Analysis of Motor Vehicle Collision Causes in Toronto from 2006 to 2023*

Assessing the Interplay of Human behaviour and Environmental Factors in Predicting Accident Severity and High-Risk Scenarios

Jiwon Choi

September 27, 2024

This study analyzes motor vehicle collisions in Toronto from 2006 to 2023, focusing on the interaction between human behaviours—such as speeding, aggressive driving, and alcohol use—and environmental conditions like visibility and road surfaces. The findings show that while environmental factors increase risks, aggressive driving is the strongest predictors of severe accidents. These results highlight the need for targeted interventions in both driver behaviour and infrastructure to improve urban traffic safety.

1 Introduction

Motor vehicle collisions are a leading cause of injury and fatality in urban areas, driving ongoing efforts to improve road safety. In an expanding city like Toronto, understanding the causes of traffic accidents is critical for shaping effective policies. This paper examines traffic collisions in Toronto over 17 years, focusing on how human behaviour and environmental conditions influence the frequency and severity of accidents.

While much research has been done on factors such as speeding or poor road conditions, there is still a lack of understanding regarding the interaction between human behaviours and environmental conditions in creating high-risk situations. Most studies focus on these factors separately, missing how they combine to increase accident risks. This research addresses that gap by analyzing both human and environmental factors simultaneously, offering a more complete view of what leads to traffic accidents in an urban setting.

Using data from Toronto Police Services, this study analyzes the role of human factors like speeding, aggressive driving, alcohol use, and disability alongside environmental factors such

*Code and data are available at: https://github.com/jwonc4602/Analysis-of-Motor-Vehicle-Collision-Causes.

as visibility, lighting, traffic control, and road conditions. The results show that aggressive driving and alcohol use, even under good road conditions, are major contributors to accidents. High-risk scenarios, like aggressive driving on dry roads with clear visibility, emerge as leading causes of collisions. These findings suggest that targeted actions, such as stricter enforcement against reckless driving and improved traffic controls in key areas, could help reduce accident rates.

The rest of the paper is organized as follows. Section 2 outlines the data and methodology and analyzes factors contributing to traffic collisions. Section 3 discusses the results and potential interventions. The final sections of the discussion part critically evaluate the study's limitations and propose directions for future research.

2 Data

2.1 Data Source and Collection

This research is based on a dataset published by Toronto Police Services, which is accessible through the City of Toronto's OpenDataToronto Library (Gelfand 2022). The specific dataset includes all traffic collision events in Toronto where a person was either Killed or Seriously Injured (KSI) since 2006. The data was compiled and examined using the R statistical programming software (R Core Team 2023a), supplemented by various tools such as `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2023), `readr` (Wickham, Hester, and Bryan 2023), `gridExtra` (Auguié 2017), `grid` (R Core Team 2023b), `knitr` (Xie 2014), and `here` (Müller and Bryan 2020).

This traffic accident dataset contains 18,957 entries from a citywide survey, detailing various factors involved in incidents. It includes 50 variables covering geographic data, such as accident locations, and behavioral data, like driver conditions and alcohol involvement. The dataset spans several years and captures environmental conditions, road types, and the actions of drivers and pedestrians during accidents. It aims to provide a comprehensive view of research and policy efforts related to urban traffic safety and accident prevention. Although the dataset was last updated September 27, 2024, but only contains observations from 2006 to 2023.

The `acclass` category classifies the severity of traffic collisions into "Fatal," "Non-Fatal Injury," and "None or Others." The original categories "None" and "Property Damage Only" were combined into "None or Others" since both involve accidents without human injury. This consolidation simplifies the analysis, focusing on accidents with a direct human impact. "Fatal" refers to collisions causing severe injury or death, while "Non-Fatal Injury" includes incidents with injuries but no fatalities. The accident classification data shows that most collisions result in non-fatal injuries (16,268), while fatal accidents account for 2,670 cases. The 'None or Others' category has the fewest incidents, with only 19 cases recorded. This highlights the high frequency of non-fatal injuries, emphasizing the need for policies focused on injury prevention. (see Figure 1)

