Traffic Safety Data Visualization: Examining Contributing Factors, Crash Density, and Impact Outcomes

Thi Ngan Ha Do
Faculty of Information Technology
Swinburne University of Technology
Melbourne, Australia
103128918@student.swin.edu.au

***Abstract*— This paper presents a visualization-based approach to analyzing key factors contributing to road traffic accidents (RTAs) within urban settings. By examining a comprehensive dataset, this study addresses three main questions: the influence of driver behavior, infrastructure, and environmental factors on accident frequency and severity; the spatial and temporal distribution of traffic crashes and its implications for police response; and the factors affecting vehicle damage costs and injury outcomes. The report outlines data sources, processing techniques, and an in-depth visual analysis focused on highlighting high-risk periods and locations. Visual analysis is achieved through heatmaps for crash density, treemaps and stacked bar charts to illustrate contributory factors and injury outcomes, and time-based line charts to analyze crash frequency by time. These user-centric visualizations enable policymakers, safety professionals, and researchers to extract actionable insights for improved traffic safety strategies.**

Keywords—road traffic accidents, crash frequency, crash density, contributory factors, spatio-temporal analysis, damage cost, injury outcome, data visualisation

I. INTRODUCTION

A. Background and Motivation

Road traffic accidents (RTAs) impose significant human and economic costs, affecting both individuals and society as a whole. These incidents lead not only to immediate health repercussions but also contribute to substantial economic burdens due to healthcare costs, loss of productivity, and damage to property [1]. The severity of traffic accidents extends beyond the physical injuries, often resulting in a range of social and psychological issues, as well as extensive expenses related to recovery and repair of the damaged environment and infrastructure [2]. Given the vast loss of life and property associated with RTAs globally, road safety has become a critical socio-economic issue, emphasizing the need for a thorough understanding and improvement of traffic safety systems within transportation engineering.

On the one hand, RTAs represent a profound human cost, claiming lives and leaving many with lasting physical and psychological impacts. In the United States, deaths due to roadway crashes have remained one of the leading causes of unnatural fatalities for decades, highlighting the ongoing risk to public safety [3]. The World Health Organization reports that approximately 1.35 million people worldwide lose their lives in RTAs each year, with an average of 3,700 deaths daily, and an additional 20 to 50 million individuals suffer non-fatal injuries, many resulting in lifelong disabilities. Regional statistics further emphasize the human toll of RTAs, with death rates per 100,000 population reaching 15.6 in America, 9.3 in Europe, and a striking 20.7 in Southeast Asia [4]. Particularly vulnerable are children and adolescents, for whom motor vehicle accidents have been the deadliest cause of death in the U.S. from 1999 to 2020. Even with a reduced death rate of around 5 per 100,000 by 2020, these incidents remain a significant threat to young lives [5].

On the other hand, RTAs impose a substantial economic burden on society, impacting various sectors through both direct and indirect costs. Financially, RTAs result in significant losses due to reduced productivity, medical expenses, and property damage, as well as nonmonetary costs associated with the pain and suffering endured by victims and their families [6]. In the United States, the economic impact of traffic incidents is particularly severe; in 2020 alone, approximately 40,000 fatalities and over 2.1 million emergency room visits due to car accidents contributed to an estimated $430 billion in costs, factoring in medical expenses and the diminished quality of life [7]. While efforts to improve road safety have shown progress, RTAs continue to result in around 1.35 million deaths worldwide each year, with an annual economic cost exceeding $65 billion USD.

The causes of RTAs are complex and multifactorial, with interrelated factors contributing to the rising incidence of crashes. Key factors include an inadequate road infrastructure that fails to accommodate growing traffic demands, leading to congestion and increased collision risk [8]. Additionally, the rapid surge in motorized vehicle ownership has intensified traffic density, further heightening accident probabilities. Insufficient road safety policies and enforcement also play a critical role, as they allow unsafe driving practices to persist unchecked. Moreover, reckless driving behaviors, including speeding, distracted driving, and disregard for traffic rules, significantly contribute to crash rates [9].

Urban areas are often associated with unique patterns and characteristics in RTAs due to high population density, diverse traffic flows, and complex road networks [10]. As a major urban center, Chicago experiences these challenges, with its dense infrastructure, large volume of motor vehicles, and high pedestrian activity creating a dynamic and sometimes hazardous traffic environment [11]. Given these conditions, analysing RTAs in Chicago is crucial for understanding the specific factors that contribute to crashes in urban contexts. Such an analysis could inform targeted interventions, enabling city planners and policymakers to address the safety concerns of urban settings such as Chicago and work towards reducing accident rates and improving overall road safety.

B. Project Objectives

The primary objective of this project is to leverage visual analysis techniques to enhance the understanding of RTA patterns and contributing factors. Recognizing that the effectiveness of accident prevention methods relies heavily on the accuracy of data and the appropriateness of analytical approaches [12], this study aims to explore various dimensions of RTA data and aid in the answer of pertinent questions in the area of RTAs, such as:

- Is it possible to identify key factors in driver behavior, infrastructure, and environmental conditions that contribute to the frequency and severity of traffic accidents?
- Are there observable patterns in the distribution of traffic crashes across different locations and times, and how do these patterns inform police response effectiveness?
- What factors most significantly impact the extent of vehicle damage and injury severity in traffic crashes?

By addressing these key questions through data visualization, this project aspires to provide insights that can inform targeted interventions and guide improvements in traffic safety.

C. Structure

This report is structured as follows. Section 2 provides the theoretical foundation for key concepts discussed. Section 3 reviews relevant studies. Section 4 outlines the methods used in this research, while Section 5 presents the results and analysis. The report concludes with Section 6, summarizing findings and suggesting directions for future work.

II. THEORETICAL FOUNDATION

A. Concept Definitions

RTA is defined as an incident on a public road involving at least one vehicle, resulting in injury or fatality [6]. Such accidents are complex events influenced by multiple interacting factors, including vehicle and driver characteristics, road infrastructure, environmental conditions, and traffic control elements [13].

