Factors Contributing to Pedestrian and Cyclist Collisions in Toronto*

Yuxuan Wei

September 27, 2024

This paper examines the key factors causing the pedestrian and cyclist collisions in Toronto from 2010 to 2023. Using data from the Toronto Police Service published on Open Data Toronto. Analyzing trends in collisions by the time of day, weather conditions, and neighborhood. The finding suggests that time of day, visibility, and road conditions significantly affect collision rates. This paper aims to inform urban planning and traffic safety policies to reduce accidents involving vulnerable road users.

1 Introduction

Ensuring the safety of bicyclists and pedestrians has grown in importance as cities like Toronto seek to promote active transportation (Pucher and Buehler (2017)). This paper investigates the key factors contributing to pedestrian and cyclist collisions in Toronto from 2010 to 2023, aiming to identify patterns and risk factors that can inform policy decisions and urban planning strategies. The remainder of this paper is structured as follows: Section Section 2 describes the dataset I used in the analysis. Section Section 3 presents findings, which are focusing on temporal patterns, environmental factors, and spatial distribution of collisions. Section Section 4 discusses the implications of these results, and Section Section B summarizes the key findings and their potential applications.

*Code and data are available at: https://github.com/wyx827/cyclist_collisions.git

2 Data

2.1 Measurement

This study utilizes data from the "Motor Vehicle Collisions involving Killed or Seriously Injured Persons" dataset, obtained from Open Data Toronto (**opendatatoronto2023?**). We accessed this data using the opendatatoronto package (Gelfand 2022). The dataset focuses specifically on incidents involving pedestrians and cyclists, providing detailed information on collision dates, times, weather conditions, lighting, road surface conditions, and injury severity. Table 1 shows part of columns in cleaned dataset.

Data cleaning and analysis were performed using R (R Core Team 2023), with several key packages: tidyverse (Wickham et al. 2019) for data manipulation and visualization, knitr (Xie 2023) for table generation, lubridate (Grolemund and Wickham 2011) for date and time handling, and here (Müller 2020) for managing file paths to ensure reproducibility. All visualizations were created using ggplot2 (Wickham 2016), which is part of the tidyverse package.

Table 1: Sample of the 2010-2023 cleaned collision data with a part of variables

Date	Hour	Weather	Visibility	Light	Injury	Type
2010-01-07	21	Clear	Clear	Dark	Major	Cyclist
2010-01-08	18	Clear	Clear	Dark	Fatal	Pedestrian
2010-01-09	10	Clear	Clear	Daylight	Major	Pedestrian
2010-01-09	11	Unknown	Clear	Daylight	Major	Pedestrian
2010-01-12	11	Clear	Clear	Daylight	Major	Pedestrian
2010-01-12	12	Clear	Clear	Daylight	Minor	Pedestrian
2010-01-12	12	Clear	Clear	Daylight	Fatal	Pedestrian
2010-01-12	13	Clear	Clear	Daylight	Minor	Pedestrian
2010-01-12	13	Clear	Clear	Daylight	Fatal	Pedestrian
2010-01-12	14	Clear	Clear	Daylight	Major	Pedestrian

3 Results

3.1 Temporal Patterns

The data shows clear trends in the frequency of collisions between cyclists and pedestrians at various times of the day and year. Figure 1 shows the number of collisions involving pedestrians (in blue) and cyclists (in green) at different hours of the day. The frequency of pedestrian collisions peaks during the afternoon and evening hours, particularly between 15:00 and 19:00, while cyclist collisions are more distributed throughout the day, with noticeable peaks around 10:00, 16:00, and 18:00, which means the timing of the collisions corresponds to high-traffic hours, especially the evening rush hours, when both pedestrians and cyclists are more active.

