Age and Gender Dynamics in Traffic Accidents: A 2019 Canadian Case Study*

Sinan Ma

April 16, 2024

This study examines the dynamics of traffic collisions across Canada for the year 2019, employing an extensive dataset retrieved from Open Canada's national collision database. Through robust data cleaning and analysis, we identified critical factors contributing to the severity of traffic accidents, encapsulating demographic variables, environmental conditions, and temporal patterns. The methodology centered on a comprehensive examination of collision records, integrating Bayesian statistical models to interpret the nuanced interplay between predictors such as age, gender, weather, and road surface conditions. The results illuminate significant correlations, with adverse weather and certain road conditions markedly increasing the likelihood of fatal outcomes. These findings offer valuable insights into preventive measures and policy-making, aiming to enhance road safety and reduce the incidence of traffic-related fatalities. The implications of this research underscore the necessity for targeted interventions and bolster the framework for data-driven decision-making in national traffic management and infrastructure planning.

Introduction

Road traffic injuries represent a global public health crisis, claiming approximately 1.19 million lives annually and ranking as the leading cause of death among children and young adults aged 5–29 years (World Health Organization 2023). These staggering figures from the World Health Organization underscore the urgent need for concerted efforts in traffic collision research and policy intervention. In Canada, the intricacies of traffic incidents captured within the 2019 national collision dataset open a window into understanding these tragedies at a granular level. This research harnesses the depth of this data to unravel the demographic determinants—age, gender, and seating position—that predispose individuals to varying collision severity outcomes.

---

*Code and data are available at: https://github.com/Sinanma/Traffic_Collision_Analysis.git.

This study delves into statistical methodologies to decode the patterns embedded within the incidents, mapping out the demographic landscape of traffic collision severity. Utilizing a sophisticated Bayesian regression model, implemented through R—a comprehensive environment for statistical computing and graphics (R Core Team 2023), we dissect large-scale data to articulate the distinct demographic factors that significantly modulate the likelihood of severe traffic collisions. The methodology and analytical strategies employed are learned from *Telling Stories with Data* by Rohan Alexander (Alexander 2023), which provides the statistical framework and R code that underpin our analysis.

The estimand of this paper is the nuanced effect of these demographic variables—age, gender, and seating position—on the severity of traffic incidents, with a focus on the contextual interplay of environmental conditions such as weather and road surfaces.

Our investigation is vital for an array of stakeholders, from traffic safety regulators to urban planners and the public at large. Traditional traffic studies have often overlooked the multifaceted relationships inherent within demographic data and collision severity. This study fills this lacuna by employing a robust Bayesian approach, thereby enabling precise inferences on the complex interdependencies at play. The insights gleaned offer a potent tool for devising tailored preventive measures and enhancing road safety strategies, with the potential to mitigate the human and economic toll of traffic collisions.

Our findings present a nuanced analysis of collision severity disparities across various demographics. This study reveals the differential patterns of collision impacts among age groups, genders, and occupants based on their seating positions within vehicles. We discover the heightened risk factors associated with specific demographics, offering a critical assessment of the stereotypes surrounding vehicular accidents and safety measures.

The significance of this study transcends academic boundaries, providing actionable insights for strategic safety enhancements. It sheds light on the demographic dynamics that influenced Canadian traffic collisions in 2019, aiding the development of informed and effective interventions to reduce road traffic injuries and fatalities.

The paper begins by introducing the broader context and motivation behind our study. Section introduces the dataset utilized for analysis, explores the dataset and key variables, enhanced by visual data representations. Section describes the Bayesian regression model used to analyze these factors, elucidating its relevance and utility. **?@sec-results** presents the findings, offering detailed interpretations of the data. Section discusses the implications, study limitations, and potential future research directions.

Data

This research utilizes the 2019 National Collision Database, publicly available via Open Canada's data portal. The dataset encompasses detailed records of traffic collisions throughout Canada, capturing various environmental and demographic factors associated with each

incident. The comprehensive dataset for 2019 can be accessed directly through National Collision Database (Transport Canada 2019b), and additional details about the dataset are available on its main overview page here (Transport Canada 2019a).

Dataset Description

The 2019 dataset provides extensive data points for over 200,000 traffic incidents, detailing the date, time, geographic coordinates, environmental conditions, vehicle types, and personal demographics involved in each collision. This rich dataset offers a unique opportunity to explore the nuanced factors that contribute to traffic collision severities in Canada, such as the specific impact of road surface conditions during different weather scenarios.