Crash frequency refers to the number of accidents occurring within a specified period, often used as a key metric to assess accident trends on highways or particular road sections [14], while crash severity is categorized by the extent of damage and human impact, ranging from property damage only (PDO) to more serious outcomes such as incapacitating injuries and fatalities [15].

B. Visual Analysis

Visual analytics is the science of combining data analysis, visualization, and user interaction to enhance understanding of complex data through interactive visual interfaces [16]. By merging human and computational insights, it enables users to explore data, identify patterns, and make informed decisions efficiently [17].

This project applies various visualization techniques to make complex road traffic accident data more accessible and meaningful. Selected techniques highlight key patterns, such as time-based accident frequency, spatial hotspots, and response times by severity, ensuring both categorical and numerical data are effectively represented in line with visual analytics principles.

Given the time-series characteristics of our dataset, we chose heat maps for our spatial data visualisation, as this method is widely used for interpreting spatial patterns. The heat map technique is an efficient method for visualizing geospatial data on a map, using color variations to indicate areas with varying point densities, which highlights patterns in distribution and intensity [18].

A stacked area chart, similar to a stacked bar chart, represents the relationship between individual components and the cumulative whole [19]. Within road traffic accident analysis, this visualization technique proves valuable in illustrating how proportions of various injury types fluctuate across age groups, while also conveying a comprehensive view of overall injury outcomes.

Our visual analysis also relies on the usage of bubble plots, which is a variation of a scatter plot where markers are replaced with bubbles, allowing it to display relationships among at least three variables [20]. This type of chart effectively highlights areas of high activity, making it useful for visualizing data distribution across space [21]. Furthermore, varying the color intensity of bubbles enhances visual clarity, enabling users to quickly identify trends and patterns, which makes bubble charts particularly suitable for illustrating relationships and distinctions within complex datasets involving both categorical and numerical data.

Using interactive methods, the analysis aims to enhance understanding and support decision-making in road safety, by integrating theoretical concepts, principles, and data visualization methods learned in our coursework. Our visualizations prioritize conceptual rigor over aesthetics alone, strengthening our capacity to derive meaningful insights from the data.

III. RELATED LITERATURE

A. Road Traffic Accident Contributory Factors

Table I. Road Traffic Accident Contributory Factors

Article's title	***Author***	***Year***	***Result***
Managing Road Traffic Accidents: A Review on Its Contributing Factors [22]	Cruz, Orlean G. Dela, Juland A. Padilla, and Armando N. Victoria	2021	This paper reviews factors contributing to road traffic accidents, identifying driver behavior, road infrastructure, and vehicle conditions as primary influences.
Identification of risk factors influencing road traffic accidents [23]	Touahmia, Mabrouk	2018	This paper identifies key risk factors contributing to road traffic accidents as human behavior, road conditions, and inadequate compliance with safety regulations.
Road traffic accident data analysis and its visualization [24]	Rabbani, Muhammad Babar Ali	2021	This study identifies key contributing factors as driver demographics, vehicle types, accident locations, and timing, highlighting high-risk groups and accident hotspots.

Road accidents often result from a combination of systemic issues within the driving ecosystem, which includes factors related to vehicles, road infrastructure, and road users, along with the interactions between these elements [23, 25]. Among these, demographic factors and human errors play a central role, with most incidents occurring on urban roads where non-compliance with speed limits is common [23]. Furthermore, males are disproportionately affected as victims in these accidents, emphasizing how driver demographics and behaviors, like speeding, are critical risk factors.

Driver behavior, particularly at intersections, also contributes to accident frequency. Left-turn and crossing conflicts are frequent at unsignalized intersections, while rear-end and left-turn crashes occur most often at signalized intersections, reflecting the unique risks that different intersection types pose [22].

Road infrastructure and environmental factors further impact crash rates. Approximately one-third of accidents are linked to road conditions, including poor pavement, inadequate lane markings, and visibility issues, underscoring the need for well-designed infrastructure. Weather also plays a role, with studies indicating that adverse conditions like rain increase crash likelihood [14].

In existing data visualization works such as [24], the methodology included visual tools such as pie charts to represent accident types, bar charts for demographic and time-based analysis, and heat maps to identify accident hotspots on geographic maps.

B. Traffic crash density and Police response times

Table II. Traffic crash density and Police response times

Article's title	*Author*	*Year*	*Result*
Analysis of intra-urban traffic accidents using spatiotemporal visualization techniques [9]	Soltani, Ali, and Sajad Askari	2014	This paper conducts a spatiotemporal analysis of intra-urban traffic accidents, identifying accident hotspots and high-risk hours during traffic congestion.
Modeling crash frequency and severity with spatiotemporal dependence [26]	Chiou, Yu-Chiun, and Chiang Fu	2015	This paper investigates the role of time and location and examines the relationship between crash occurrences and police response times
Exploration of the police response time to motor-vehicle crashes in Pennsylvania, USA [27]	Liu, Chenhui	2022	This paper investigates the correlation between police response times and crash outcomes, finding that delayed responses are linked to increased fatalities.

Both studies [9, 28] agreed that spatial and temporal factors significantly influence crash frequency and severity. Crash occurrences were found to be concentrated near major arterial roads and tend to vary by time of day, with higher crash rates during periods of heavy traffic. Areas with increased traffic speeds and volumes are more likely to experience higher accident severities [9]. Concerning the seasonal effects on crash frequency, there is evidence that holiday months and summer seasons may present increased crash risks due to higher tourist activity and congestion.

Article [27] highlights a clear link between police response delays and crash outcomes, particularly regarding injury severity and fatalities. Longer police response times correlate with an increased likelihood of fatal outcomes, both at individual and county levels. Factors influencing police response times include illumination, weather, and area type, with crashes at night, in adverse weather, or in rural areas typically experiencing slower responses. These delays may compound injury severity and fatality risks, as prompt police intervention is critical for scene management and facilitating medical aid.

Existing visualization works incorporated techniques such as kernel density estimation, which highlights high-density crash zones, effectively identifying accident hotspots in urban areas, and heat maps that illustrate spatial distribution incorporate map animations and co-maps, which provide a visual breakdown of crash incidents across sequential time intervals, helping to observe temporal variations and patterns.