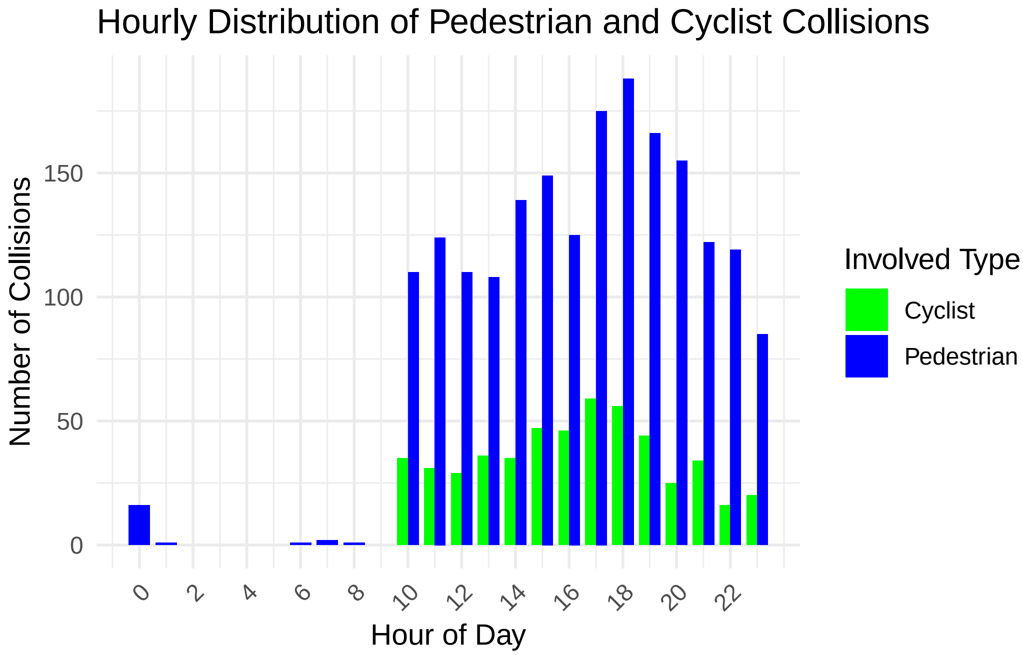

Figure 1: Hourly distribution of pedestrian and cyclist collisions

3.2 Environmental Factors

Weather conditions and visibility may not play crucial roles in the safety of pedestrians and cyclists (Reynolds et al. (2009)). Figure 2 shows the number of collisions under different weather conditions (Adverse, Clear, and Unknown) and visibility types (Clear, Rain, Snow, Other). The majority of collisions occur during clear weather conditions, with visibility marked as clear. A smaller number of collisions occur in rainy conditions, while snow-related collisions are rare. Despite the higher risk associated with adverse weather, the largest number of collisions happens during clear weather, suggesting that traffic volume may be a more significant factor than weather conditions. However, rain still shows a notable increase in collision frequency, indicating that wet conditions are a contributing factor.

##Spatial Distribution

The frequency of collisions varies significantly across different neighborhoods in Toronto. Figure 3 highlights the neighborhoods with the highest collision frequency in Toronto. The Yonge-Bay Corridor has the highest number of collisions, followed by Kensington-Chinatown and South Riverdale. The spatial distribution of collisions indicates that certain high-density areas or neighborhoods with heavy pedestrian and cyclist traffic see significantly more incidents.