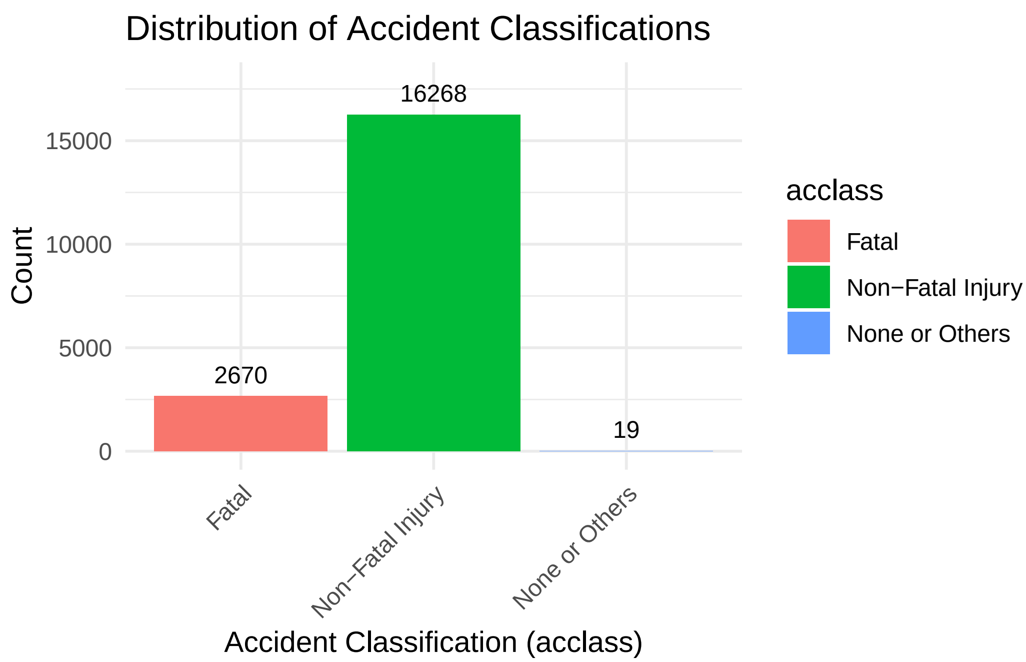

Figure 1: Distribution of Accident Classifications

2.2 Human Factors

The cleaned data is split into two sets. The first focuses on human factors, particularly driver behaviours, and conditions that may have contributed to accidents. Four key categories—`speeding`, `ag_driv`, `alcohol` and `disability`—were selected due to their known impact on traffic accidents. Each is recorded as either "None" or "Yes," indicating whether that factor was present in the incident. These categories focus on preventable human actions or conditions that are strongly linked to traffic risks. (see Table 1)

Table 1: Sample of Human Factors of Collision Data

date	speeding	ag_driv	alcohol	disability
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet

1. **Speeding**: Tracks whether speeding was a factor in the collision. Speeding reduces reaction time and increases crash severity, making it a leading cause of accidents.

2. **Aggressive Driving (ag_driv)**: Refers to reckless or distracted behaviours, such as texting while driving. These actions impair judgment and reaction time, significantly increasing accident risk.

3. **Alcohol**: Captures whether alcohol was involved in the collision. Alcohol impairs motor skills and decision-making, contributing to many severe and fatal accidents.

4. **Disability**: Notes if a medical or physical disability affected the driver's ability to control the vehicle. Disabilities can impair a driver's capacity to respond to road conditions, making this a key safety factor.

Figure 2 illustrates the frequency of human factors contributing to traffic collisions. Aggressive and disturbed driving (ag_driv) appears most frequently, with 9,836 occurrences, followed by speeding, which accounts for 2,694 cases. Alcohol use and disability are less common, with 808 and 493 incidents, respectively. Understanding these factors is important for identifying the human behaviours that increase traffic accident risks and for shaping strategies to reduce both the frequency and severity of collisions.