C. Damage Cost and Injury Outcomes

Table III. Damage Cost and Injury Outcomes

Article's title	*Author*	*Year*	*Result*
Innovative approaches in assessing social and economic damage from road accidents [29]	Bychkov, V. P., M. V. Drapalyuk, I. Yu Proskurina, and E. N. Busarin	2018	This paper explores methods for assessing the social and economic costs of road accidents, considering factors such as vehicle damage, injury severity across age groups, and infrastructure impact.
The methodology for calculation of road accident costs [28]	Pukalskas, Saugirdas, Robertas Pečeliūnas, and Vigilijus Sadauskas	2015	This paper presents a methodology for calculating the costs associated with road accidents, considering factors such as human health impacts, property damage, investigation costs, and lost productivity.
Human factors contributing to the road traffic accident occurrence [25]	Bucsuházy, Kateřina, Eva Matuchová, and Robert Zůvala	2020	This paper analyzes injury outcomes in road traffic accidents, focusing on demographic variables to better understand accident causation.

The economic costs associated with RTAs encompass both direct and indirect expenses [29]. Direct costs include measurable outlays such as vehicle repairs, medical treatment, and emergency services, while indirect costs reflect the broader societal impact, including productivity losses resulting from fatalities, injuries, and disabilities. Human health costs, in particular, account for reduced workforce productivity, temporary or permanent disabilities, and the social and moral tolls associated with these incidents [26].

[25] highlight demographic factors affecting injury severity in RTAs. Young drivers, particularly those under 25, are more frequently involved in accidents due to inexperience and a propensity for high-risk behaviors, such as speeding. Conversely, older drivers (over 65) often face accidents linked to age-related cognitive and physical decline, along with medical conditions that may impair driving abilities. Gender differences further contribute, with male drivers more likely to take aggressive risks and misjudge road scenarios, while female drivers may experience accidents related to inattentiveness or panic reactions.

Regarding data visualization, the articles emphasize the use of statistical tables and figures to represent accident costs by category, severity, or affected demographic groups. They do not, however, explore advanced visualization techniques specific to analyzing spatio-temporal patterns or crash-type comparisons.

IV. METHODOLOGY

Data processing is a critical step to ensure accuracy and reliability in the insights derived. For this project, data was consolidated from three distinct sources. By integrating these sources, the analysis benefits from a more comprehensive view, allowing for a deeper examination of patterns and factors that influence crash occurrences and outcomes.

A. Data Sources

This project applied three datasets, all sourced from the City of Chicago's publicly accessible portal, each capturing distinct aspects of RTAs and contributing to a general analysis of crash impacts:

Traffic Crashes – Crashes

The Crashes dataset records all police-reported RTAs under the jurisdiction of the Chicago Police Department, utilizing an electronic reporting system for accuracy and timeliness. According to Illinois law, only incidents involving property damage over $1,500 or any bodily injury on a public road involving at least one moving vehicle are considered reportable crashes.

Traffic Crashes – People

The People dataset provides detailed information on individuals involved in each crash, whether they were vehicle occupants, pedestrians, bicyclists, or users of other non-motorized transportation. Each record within the People dataset corresponds to an occupant recorded in the Crashes dataset and includes details on any injuries sustained, as documented by the police at the crash scene.

Traffic Crashes – Vehicles

The Vehicles dataset contains data on each vehicle, or "unit," involved in a crash, covering motorized vehicles, bicycles, and pedestrians. Each vehicle or non-motorized unit is recorded as an independent entry, capturing a diverse range of transportation modes and allowing for a detailed analysis of each vehicle type's involvement in crash scenarios.

B. Data processing

The data processing workflow in this study involves multiple stages, progressing from raw data collection to the creation of processed datasets, which are then transformed into visual symbols for effective visualization. Raw data was gathered from various sources, encompassing temporal, spatial, spatio-temporal, and multivariate properties essential for comprehensive analysis [30]. The cleaning process typically includes three main phases: initially auditing the data to detect any inconsistencies, then determining appropriate transformations to address these issues, and finally applying these transformations to refine the dataset for usability.

Following this workflow, several key steps were taken to prepare the data for visualization:

Rows with missing values exceeding 10% were removed, and nominal column gaps were filled with "False" to handle missing data effectively. Date formats were standardized from "mm/dd/yyyy" to "dd/mm/yyyy", and duplicate records, filtered by unique identifiers, were eliminated to maintain dataset integrity.

To optimize the dataset for Tableau, large files were downsized by selecting every third row, preserving essential primary – foreign references. Three datasets from several sources were merged using a common referencing key "crash_record_id" to provide a homogenous dataset.

Data aggregation is effective in reducing the data size and provides convenience in subsequent analysis [30]. To optimize the dataset for Tableau, large files were downsized by selecting every third row, preserving essential primary – foreign references. Three datasets from several sources were merged using a common referencing key "crash_record_id" to provide a homogenous dataset.

Feature selection further streamlined the dataset by excluding repetitive or unnecessary columns from both the crashes and vehicles datasets, reducing file size without losing crucial information.

Calculated fields were created to enhance analysis, such as grouping time of day, categorizing accident severity based on injury and damage levels, and determining vehicle age groups. Additionally, formulas were applied to identify common values in categorical columns and calculate the count of specific injury types, contributing to a more insightful visualization of accident trends and contributing factors.

C. Selection of Design

1) Colour scheme: We utilized intuitive color schemes and opacity variations to differentiate key elements, such as gender, crash frequency, and crash severity. To facilitate focus and prevent information overload, each chart corresponding to distinct research questions was assigned a unique color scheme. Recognizing the challenges faced by individuals with color vision deficiencies, we selected a blue-orange palette tailored for color blindness [31], enhancing accessibility. Additionally, shapes and patterns were incorporated to supplement color, ensuring that all visuals remain clear and interpretable across user groups.

Aesthetics are maintained through a clean layout with a light background that does not detract from the data. The use of consistent and thematic color schemes enhances visual appeal and aligns with the overall professional tone of the project. While aesthetics was prioritized, the functionality and clarity of the visualizations remained the primary focus, ensuring that the design is not only visually pleasing but also effective in conveying insights.