Variables of Interest

The primary variables analyzed in this study include: - **Month**: Categorical; the month when the collision occurred, from "January" to "December". - **Day of the Week**: Categorical; the day of the week on which the collision occurred, from "Monday" to "Sunday". - **Weather Condition**: Categorical; the observed weather conditions during the incident, such as "Clear", "Rain", "Snow", etc. - **Road Surface Condition**: Categorical; the condition of the road during the incident, including "Dry", "Wet", "Icy", etc. - **Collision Severity**: Categorical; the severity of the collision, categorized as "Fatal" or "Non-Fatal".

Each variable is crucial for understanding the dynamics of traffic accidents and their outcomes, which can significantly aid in preventative measures and policy making.

Data Cleaning and Preparation

Data was cleaned and prepared for analysis, which involved handling missing values, correcting data entry errors, and formatting date and time entries for consistency. The cleaned dataset was then used for all subsequent analyses, ensuring reliability and accuracy in the findings.

Graphical and Statistical Analysis

To elucidate the data's characteristics and the relationships between various factors and traffic collision outcomes, extensive graphical and statistical analyses were performed. Each figure provides insights into specific aspects of the data, helping to visualize the complex interactions and distributions that underpin traffic collision dynamics.

- **Figure 1**: Distribution of Accidents by Driver Age reveals the age profile of drivers involved in accidents, highlighting specific age groups that are more susceptible to being in collisions.

- **Figure 2**: [Distribution of Accidents by Driver Gender] (Figure 2) explores the gender distribution among drivers involved in traffic accidents, indicating potential gender-specific patterns in accident rates.

- **Figure 3**: Impact of Weather Conditions on Collision Severity: Fatal vs. Non-Fatal presents a side-by-side comparison of how different weather conditions affect the severity of traffic collisions, contrasting fatal and non-fatal outcomes.

- **Figure 4**: Impact of Road Surface Conditions on Collision Severity: Fatal vs. Non-Fatal illustrates the relationship between road surface conditions and the severity of collisions, providing a parallel analysis for both fatal and non-fatal incidents.

- **Figure 5**: Distribution of Traffic Accidents by Day of the Week depicts the frequency of accidents across different days, identifying any trends related to the day of the week on accident rates.

- **Figure 6**: Distribution of Traffic Accidents by Month illustrates the monthly distribution of traffic accidents, helping to uncover seasonal or monthly variations in collision frequencies.

These visualizations are pivotal for understanding the underlying patterns in the data and assist in making informed conclusions about traffic safety and accident prevention strategies.
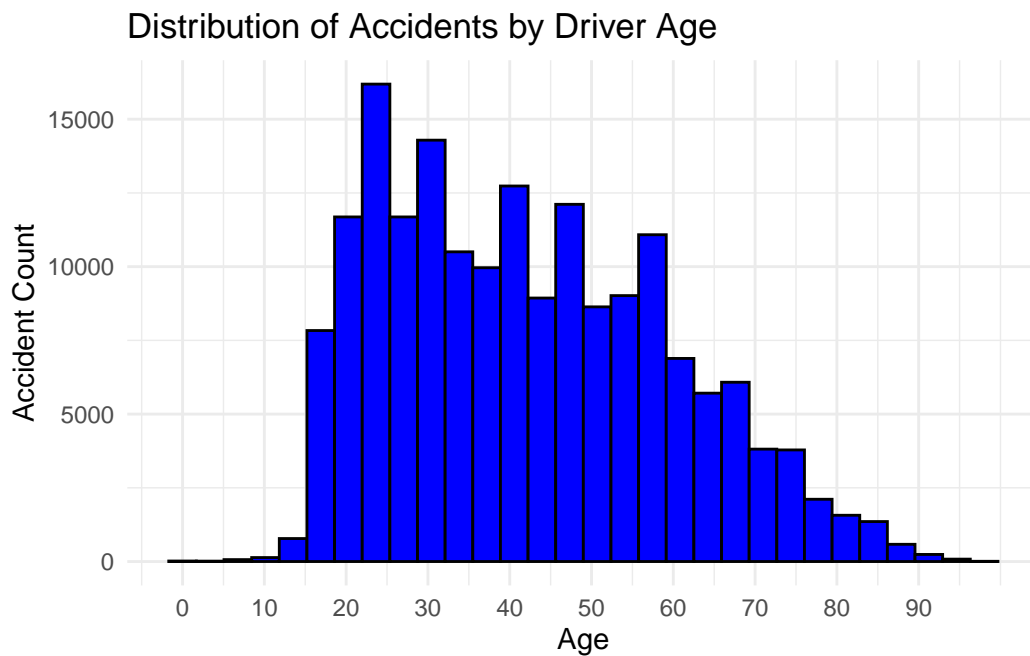