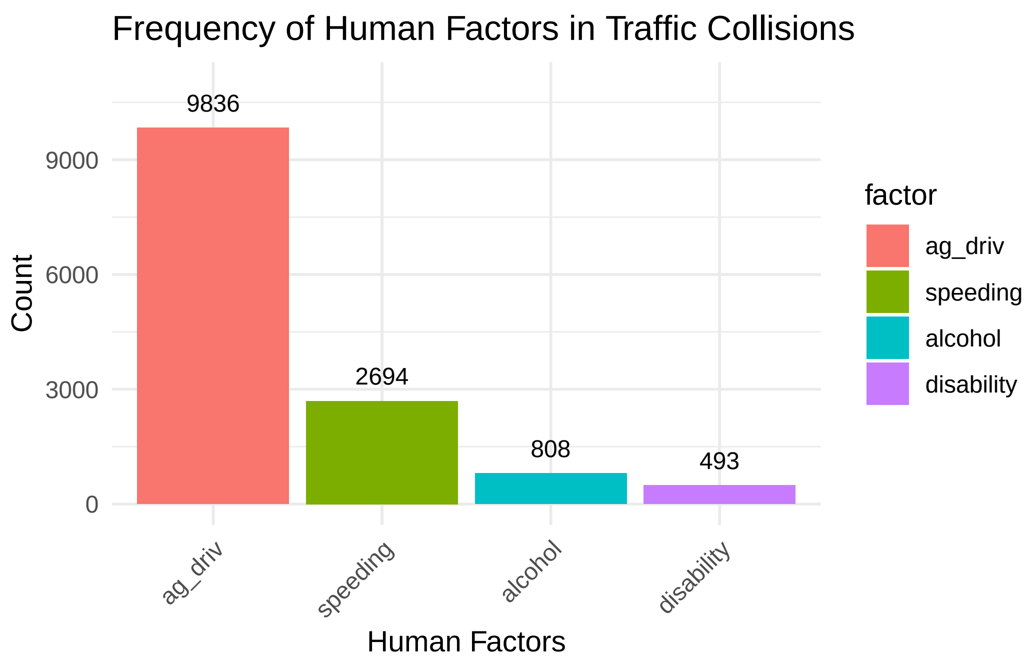

Figure 2: Frequency of Human Factors in Traffic Collisions

2.3 Environmental Factors

The second dataset focuses on environmental factors that influence traffic collisions, including conditions such as traffic control, visibility, lighting, and road surface. These elements shape driver behaviour and impact accident risk. To simplify analysis and improve consistency, categories were consolidated, providing a clearer understanding of how these factors affect accidents and where safety improvements are needed. (see Table 2)

Table 2: Sample of Environmental Factors of Collision Data

date	traffctl	visibility	light	rdsfcond
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet
2006-01-01	No Control	Clear	Dark	Wet

1. **Traffic Control Type (traffctl)**: This was simplified into "No Control," "Traffic Signal," and "Others." The "Others" group combines less common traffic controls like stop signs and pedestrian crossovers. This distinction highlights the impact of varying levels of traffic regulation on driver behaviour, focusing on scenarios where traffic control presence may influence accident risks.

2. **Visibility**: Conditions were grouped into "Clear" and "Disturbed," with "Disturbed" covering adverse weather such as fog, rain, and snow. This division highlights the contrast between ideal and hazardous visibility, which directly affects a driver's ability to navigate safely. Poor visibility is a known factor in accidents, making this an essential category for identifying environmental hazards.

3. **Light**: Lighting conditions were combined into "Dark," "Daylight," and "Dim Light." These reflect the most relevant lighting environments, with "Dark" including both natural and artificial lighting, and "Dim Light" covering intermediate conditions like dawn and dusk. Lighting affects visibility and reaction times, and this grouping helps assess its role in accidents.

4. **Road Surface Condition (rdsfcond)**: This was grouped into "Dry," "Wet," and "Others." The "Others" category includes ice, snow, and slush, conditions that greatly reduce vehicle traction. Focusing on major surface conditions helps clarify how road conditions influence accident risk, especially in wet or slippery environments.

Figure 3 and Figure 4 show that there were a similar number of collisions with no traffic control, totaling 9,021, and with traffic signals, at 8,035. The majority occurred under clear visibility with 16,373 collisions, during daylight with 10,930 collisions, and on dry road surfaces with 15,231 collisions. In contrast, there were 2,584 collisions under disturbed visibility and 729 under dim light conditions. Wet surfaces accounted for 3,140 collisions, while dark conditions saw 7,298 collisions, highlighting the importance of environmental and visibility factors in traffic accidents.