2) Interactivity: Interactive features include filters for crash types, age, and year, allowing users to engage with the data dynamically. Additionally, on-hover tooltips add a layer of interactivity that offers detailed information without cluttering the main visualization, supporting an exploratory analysis approach.

3) Storytelling: Our visualisations used scene-based storytelling, with each dashboard focused on specific topics like contributory factors, spatio-temporal patterns, and crash impact. Interactive filters allow users to explore data dynamically, enhancing engagement and personalization. The narrative follows a logical sequence, guiding viewers from broad analyses of driver behaviors and infrastructure to detailed crash outcomes.

4) Design principles: Gestalt principles in data visualization concentrate on how people perceive and interpret visual components [32]. Our visual analysis considered Gestalt principles including proximity, similarity, and figure-ground to enhance visual coherence. Related charts are grouped together to allow for an intuitive understanding of connections between variables.

V. ANALYSIS OF SCENARIOS

A. Contributory Factors Analysis: Driver Behaviour, Infrastructure, and Environmental Conditions

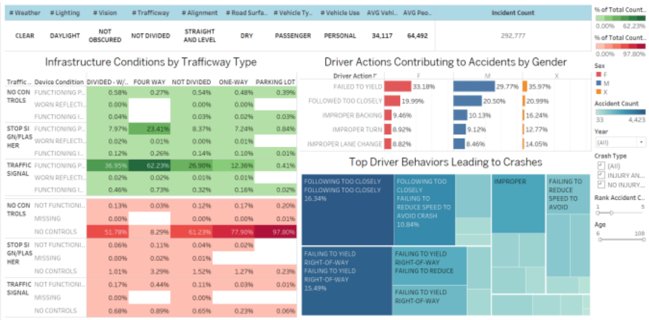

Fig. 1. Contributory Factors Analysis: Driver Behaviour, Infrastructure, and Environmental Conditions

The dashboard integrates textual and visual elements to provide structured context and support the contributory factors analysis of traffic crash data. The top multi-row cards, a visualization technique for grouping and summarizing values [33], provide an overview of crash characteristics, covering factors such as environmental conditions, road types, and vehicle categories, giving users an accessible overview of conditions frequently associated with crashes. Regarding weather conditions, the majority of recorded accidents occurred under clear weather, dry road surfaces, and daylight conditions. This finding contrasts with existing studies, which suggest a positive correlation between rainfall, wet road conditions, and accident frequency [34]. Additionally, the analysis identifies undivided trafficways and straight-level road alignments as particularly hazardous, contributing significantly to road traffic accidents. This aligns with surveys conducted by [23], where 63% of participants reported experiencing road crashes, with 67% of these incidents attributed to human error, 29% to road conditions, and 4% to vehicle defects.

This visual analysis follows a Z-pattern layout, guiding viewers to begin scanning from the top left, proceed horizontally to the top right, then diagonally down to the bottom left, and finally move horizontally to the bottom right [35]. A high-level summary is positioned at the top to establish context and present an overview of the typical conditions under which incidents occur. Below this summary, three visualizations provide a more in-depth exploration of the data, examining the interactions between road infrastructure and safety device usage, demographic factors, and driver behaviors as assessed by on-site police.

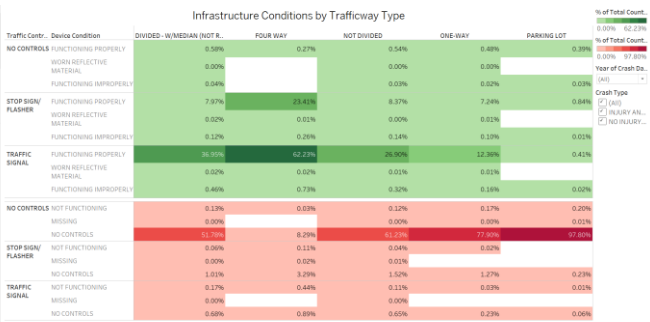

Fig. 2. Heatmap - Infrastructure Conditions by Trafficway Type

The heatmap illustrates the relationship between road infrastructure and crash frequency. The quality of road infrastructure, categorized by trafficway type, is evaluated based on its traffic control measures and the state of safety device. The visualization reveals that the high incident rates on undivided and straight-level trafficways, particularly where traffic signals are not functioning or entirely absent. This suggests a correlation between insufficient traffic control devices and elevated crash rates, emphasizing infrastructure as a crucial contributory factor in road safety.

Table IV. Data Encoding Heatmap - Dashboard 1

Attributes	***Mark***	***Channel***	***Encoding***
Trafficway Type	Rectangles	Positions along the x-axis	Encoded using positions along the x-axis
Traffic Control Device	Rectangles	Positions along the y-axis	Encoded using positions along the y-axis
Device Condition	Rectangles	Secondary categorization on the y-axis. Colour	Distinguish between functioning and non-functioning conditions
Percentage of Accidents	Text labels	Colour intensity	Darker shades indicate higher percentages

The heatmap design utilizes a green-to-red color gradient to represent crash frequency, which supports quick identification of high-risk areas through intuitive color coding. In line with the "maximize the data-ink ratio" principle, data elements are minimized to maintain clarity and comprehensiveness. The chart's dual axes provide a multidimensional view, following the details-on-demand approach. The varying color saturation also aids in grouping related data, allowing users to examine both traffic control conditions and road configurations linked to higher crash rates.

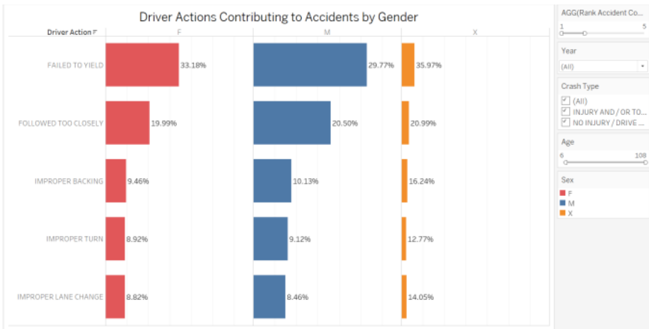

Fig. 3. Bar Chart - Driver Actions Contributing to Accidents by Gender

The bar chart provides insights into the influence of driver behavior on road accidents, segmented by gender. Key observations indicate that males are most frequently involved in traffic incidents [25], with hazardous driving behaviors such as "Failure to Yield" and "Following Too Closely" [12], being notably prevalent among both male and female drivers.