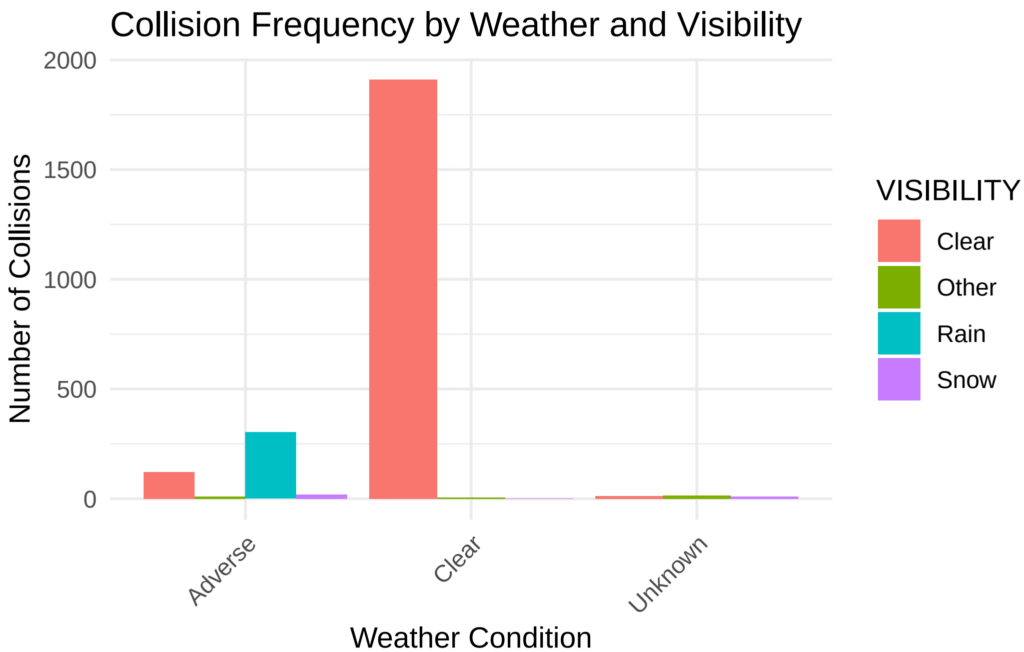

Figure 2: Collision frequency by weather and visibility conditions

These areas may require targeted safety interventions like improved crosswalks, signage, or dedicated cycling lanes.

```
Selecting by n
```

4 Discussion

Our findings highlight several key factors contributing to pedestrian and cyclist collisions in Toronto:

1. Time of Day: As seen in the hourly distribution of collisions, there is a clear spike during morning (10:00) and evening rush hours (16:00–19:00). This trend aligns with higher traffic volumes during these periods. The peak in both pedestrian and cyclist collisions suggests a need for enhanced safety measures, such as better traffic management, additional crossing guards, and the implementation of rush-hour slow zones.

2. Weather and Visibility: Contrary to the expectation that adverse weather would lead to more collisions, the majority of incidents occurred under clear weather and visibility conditions. This could be due to the increased number of pedestrians and cyclists when weather conditions are favorable. However, rain does show an elevated risk, implying that