Figure 1: Distribution of Accidents by Driver Age
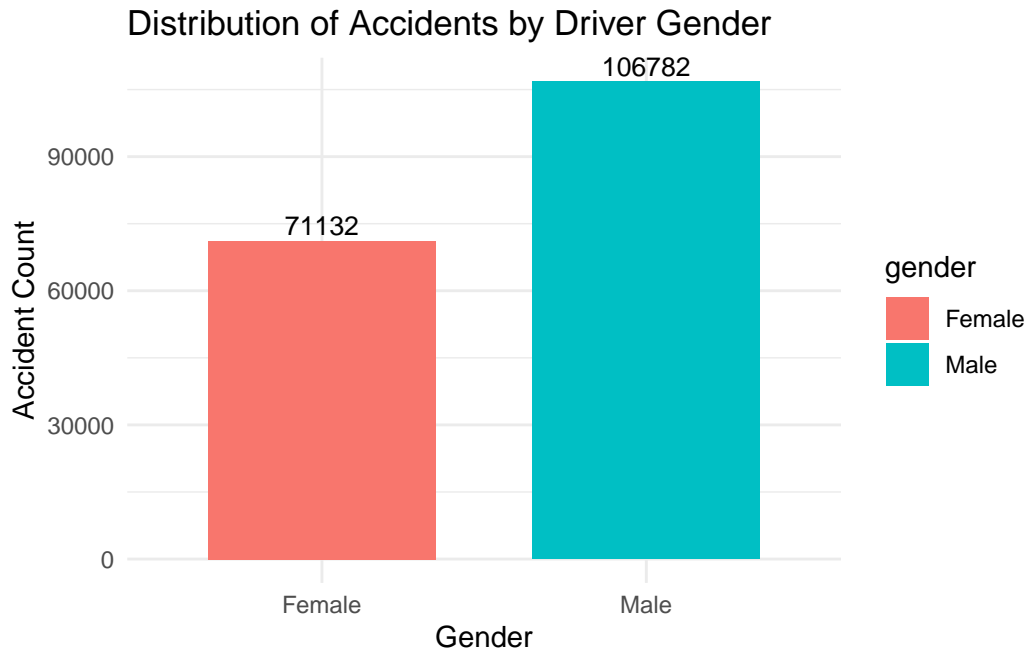
Talk more about it.

Figure 2: Distribution of Accidents by Driver Gender

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix .

Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.
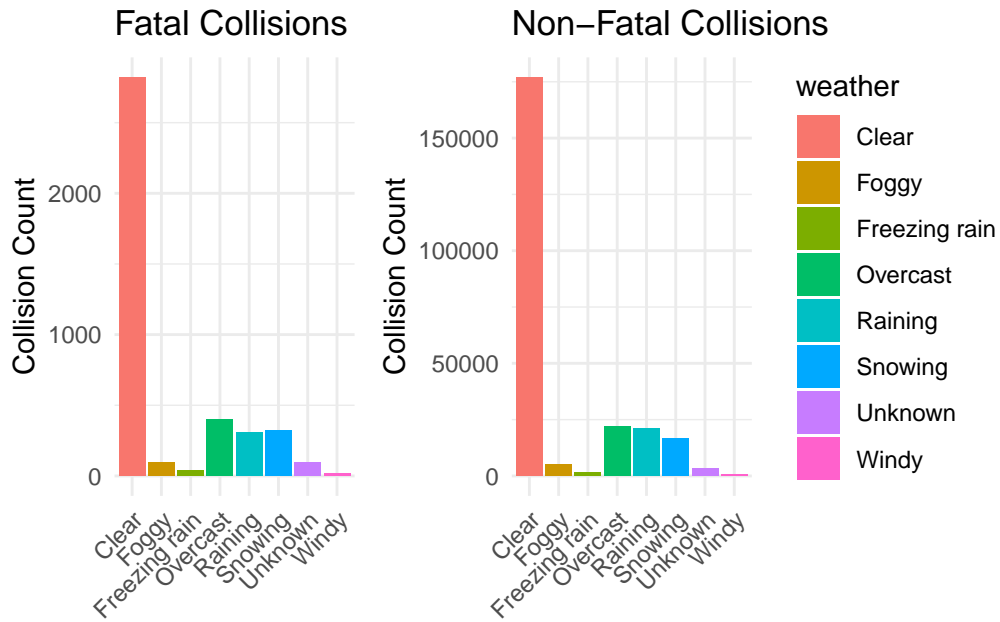
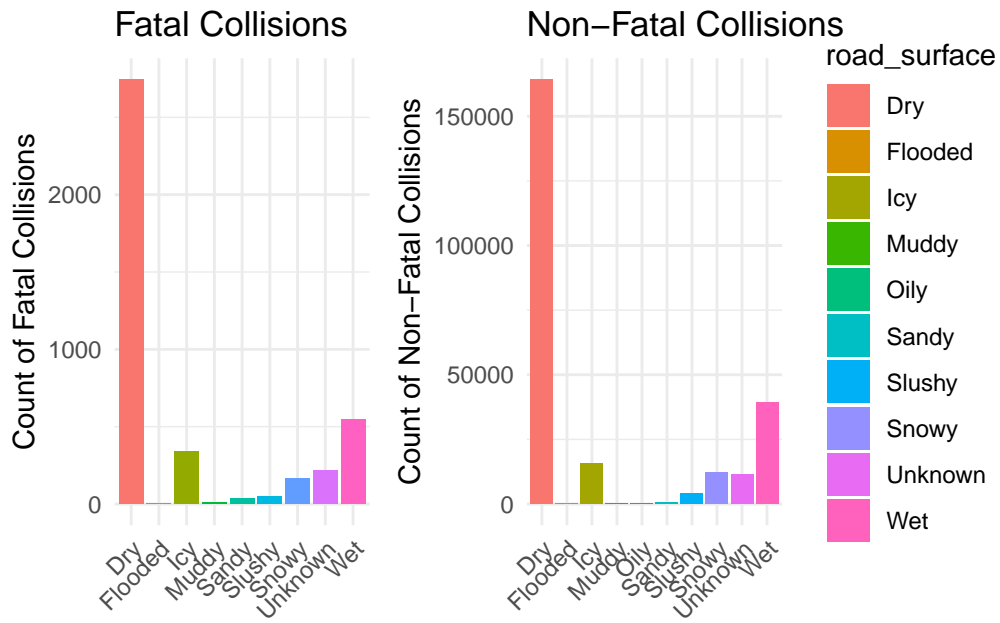Figure 3: Impact of Weather Conditions on Collision Severity: Fatal vs. Non-Fatal