Table V. Data encoding Bar Chart - Dashboard 1

Attributes	***Mark***	***Channel***	***Encoding***
Driver Action	Bar	Positions along the y-axis	Encoded using positions along the y-axis

Sex	Bar	Positions along the x-axis. Colours coded	Encoded using positions along the x-axis and colours to differentiate between gender groups
Percentage of Accidents	Text label	Length of bars	Encoded using the length of bars and text labels

This visualization serves a narrative purpose by guiding viewers through an exploration of driver behavior patterns and their safety implications. The stacked graph is a widely utilized visualization method, effective for displaying both the individual values of each variable and their proportions relative to the total in an accumulated form along the y-axis [30]. Distinct colors are applied for each gender category to facilitate clear differentiation and enable identification of behaviors linked to each group. The horizontal bar layout supports comparative analysis of behavior frequencies across genders, with intensified colors highlighting the most prevalent risky behaviors.

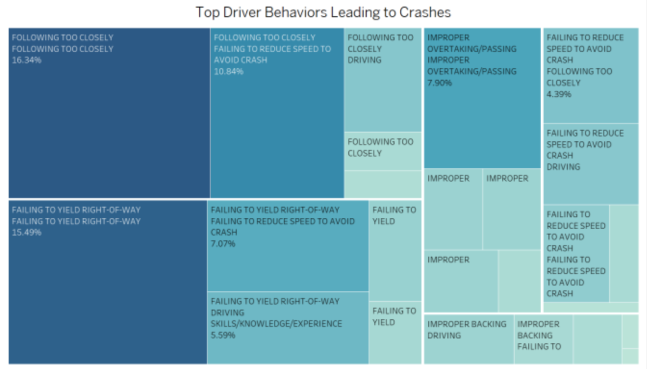

Fig. 4. Treemap - Top Behavioural Contributory Causes Leading to Crashes

The treemap visualization highlights the most common driver behaviors that contribute to crashes, "Following Too Closely" and "Failing to Yield Right-of-Way". This insight aligns with the broader theme of contributory factors in road safety, demonstrating the impact of specific driving actions on accident rates. The size of each box corresponds to the frequency of each behavior, effectively summarizing the primary risky actions that lead to road incidents.

Table VI. Data Encoding Treemap - Dashboard 1

Attributes	***Mark***	***Channel***	***Encoding***
Primary Contributory Cause	Text label	Positions within the treemap	Encoded using the text labels inside the rectangles
Secondary Contributory Cause	Text label	Positions within the treemap	
Percentage of Accidents	Rectangle sizes, colour intensity	Size of the rectangles. Colour coded	Encoded using the size of the rectangles and text labels

This visualization plays a key role in the narrative by highlighting driving behaviors that have a substantial impact on road safety, underscoring areas that could benefit from focused interventions. The treemap design showcases hierarchical data with nested shapes, where each area is scaled according to the values of numeric attributes [36]. Additionally, a monochromatic color palette with varying shades is applied to reflect the intensity of each behavior, which aids in visual clarity and ensures a clean, organized appearance. By clustering similar behaviors together, the treemap leverages principles of human perception to enhance understanding.

B. *Spatio-Temporal Analysis: Traffic Crash Density and Police Response Times*

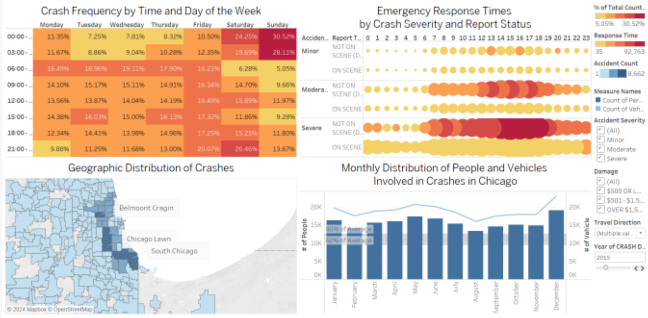

Fig. 5. Spatio-Temporal Analysis: Traffic Crash Density and Police Response Times

The Spatio-Temporal dashboard provides insights into the density of traffic crashes across various times and locations, highlighting high-risk periods and areas with the most significant accident frequency. The heatmap on the top left identifies peak crash times throughout the week, showing a notable increase in incidents on weekends and during late hours of the day [26]. The geographic distribution map pinpoints neighborhoods in Chicago with higher crash densities. The bubble chart on the top right correlates police response times with crash severity, illustrating how longer response times are associated with more severe accidents [27], underscoring the importance of prompt emergency response. Lastly, the bar chart at the bottom right displays monthly fluctuations in the number of people and vehicles involved in crashes, offering a seasonal perspective on crash trends.

This visual analysis uses an F-pattern layout, guiding viewers from the top-left corner across and down the screen [35]. The dashboard is divided into two sections for spatio-temporal analysis, with warm colors (orange) representing temporal insights and cool colors (blue) for spatial insights. The proximity principle organizes related visualizations, with temporal analysis on top and spatial analysis below, though this structure differs from previous layouts and may disrupt viewer familiarity.

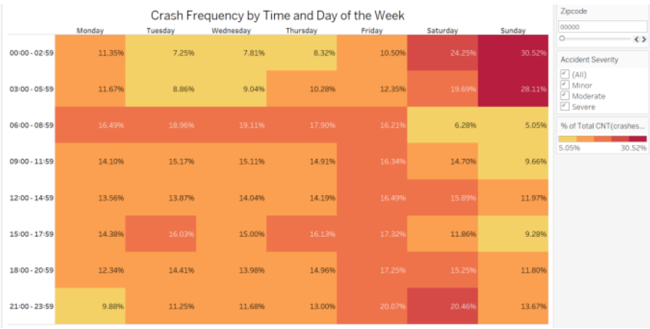

Fig. 6. Heatmap - Crash Frequency by Time and Day of the Week

The heatmap visualization effectively illustrates the density of traffic crashes across different times of day and days of the week. Key findings indicate that crash incidents are highest during weekends, with a marked increase on Saturdays and Sundays, especially in the late-night to early-morning hours (00:00 to 05:00). This pattern implies that these periods are particularly high-risk, potentially due to factors such as reduced visibility, driver fatigue, or impaired driving [26]. support this analysis, time bins were calculated based on the concept of periodic time, recognizing that many natural processes are recursive and occur in cycles, such as daily intervals [30].