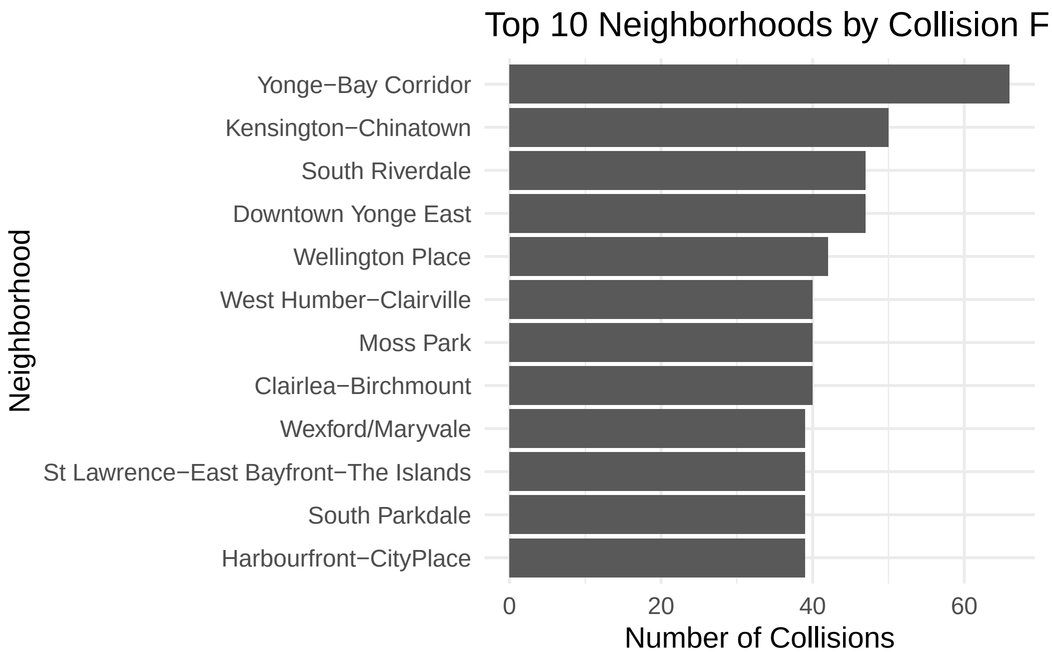

Figure 3: Top 10 neighborhoods by collision frequency

wet conditions negatively impact road safety. This suggests the need for infrastructure improvements like anti-slip surfaces and better drainage systems to mitigate the effects of rain on road safety.

3. Neighborhood Characteristics: Certain neighborhoods, particularly the Yonge-Bay Corridor and Kensington-Chinatown, have significantly higher collision rates. These areas may experience heavier traffic, higher pedestrian density, or insufficient infrastructure for cyclists and pedestrians. Addressing these issues with neighborhood-specific interventions—such as dedicated bike lanes, improved pedestrian crossings, or traffic calming measures—could help reduce collisions in these high-risk areas.

Based on these findings, there are a complex interplay of factors influencing collision risk for pedestrians and cyclists. They point to the need for targeted interventions, such as improved lighting in high-risk areas, weather-responsive traffic management, and neighborhood-specific safety initiatives.

4.1 Temporal Patterns and Urban Mobility

The analysis of temporal patterns in pedestrian and cyclist collisions reveals critical insights into urban mobility and safety challenges in Toronto. The peak in collisions during morning and evening rush hours, as shown in Figure 1, aligns with typical commuting patterns. This correlation suggests that the increased volume of both vehicular and pedestrian/cyclist traffic during these times creates a higher risk environment.

These findings echo broader research on urban transportation patterns, such as the work by Pucher and Buehler (2017), which emphasizes the need for integrated transportation planning that considers the safety of all road users. The concentration of collisions during peak hours points to potential inadequacies in current infrastructure to safely accommodate the surge in mixed-mode traffic.

To address this, cities like Copenhagen and Amsterdam have implemented time-separated traffic signals for cyclists and pedestrians, reducing conflict points with vehicles during busy periods. Toronto could consider similar strategies, along with expanding dedicated cycling infrastructure and pedestrian pathways to create safer corridors during high-traffic times.

4.2 Environmental Factors and Road Safety

The impact of weather conditions and visibility on collision rates, as illustrated in Figure 2, underscores the complex relationship between environmental factors and road safety. The data shows a higher frequency of collisions in clear weather conditions, which might seem counterintuitive at first glance. However, this pattern likely reflects higher overall road usage during favorable weather, particularly by cyclists and pedestrians.

This finding aligns with research by Reynolds et al. (2009), which highlights the multifaceted nature of environmental influences on cycling safety. While adverse weather conditions like rain or snow might intuitively seem more dangerous, they often result in reduced overall traffic and more cautious behavior from all road users.

The implications of these findings are twofold: 1. There's a need for year-round safety measures that don't solely focus on adverse weather conditions. 2. Public education campaigns should emphasize vigilance in all weather conditions, countering the potential complacency that might occur during seemingly "safer" clear weather.

4.3 Spatial Distribution and Urban Planning

The spatial analysis revealing collision hotspots in specific neighborhoods (Figure 3) provides crucial information for targeted urban planning interventions. This uneven distribution of collision risks across the city points to potential disparities in infrastructure quality, traffic management, or local road use patterns.