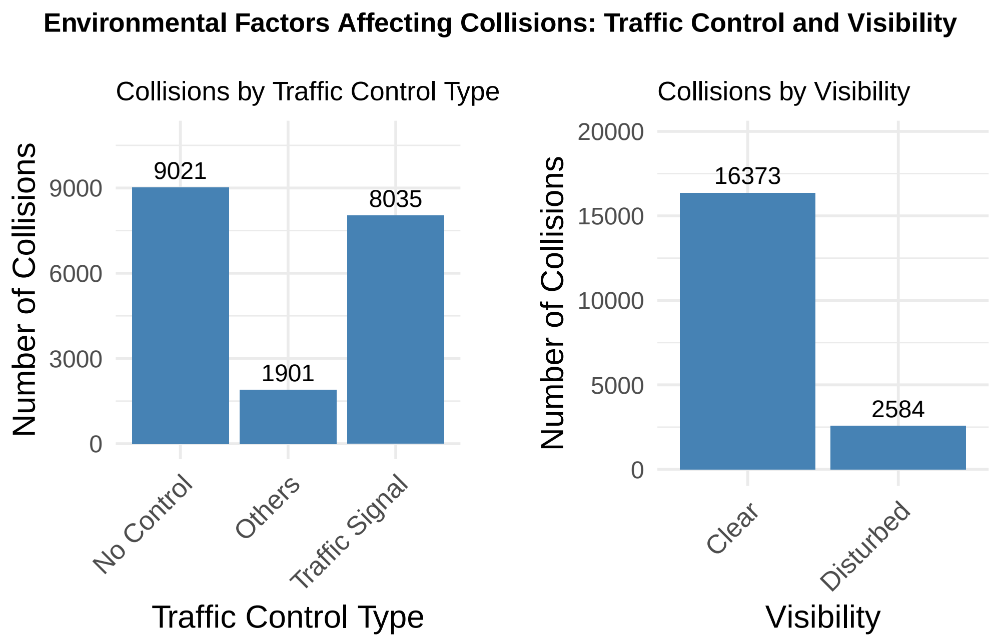

Figure 3: Environmental Factors Affecting Collisions - Traffic Control and Visibility

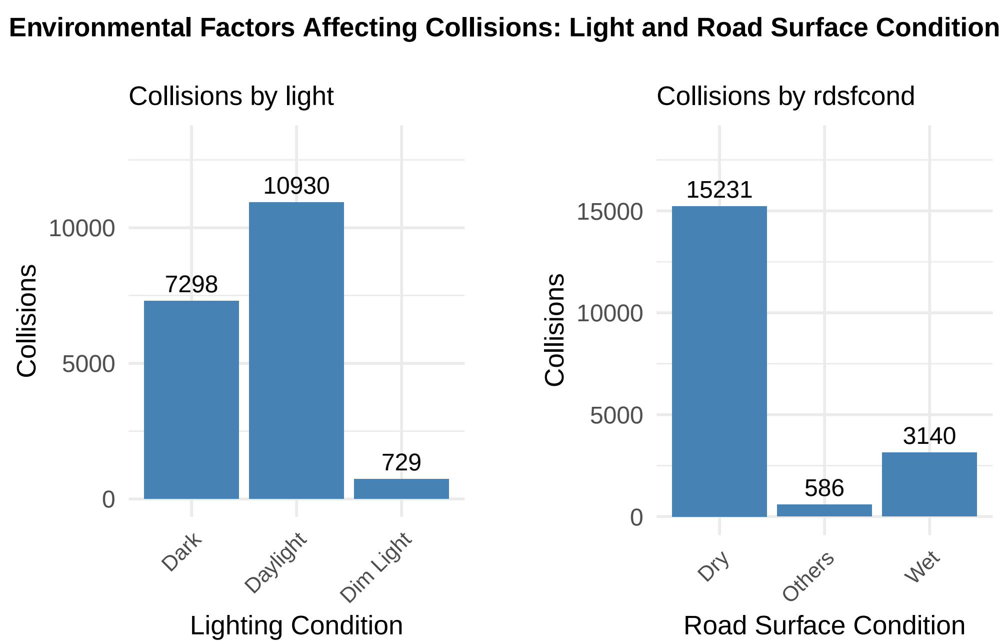

Figure 4: Environmental Factors Affecting Collisions - Light and Road Surface

3 Discussion

3.1 High-Risk Scenarios for Road Accidents

Figure 5 shows the relationship between environmental factors—traffic control, visibility, lighting, and road surface—and human behaviours like speeding, aggressive driving, alcohol use, and disability. It highlights clear patterns where certain combinations lead to higher accident rates.

A fatal collision frequently occurs when traffic signals are present, clear visibility, daylight, the road is dry, and aggressive driving is involved. Even in ideal environmental conditions, aggressive driving leads to the highest number of accidents. This suggests that behaviours like tailgating and reckless overtaking pose significant risks, even on dry roads with good visibility. To reduce accidents in such areas, stricter enforcement of speed limits and penalties for aggressive driving may be necessary.

Another high-risk combination is no traffic control, clear visibility, daylight, and dry roads with no human factors involved. The lack of traffic control in this scenario appears to give drivers a false sense of safety, leading to reckless behaviour. Installing traffic controls, like stop signs or roundabouts, in areas with high accident rates could help reduce these incidents.