Table VII. Data Encoding Heatmap - Dashboard 2

Attributes	***Mark***	***Channel***	***Encoding***
Time of the Day	Rectangles	Positions along the y-axis	Encoded using positions along the y-axis
Day of the Week	Rectangles	Positions along the x-axis	Encoded using positions along the x-axis
Percentage of Accidents	Colour, text labels	Colour intensity	Darker shades indicate higher percentages

From a design perspective, the heatmap uses a gradient color scheme, ranging from lighter shades to intense red, which signifies areas of greater crash frequency. The design balances clarity and emphasis to highlight relevant data points while maintaining a comprehensive view of all time intervals. As part of the dashboard's narrative, this heatmap guides users through a spatio-temporal analysis, emphasizing high-risk periods and aiding in understanding how time-based factors correlate with crash likelihood.

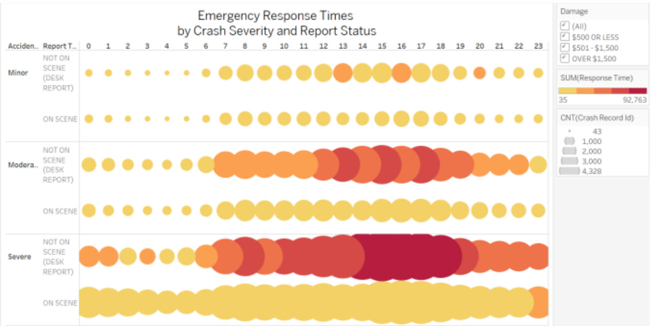

Fig. 7. Bubble Chart - Emergency Response Times by Crash Severity and Report Status

The bubble chart illustrates police response times based on crash severity and report status. The longer response times associated with severe crashes, particularly when the report status is on scene, suggesting that critical incidents not only require more immediate attention but also that response delays are more pronounced in severe cases, potentially impacting survival rates and injury outcomes [27].

Table VIII. Data Encoding Bubble Chart - Dashboard 2

Attributes	***Mark***	***Channel***	***Encoding***
Crash Severity	Circles	Positions along the y-axis	Encoded using positions along the y-axis
Report Type	Circles	Positions along the y-axis (secondary level)	Encoded using positions along the y-axis
Crash Hour	Circles	Positions along the x-axis	Encoded using positions along the x-axis
Number of Accidents	Circle sizes	Size of circles	Larger circles represent higher counts of crashes
Response Time	Colour	Colour intensity	Darker shades indicate higher percentages

In terms of design, this design choice follows visual encoding principles. The placement of severity levels in a descending order (minor to severe), and the consistent use of color shades supports the dashboard's narrative, emphasizing the critical role of timely response in severe incidents.

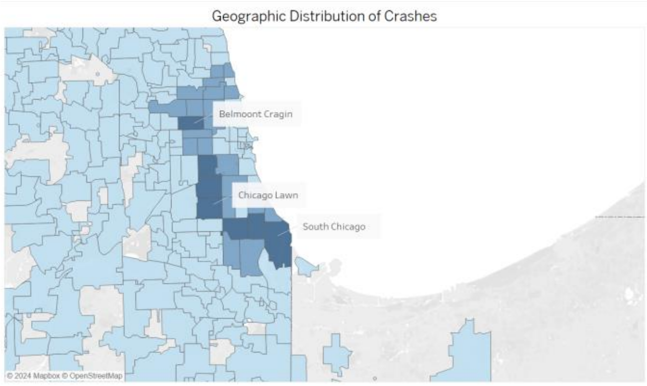

Fig. 8. Map - Spatial Distribution of Crashes in Illinois

This geographic heatmap offers an overview of traffic crash density across Chicago neighborhoods, with darker shades representing areas with higher crash frequencies. Observations from the map show a concentration of incidents in certain neighborhoods, namely Belmont Cragin, Chicago Lawn, and South Chicago.

The visualization utilizes a blue color gradient, from light to dark, to intuitively communicate crash density, adhering to human perception principles that associate color intensity with frequency. Labeling high-crash neighborhoods enhances accessibility and comprehension, allowing users to quickly identify key areas of interest. Overall, this spatial visualization contributes to the narrative by guiding users through a visual exploration of crash hotspots, fostering a clearer understanding of the city's accident distribution and prompting considerations for localized safety measures.

C. Crash Impact Analysis: Vehicle Damage and Injury Outcomes

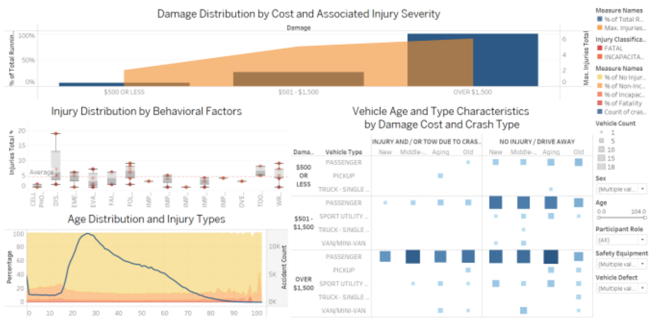

Fig. 9. Crash Impact Analysis: Vehicle Damage and Injury Outcomes

This dashboard provides a general analysis of the impact of vehicle crashes on damage costs and injury outcomes, highlighting trends related to vehicle type, driver behavior, and injury severity. The main insights indicate that higher damage costs, particularly over $1,500, correlate with severe injuries. Vehicle types such as sport utility and pickups may have a more significant impact on both financial and health outcomes in crashes. The injury distribution by behavioral factors and the age distribution of injury types further emphasizes the role of driver behavior and age in crash severity.