Urban planning theories, such as those proposed by Jane Jacobs, emphasize the importance of street-level design in promoting safety and vibrant urban environments. The concentration of collisions in certain neighborhoods suggests that a one-size-fits-all approach to road safety is insufficient. Instead, localized strategies that consider the unique characteristics of each neighborhood are necessary.

Potential interventions could include: 1. Neighborhood-specific traffic calming measures 2. Customized street redesigns that prioritize pedestrian and cyclist safety in high-risk areas 3. Community engagement initiatives to understand local mobility patterns and safety concerns

4.4 Weaknesses and next steps

While this study provides valuable insights, it has several limitations that should be addressed in future research:

Lack of Exposure Data: The analysis doesn't account for the varying levels of pedestrian and cyclist activity across different times and locations. Incorporating data on foot and bicycle traffic volumes would provide a more accurate picture of relative risk. Limited Scope of Factors: The study focuses primarily on temporal and environmental factors. Future research should consider additional variables such as road design, speed limits, and demographic factors of those involved in collisions. Potential Reporting Bias: The dataset only includes reported collisions, which may not capture all incidents, particularly minor ones. This could lead to an underestimation of the true collision rates. Absence of Longitudinal Analysis: While the data spans from 2013 to 2023, this study doesn't deeply explore year-over-year trends or the impact of specific interventions over time.

Next steps for research could include:

Conducting a comparative analysis with other cities to benchmark Toronto's performance and identify best practices. Integrating data on infrastructure changes (e.g., implementation of bike lanes) to assess their impact on collision rates. Employing machine learning techniques to predict high-risk areas and times, enabling proactive safety measures. Collaborating with public health researchers to link collision data with health outcomes, providing a more comprehensive understanding of the impact of these incidents.

By addressing these limitations and pursuing these next steps, future research can build upon this study to provide even more robust and actionable insights for improving pedestrian and cyclist safety in Toronto.

A Additional data details

A.1 Selected Columns for Analysis

There are more than 40 columns in the raw dataset at first

The following 11 columns were selected from the raw dataset for further analysis: DATE: The date the collision occurred. TIME: The time the collision occurred. ACCLASS: The classification of the accident, indicating whether it resulted in an injury or fatality. VISIBILITY: The environmental condition at the time of the collision (e.g., clear, foggy, etc.). LIGHT: The light condition (e.g., daylight, darkness) at the time of the collision. RDSFCOND: The road surface condition (e.g., dry, wet, icy). INJURY: The severity of the injury, categorized as minor, major, or fatal. INVTYPE: The involvement type, specifying whether the individual involved was a pedestrian or cyclist. INITDIR: The initial direction of travel for the vehicle involved in the collision. NEIGHBOURHOOD_158: The City of Toronto neighborhood where the collision occurred, identified by its name. DIVISION: The Toronto Police Service Division responsible for the area where the collision occurred.

A.2 Column Value Transformations

The selected columns were further cleaned and transformed to facilitate analysis, as follows:

COLLISION_DATE: Created by converting the DATE field into a standard date format (YYYY/MM/DD). COLLISION_YEAR: Extracted from COLLISION_DATE, representing the year of the collision. COLLISION_TIME: Created by combining DATE and TIME and parsing them into a datetime format. HOUR: Extracted from COLLISION_TIME, representing the hour of the collision. WEATHER: Derived from RDSFCOND to simplify weather conditions: - "Clear" for dry conditions. - "Adverse" for wet, snowy, or icy conditions. - "Unknown" for other or missing conditions. VISIBILITY: Simplified to include common visibility conditions (Clear, Rain, Snow, Fog) and an "Unknown" category for missing values. LIGHT: Cleaned and categorized into standard lighting conditions (Daylight, Dark, Dawn, Dusk) and an "Unknown" category for missing or unclear data. INJURY: Missing values for injury severity were replaced with "None." NEIGHBOURHOOD: Transformed from NEIGHBOURHOOD_158 to represent the name of the neighborhood where the collision occurred.

