safety Evaluation of Transit Signal Priority with Bus Speed Volatility as a Surrogate Measure: Case Study in Minnesota. pp 479–490, https://doi.org/10.1061/9780784483534.041

Strathman J, Wachana P, Callas S (2010) Analysis of bus collision and non-collision incidents using transit ITS and other archived operations data. Journal of Safety Research 41:137–144. https://doi.org/10.1016/j.jsr.2010.02.003

Yin S, Fischer L, Wang Y, et al (2024) Business Case for Transit Bus Automated Collision Avoidance Systems. Transportation Research Record: Journal of the Transportation Research Board 2678:891–898. https://doi.org/10.1177/03611981231176810

Zheng L, Yang Y, Ding T, et al (2024) Corporate driver management perspectives: A PLS-SEM-based analysis of the mechanism of bus crash injury severity factor effects. Heliyon 10:e39959. https://doi.org/10.1016/j.heliyon.2024.e39959

Zhu T, Qin D, Jia W (2023) Examining the associations between urban bus drivers' rule violations and crash frequency using observational data. Accident Analysis & Prevention 187:107074. https://doi.org/10.1016/j.aap.2023.107074