The dashboard is structured to guide viewers through different aspects of crash impact, with each section contributing to a narrative on vehicle damage and injury severity. This approach, similar to Dashboard 1, places an overview at the top, summarizing how vehicle damage costs relate to injury severity, and then divides the analysis into two distinct sections below. The left side focuses on injury outcomes, highlighting contributing behavioral factors and age distribution among victims, while the right side delves into vehicle damage cost analysis, including breakdowns by vehicle type and crash type. However, the lack of clear enclosure around related visualizations and the imbalance created by the smaller size of the area chart affect the layout's continuity and symmetry.

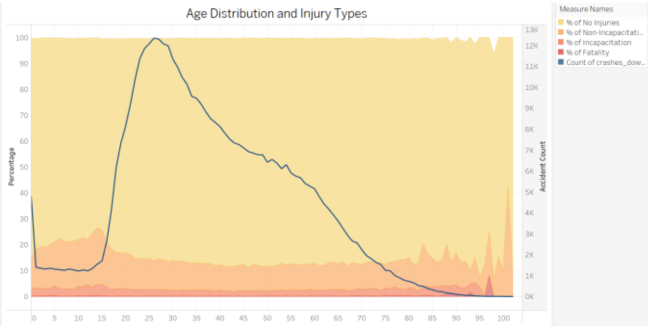

Fig. 10. Stacked Area Chart - Age Distribution and Injury Types

This stacked area chart presents the relationship between age and types of injuries sustained in crashes. Younger drivers, especially those between ages 20 and 35, experience a high frequency of crashes, with a significant proportion resulting in non-incapacitating injuries. As age increases, the incidence of crashes and the rate of injury severity decrease, yet incapacitating injuries and fatalities show a gradual increase among older age groups, reflecting greater vulnerability.

Table IX. Data Encoding Stacked Area Chart - Dashboard 3

Attributes	***Mark***	***Channel***	***Encoding***
Age	Area, line	Positions along the x-axis	Encoded using positions along the x-axis
Injury Types	Area, colour	Height of area, colour intensity	Encoded using the height of the stacked areas, darker shades indicate more severe injury outcomes
Number of Accidents	Line	Height of line	Encoded using the height of the line chart

In terms of design, a stacked area chart was chosen to emphasize both the percentage distribution of injury types by age and the overall accident count. The warm color gradient, transitioning from light orange for non-injury cases to darker reds for more severe injuries, visually encodes injury severity, following the idiom of color intensity to signify increased risk. The line chart overlay for accident count reinforces the peak ages for crash incidents, providing a dual layer of insight. The narrative is further supported by this clear, comprehensive approach, which guides viewers to understand the dual impact of age on crash frequency and injury severity.

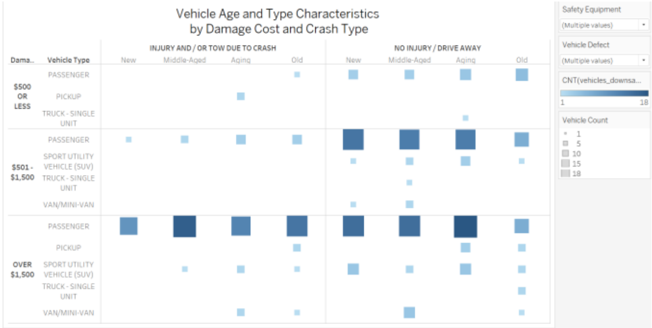

Fig. 11. Heatmap - Vehicle Age and Type Characteristics by Damage Cost and Crash Type

This heatmap provides insights into how vehicle type, age, and damage cost correlate with crash outcomes. The heatmap-style matrix categorizes vehicles by age groups and types across different damage cost tiers and crash consequences. A notable pattern is observed in the higher frequency of significant damage costs associated with passenger and pickup vehicles, particularly in older age categories. This distribution suggests that vehicle age and type might influence both the cost of damage and the likelihood of severe crash outcomes.

Table X. Data Encoding Heatmap - Dashboard 3

Attributes	***Mark***	***Channel***	***Encoding***
Damage	Squares	Positions along the y-axis	Encoded using positions along the y-axis
Vehicle Types	Squares	Positions along the y-axis	Encoded using positions along the y-axis
Vehicle Age Groups	Squares	Positions along the x-axis	Encoded using positions along the x-axis
Crash Types	Squares	Secondary categorization on the x-axis	Encoded as a secondary variable on the x-axis
Number of Vehicles	Square sizes, colour intensity	Size and colour of squares	Larger and darker squares represent higher counts of vehicles

From a design perspective, the layout, arranged by vehicle age and type across damage tiers, follows principles of proximity and continuity, grouping similar data points to facilitate logical interpretation. As a narrative device, this visualization emphasizes the role of vehicle characteristics in crash severity and cost outcomes.

VI. CONCLUSION

A. Results

The visualizations in this project effectively highlighted key factors influencing road traffic accidents, such as driver behavior, infrastructure, and environmental conditions. Dashboard 1 illustrated the impact of infrastructure and driver actions, showing that most accidents occur on undivided roads in clear conditions, often due to risky behaviors like failing to yield. Dashboard 2 identified high crash rates during late-night hours on weekends, with concentration in areas like Belmont

Cragin, and highlighted how delayed police response correlates with crash severity. Dashboard 3 linked injury severity and damage costs, showing older vehicles and behaviors like improper turns were associated with higher costs and severe injuries. These insights emphasize where targeted safety interventions could reduce crashes.

B. Limitations

This analysis faced limitations, such as data density issues in heatmaps and inconsistencies in color schemes across dashboards, which could challenge interpretation. Future improvements could include streamlining color usage for clarity, adding hover-over details to reduce clutter, and using more granular geographic data to support policymakers in targeting high-risk areas. Incorporating predictive analytics could also enhance these dashboards' effectiveness for safety planning and resource allocation.

References

[1] Abdulhafedh, A. (2016). Crash frequency analysis. Journal of Transportation Technologies, 6(04), 169.

[2] Ahmed, B. Y., Majeed, Z. H., & Najim, S. A. (2020). Applying the visualization technique to solve the human color blindness.