A.3 Data Cleaning

The raw data from the Toronto Police Service required several preprocessing steps to prepare it for analysis:

1. Filtering for pedestrian and cyclist incidents only.
2. Converting date and time strings to appropriate datetime objects.

3. Categorizing weather conditions into broader groups for easier analysis.
4. Handling missing values, particularly in the visibility and light condition fields.

These steps ensured that our analysis focused specifically on pedestrian and cyclist collisions and that the data was in a format conducive to the types of analyses we performed.

A.4 Diagnostics

?@fig-missing-data is a visualization of missing data in our dataset. It shows that most variables have complete data, with some missingness in the visibility and light condition fields. This missingness was addressed through imputation based on related variables like weather conditions.

?@fig-collision-trends shows the trend of collisions over the years covered by our dataset. This plot helps us understand if there are any long-term trends or anomalies in the data that need to be accounted for in our analysis.

This plot helps to provide additional context for interpreting my main results. Also highlights areas where further investigation or data collection might be necessary in future studies.

B Conclusion

This analysis help people understand what causes pedestrian and cyclist collisions in Toronto. By finding patterns and risk factors, we can create better policies and urban designs to improve road safety for these vulnerable groups.

Future studies could look at how well specific solutions, like protected bike lanes or safer intersections, reduce collisions. Including data on traffic volume and changes to infrastructure could also give us a clearer picture of what leads to accidents.

As Toronto encourages more walking and cycling, making sure these activities are safe is crucial. This research shows where focused actions can make the biggest difference in reducing accidents and improving the city's transportation system for everyone.

C References

City of Toronto. (2023). Motor Vehicle Collisions involving Killed or Seriously Injured Persons. https://open.toronto.ca/dataset/motor-vehicle-collisions-involving-killed-or-seriously-injured-persons/

Pucher, John, and Ralph Buehler. 2017. "Cycling towards a more sustainable transport future." Transport Reviews 37 (6): 689-694. https://www.tandfonline.com/doi/full/10.1080/01441647.2017.134023

Reynolds, Conor C. O., M. Anne Harris, Kay Teschke, Peter A. Cripton, and Meghan Winters. 2009. "The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature." Environmental Health 8 (1): 47. https://ehjournal.biomedcentral.com/articles/10.1186/1476-069X-8-47

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.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://joss.theoj.org/papers/10.21105/joss.01686

Xie, Yihui. 2023. "knitr: A General-Purpose Package for Dynamic Report Generation in R." https://yihui.org/knitr/ Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with {lubridate}." Journal of Statistical Software 40 (3): 1-25. https://www.jstatsoft.org/v40/i03/

Müller, Kirill. 2020. "here: A Simpler Way to Find Your Files." https://CRAN.R-project.org/package=here Wickham, Hadley. 2016. "ggplot2: Elegant Graphics for Data Analysis." Springer-Verlag New York. https://ggplot2.tidyverse.org

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

Jacobs, Jane. 1961. "The Death and Life of Great American Cities." New York: Random House. https://www.petkovstudio.com/bg/wp-content/uploads/2017/03/The-Death-and-Life-of-Great-American-Cities_Jane-Jacobs-Complete-book.pdf

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

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pucher, John, and Ralph Buehler. 2017. "Cycling Towards a More Sustainable Transport Future." *Transport Reviews* 37 (6): 689–94. https://www.tandfonline.com/doi/full/10.1080/01441647.2017.1340234.

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

Reynolds, Conor C. O., M. Anne Harris, Kay Teschke, Peter A. Cripton, and Meghan Winters. 2009. "The Impact of Transportation Infrastructure on Bicycling Injuries and Crashes: A Review of the Literature." *Environmental Health* 8 (1): 47. https://ehjournal.biomedcentral.com/articles/10.1186/1476-069X-8-47.

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://joss.theoj.org/papers/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.