This shows that while environmental factors like no traffic controls, wet roads and poor visibility are important, aggressive driving is a key behaviour that amplifies risk, even in favorable conditions. Focusing on improving infrastructure and stricter enforcement against aggressive driving and alcohol-related offenses could significantly lower accident rates in these high-risk scenarios.

3.2 Can Environmental and Human Factors Serve as Early Warning Indicators?

With advances in data-driven technologies, it's now possible to predict the severity of motor vehicle accidents by analyzing a combination of human behaviours and environmental conditions. This section focuses on how predictive models can be developed to assess the likelihood of severe accidents using factors such as speeding, alcohol use, poor road conditions, and low visibility.

For instance, a system could gather real-time environmental data, such as weather and road surface conditions, and combine it with information on traffic behaviours, like speeding patterns, to determine if an area is at high risk for serious accidents. This type of early warning system could be integrated into navigation apps or smart cars to alert drivers to potential dangers and help prevent accidents before they occur.

Predictive analytics can identify the most dangerous combinations of factors, such as speeding during rainy weather or aggressive driving in poor visibility. Machine learning algorithms could continuously improve these models by incorporating historical accident data, making

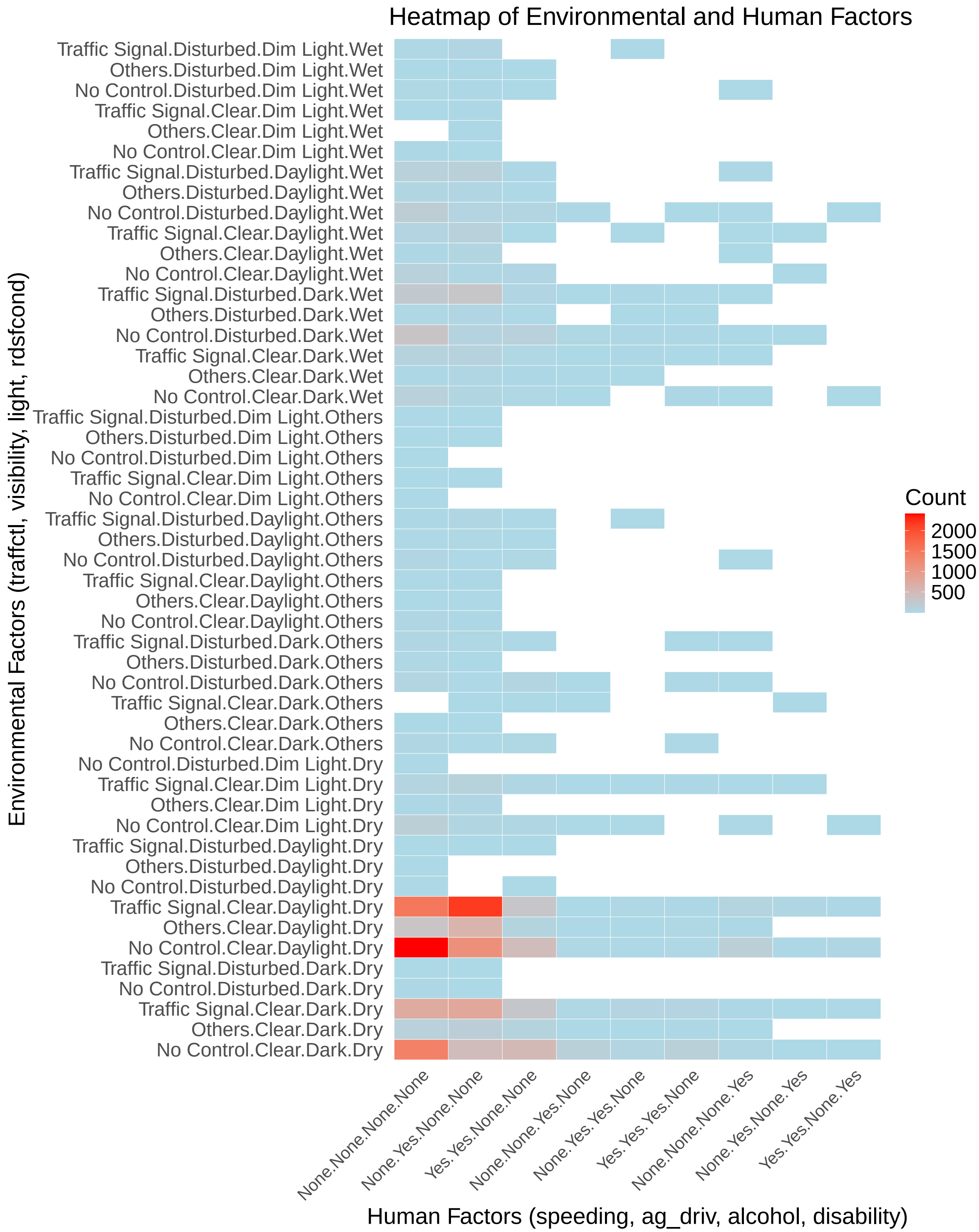

Figure 5: High-Risk Scenarios for Collisions

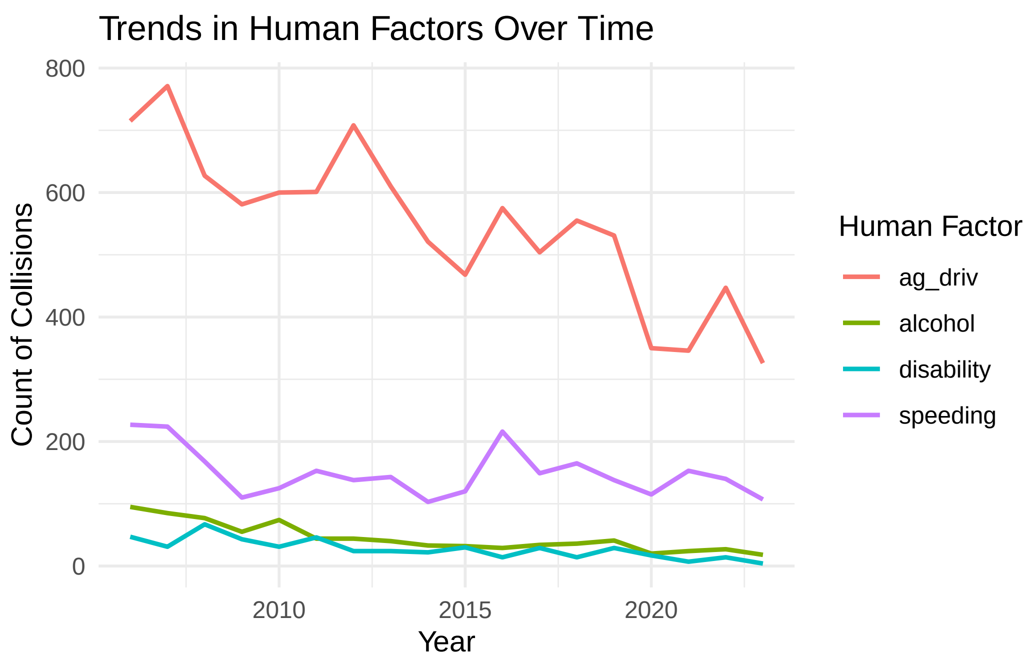

Figure 6: Trends of human factors in traffic collisions over time

the predictions more accurate over time. This approach shifts accident prevention from a reactive to a proactive strategy, which could reduce the strain on healthcare systems and improve overall road safety.

3.3 Weaknesses and next steps