[3] Ahmed, S. K., Mohammed, M. G., Abdulqadir, S. O., El‐Kader, R. G. A., El‐Shall, N. A., Chandran, D.,⋯ Dhama, K. (2023). Road traffic accidental injuries and deaths: A neglected global health issue. Health science reports, 6(5), e1240.

[4] Ajibade, S. S., & Adediran, A. (2016). An overview of big data visualization techniques in data mining. International Journal of Computer Science and Information Technology Research, 4(3), 105-113.

[5] Ala'a, B., Altarawneh, S. A., Ghazaleh, L. Z. A., Jaradat, T. M., Abujudeh, R., Almbaidin, M., & Hasan, H. (2024). Navigating Road Traffic Accidents in Jordan: A Retrospective Exploration of the Health and Economic Impacts. Cureus, 16(5).

[6] Andrienko, N., Andrienko, G., Fuchs, G., Slingsby, A., Turkay, C., & Wrobel, S. (2020). Visual analytics for data scientists. Springer.

[7] Bougna, T., Hundal, G., & Taniform, P. (2022). Quantitative analysis of the social costs of road traffic crashes literature. Accident Analysis & Prevention, 165, 106282.

[8] Bucsuházy, K., Matuchová, E., Zůvala, R., Moravcová, P., Kostíková, M., & Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. Transportation research procedia, 45, 555-561.

[9] Bychkov, V. P., Drapalyuk, M. V., Proskurina, I. Y., & Busarin, E. N. (2018). Innovative approaches in assessing social and economic damage from road accidents.

[10] Cabrera-Arnau, C., Prieto Curiel, R., & Bishop, S. R. (2020). Uncovering the behaviour of road accidents in urban areas. Royal Society open science, 7(4), 191739.

[11] Cantillo, V., Márquez, L., & Díaz, C. J. (2020). An exploratory analysis of factors associated with traffic crashes severity in Cartagena, Colombia. Accident Analysis & Prevention, 146, 105749.

[12] Chand, A., Jayesh, S., & Bhasi, A. B. (2021). Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. Materials Today: Proceedings, 47, 5135-5141.

[13] Chen, W., Guo, F., & Wang, F.-Y. (2015). A survey of traffic data visualization. IEEE Transactions on intelligent transportation systems, 16(6), 2970-2984.

[14] Chiou, Y.-C., & Fu, C. (2015). Modeling crash frequency and severity with spatiotemporal dependence. Analytic Methods in Accident Research, 5, 43-58.

[15] Christou, N. (2021). Spatial data in undergraduate statistics curriculum. Journal of Statistics and Data Science Education, 29(1), 27-38.

[16] Cruz, O. G. D., Padilla, J. A., & Victoria, A. N. (2021). Managing road traffic accidents: a review on its contributing factors.

[17] Duba, B. P. THE USE OF VISUALIZATION AND ANALYTIC TOOLS OF POWER BI IN LEARNING TASKS.

[18] García-Herrero, S., Febres, J. D., Boulagouas, W., Gutiérrez, J. M., & Mariscal Saldaña, M. Á. (2021). Assessment of the influence of technology-based distracted driving on drivers' infractions and their subsequent impact on traffic accidents severity. International journal of environmental research and public health, 18(13), 7155.

[19] Goldstick, J. E., Cunningham, R. M., & Carter, P. M. (2022). Current causes of death in children and adolescents in the United States. New England journal of medicine, 386(20), 1955-1956.

[20] Hillard, B., Amarego, J., & McGill, T. (2016). Optimising visual layout for training and learning technologies.

[21] Keim, D. A., Mansmann, F., Stoffel, A., & Ziegler, H. (2008). Visual analytics.

[22] Liu, C. (2022). Exploration of the police response time to motor-vehicle crashes in Pennsylvania, USA. Journal of safety research, 80, 243-253.

[23] Mohammed, L. T., AlHabshy, A. A., & ElDahshan, K. A. (2022). Big data visualization: A survey.

[24] Netek, R., Pour, T., & Slezakova, R. (2018). Implementation of heat maps in geographical information system–exploratory study on traffic accident data. Open Geosciences, 10(1), 367-384.

[25] Pembuain, A., Priyanto, S., & Suparma, L. (2019). The effect of road infrastructure on traffic accidents.

[26] Pukalskas, S., Pečeliūnas, R., Sadauskas, V., Kilikevičienė, K., & Bogdevičius, M. (2015). The methodology for calculation of road accident costs. Transport, 30(1), 33-42.

[27] Rabbani, M. B. A., Musarat, M. A., Alaloul, W. S., Maqsoom, A., Bukhari, H., & Waqas, R. (2021). Road traffic accident data analysis and its visualization. Civil engineering and architecture, 9(5), 1603-1614.

[28] Salman, R. E., & Alzaatreh, A. (2022). Market Basket Analysis of Chicago Road Accidents.

[29] Soltani, A., & Askari, S. (2014). Analysis of intra-urban traffic accidents using spatiotemporal visualization techniques. Transport and telecommunication journal, 15(3), 227-232.

[30] Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. Accident Analysis & Prevention, 72, 244-256.

[31] Touahmia, M. (2018). Identification of risk factors influencing road traffic accidents. Engineering, Technology & Applied Science Research, 8(1), 2417-2421.

[32] Umair, M., Rana, I. A., & Lodhi, R. H. (2022). The impact of urban design and the built environment on road traffic crashes: A case study of Rawalpindi, Pakistan. Case studies on transport policy, 10(1), 417-426.

[33] Wang, Y., & Zhang, W. (2017). Analysis of roadway and environmental factors affecting traffic crash severities. Transportation research procedia, 25, 2119-2125.

[34] Ye, Z., Xue, C., & Lin, Y. (2021). Visual perception based on gestalt theory.

[35] Zhang, H. (2010). Identifying and quantifying factors affecting traffic crash severity in louisiana. Louisiana State University and Agricultural & Mechanical College.

[36] Zhang, L., Zhang, M., Tang, J., Ma, J., Duan, X., Sun, J.,Xu, S. (2022). Analysis of traffic accident based on knowledge graph. Journal of advanced transportation, 2022(1), 3915467.