Figure 4: Impact of Road Surface Conditions on Collision Severity: Fatal vs. Non-Fatal
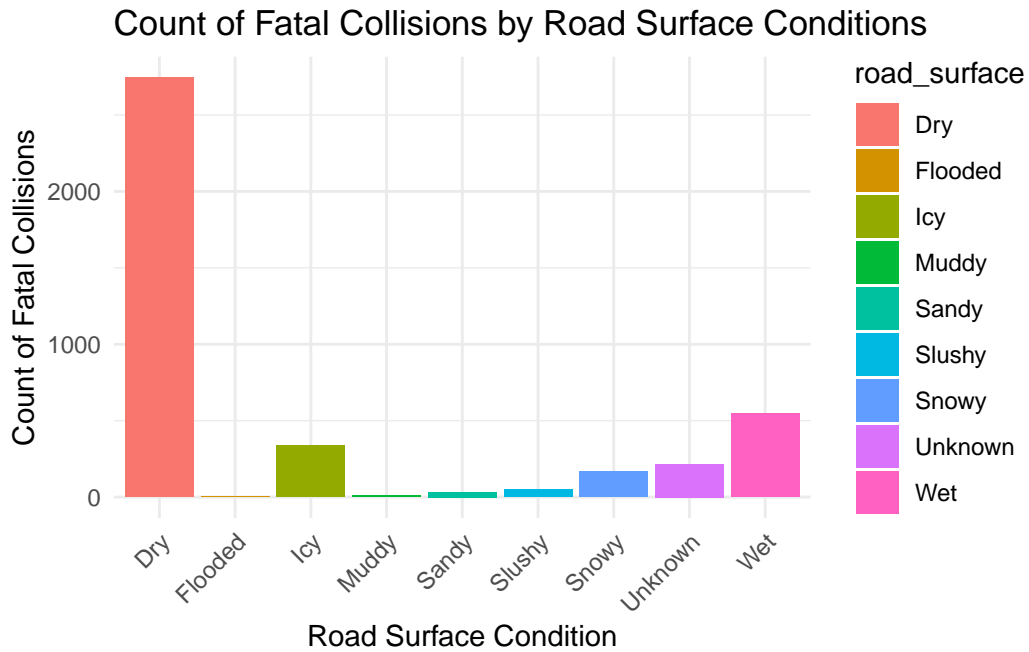
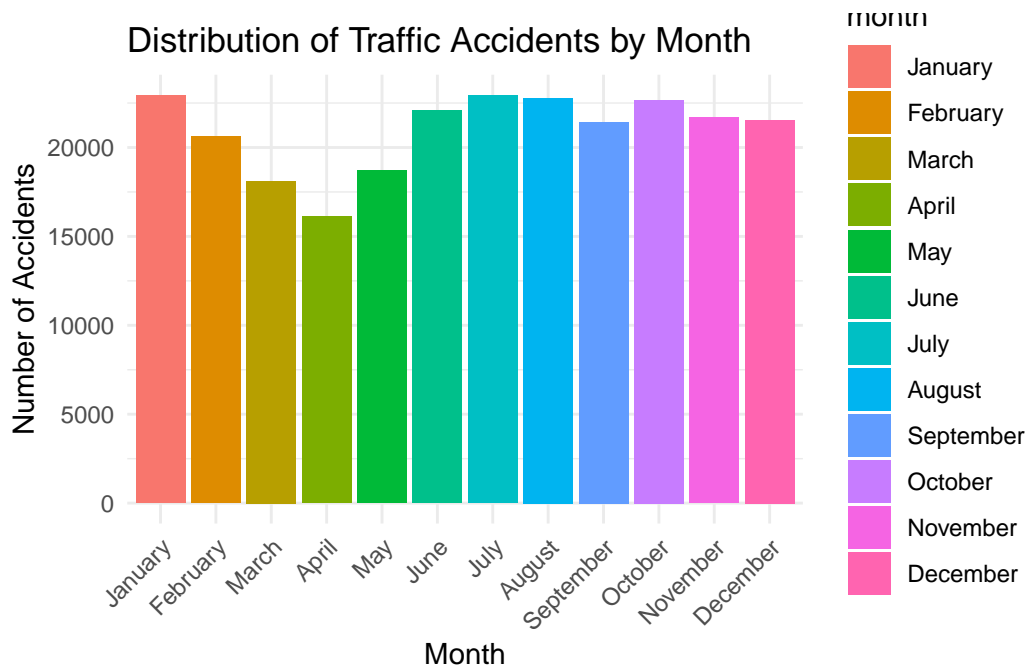Figure 5: Impact of Road Surface Conditions on Collision Severity

Figure 6: Distribution of Traffic Accidents by Month

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

Results

Our results are summarized in Table 1.

Discussion

First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Second discussion point

Third discussion point

Weaknesses and next steps

Weaknesses and next steps should also be included.

Table 1: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

Appendix

Additional data details

Model details

Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows... This suggests...

?@fig-stanareyouokay-2 is a Rhat plot. It shows... This suggests...

References

Alexander, Rohan. 2023. *Telling Stories with Data.* Boca Raton: CRC Press. https://tellingstorieswithdata.com/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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

Transport Canada. 2019a. "2019 National Collision Database (NCDB)." Open Canada. https://open.canada.ca/data/en/dataset/1eb9eba7-71d1-4b30-9fb1-30cbdab7e63a.

———. 2019b. "National Collision Database (NCDB) – 2019 Collision Data." Online. https://open.canada.ca/data/dataset/1eb9eba7-71d1-4b30-9fb1-30cbdab7e63a/resource/1e2a6aa0-c35f-4736-984a-71f2eac1c1dc/download/2019_dataset_en.csv.

World Health Organization. 2023. "Road Traffic Injuries." https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.