One limitation of this analysis is that the dataset only includes observations up to 2023. While this offers a strong historical view on traffic accidents in Toronto, it does not account for more recent shifts in road usage, traffic enforcement, or driver behaviour that may have emerged after 2023. For example, new driving technologies, ride-sharing services, and updated traffic regulations could significantly influence accident rates, but these factors are not reflected in the current data. As a result, the findings may not fully align with recent trends, limiting their relevance for current and future policy decisions.

A key area for future research involves the growing use of autonomous or driver-assist technologies, which are not yet covered by the dataset but could have a major impact on traffic accident patterns. As these technologies become more common, understanding their effect on traffic safety and accident severity will be essential. The lack of data on autonomous vehicles limits our ability to predict how these innovations might change accident trends. However, from the figure Figure 6, we see that accidents involving aggressive driving have decreased, possibly due to the rise of driver-assist features that reduce reckless behaviour. Future studies should incorporate data on autonomous vehicles and their interactions with human-driven cars to better understand how these advancements affect safety. This could help develop predictive models that account for both human and automated driving behaviours, providing important understanding for policymakers and urban planners as they adapt infrastructure and safety regulations for a future where both types of vehicles share the road.

References

Auguié, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Müller, Kirill, and Jennifer Bryan. 2020. *Here: A Simpler Way to Find Your Files*. https://cran.r-project.org/web/packages/here/index.html.

R Core Team. 2023a. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

———. 2023b. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.