Analyzing Toronto Traffic Speeds: Insights into Urban Mobility, Safety, and Speed Dynamics*

Exploring Key Traffic Patterns and Relationships Between Speed Percentiles and Volume

Andy Jiang

December 3, 2024

This paper investigates traffic speed data from Toronto collected between 2017 and 2024, focusing on trends in speed percentiles and their relationship to traffic volume. Using a Bayesian linear regression model, we predict mean vehicle speeds based on key predictors such as 5th, 50th, and 95th percentile speeds and total traffic volume. Results highlight significant changes in traffic patterns post-2020, with increased extreme speeds linked to higher traffic volumes. This study provides insights into urban mobility dynamics, offering a basis for traffic management strategies and policy-making.

Table of contents

---

*Code and data are available at: https://github.com/AndyYanxunJiang/toronto-vehicles-speed-analysis.

1 Introduction

Traffic speed analysis is a critical tool for understanding urban mobility, road safety, and congestion. In metropolitan areas like Toronto, shifting traffic patterns reflect changes in urban infrastructure, population growth, and behavioral adaptations to external events, such as the COVID-19 pandemic. The advent of technologies like radar-equipped speed signs has enabled the collection of detailed traffic data, providing new opportunities to analyze and predict traffic behavior at a granular level.

This paper focuses on traffic speed data collected in Toronto's designated Safety Zones between 2017 and 2024. The dataset comprises monthly summaries of speed percentiles, vehicle counts across specific speed ranges, and total traffic volumes. Notably, the study explores trends in extreme speeds (vehicles exceeding 100 km/h) and evaluates how they have evolved before and after 2020, a period marked by rapid population growth and post-pandemic recovery. These analyses aim to address a key gap: understanding how percentile-based speed measures and traffic volume influence overall mean speeds.

To investigate these relationships, we employ a Bayesian linear regression model to predict mean vehicle speed (`spd_mean`) based on predictors including 5th, 50th, and 95th percentile speeds and traffic volume. Bayesian modeling allows for the incorporation of uncertainty and prior information, making it well-suited for analyzing urban traffic data.

Our findings reveal significant changes in traffic behavior post-2020, including higher extreme speed occurrences at increased traffic volumes, reflecting broader shifts in urban mobility. This study highlights the importance of percentile-based metrics in predicting mean speeds and contributes to ongoing efforts to enhance traffic safety and efficiency. The paper is structured as follows: the data section provides an overview of the dataset and key variables, the model section outlines the Bayesian regression framework, results are discussed in detail, and the discussion section reflects on the implications and limitations of the findings.

The remainder of this paper is structured as follows. Section 2 discusses the raw data, cleaning process, and key variables of interest, offering visual and tabular representations of the dataset. Section 3 introduces and justifies the Bayesian linear regression model used to analyze the relationship between traffic volume, speed percentiles, and mean vehicle speed. Section 4 presents the findings of the model, highlighting significant trends and correlations within the data. Finally, Section 5 explores the broader implications of these results, offering insights into traffic behavior and potential avenues for future research. Additional information regarding data observation methodology is provided in Appendix A, and further model diagnostics are detailed in Appendix B.

2 Data

The data used in this paper came from the Open Data Toronto portal through the library open-datatoronto (Gelfand 2022), specifically the Safety Zone Watch Your Speed Program – Monthly Summary (Dumas 2024). All data analysis was conducted using R (R Core Team 2023) with the support of the following packages: tidyverse (Wickham et al. 2019), here (Müller 2020), ggplot2 (Wickham 2016), rstanarm (Goodrich et al. 2022), kableExtra (Zhu 2024), modelsummary (Arel-Bundock 2022), dplyr (Wickham et al. 2023), and lubridate (Grolemund and Wickham 2011). These packages provided a robust framework for data manipulation, visualization, Bayesian modeling, and reporting, facilitating reproducible and efficient analysis.

2.1 Overview

This paper analyzes traffic speed data collected in Toronto over the period 2017–2024. The dataset contains detailed information on percentile speeds, total traffic volume, and counts of vehicles traveling within specific speed ranges. These data provide an opportunity to study the evolution of traffic patterns in Toronto, especially in the context of changing demographics, urbanization, and post-pandemic recovery.

The dataset aggregates metrics monthly, offering insights into how traffic behavior has changed over time. This analysis is particularly focused on extreme speeds (vehicles exceeding 100 km/h) and their relationship with total traffic volume, as well as trends in key speed percentiles.

2.2 Measurement

The dataset analyzed in this paper originates from the Safety Zone Watch Your Speed Program (WYSP), a City of Toronto initiative designed to monitor and influence driver behavior in designated Safety Zones. Data are collected using radar-equipped speed display signs installed on hydro poles or streetlights, which measure the speeds of oncoming vehicles with an accuracy of ± 1 km/h and display the speeds on LED screens to encourage compliance with speed limits. These measurements are aggregated into monthly summaries, including key metrics such as percentile speeds (e.g., 5th, 50th, 95th percentiles), vehicle counts within specific speed ranges, and total observed traffic volume. The dataset is curated and published by Open Data Toronto, which ensures its metadata quality, completeness, and accessibility.

However, the dataset has certain limitations. The number of vehicles recorded by the speed signs is not equivalent to true traffic volume, as it may exclude vehicles that pass too quickly or fall outside the radar's range. Additionally, the presence of the signs might influence driver behavior, leading to changes in speed as drivers approach them. Furthermore, the data is limited to Safety Zones where the speed signs have been installed, restricting its spatial coverage.

Despite these limitations, the dataset is well-documented, complete, and up-to-date, providing valuable insights into traffic speed patterns across Toronto.

2.3 Speed Dataset Preview

The traffic speed dataset contains information collected from multiple locations and months, detailing percentile speeds, total traffic volume, and counts of vehicles traveling within specific speed ranges. The dataset aggregates these metrics monthly, offering insights into traffic behavior and variations over time.

This dataset includes:

- Percentile speeds (5th, 10th, … 95th) representing speed thresholds exceeded by corresponding proportions of vehicles.
- Traffic volume as the total number of vehicles observed monthly.
- Speed bins (e.g., 0-4 km/h, 5-9 km/h) for detailed breakdowns of vehicle counts across speed ranges.
- Table 1 showcases a simplified preview of the dataset with selected columns to highlight key features.

Table 1: Preview of analysis traffic speed dataset.

Month	5th Percentile	95th Percentile	Vehicles (>100 km/h)	Total Volume
2024-09-01	7	32	0	57758
2024-09-01	5	29	0	48365
2024-09-01	0	28	0	2116
2024-09-01	21	55	4	163591
2024-09-01	0	46	0	30606
2024-09-01	17	49	1	64690
2024-09-01	0	40	0	27790
2024-09-01	0	26	0	59467
2024-09-01	0	65	7	62706
2024-09-01	6	44	0	34309

2.4 Random Sample from the Dataset

Table 2 provides a random sample of 10 observations, offering an unbiased snapshot of key metrics such as speed percentiles, total traffic volume, and counts of vehicles exceeding 100 km/h. This selection illustrates the dataset's diversity without emphasizing specific patterns.

Table 2: Random sample of traffic speed dataset.

Month	5th Percentile	95th Percentile	Vehicles (>100 km/h)	Total Volume
2024-03-01	15	51	0	33333
2020-02-01	11	42	0	40020
2024-09-01	9	38	0	117573
2022-12-01	0	34	4	31121
2021-07-01	14	51	1	160808
2023-09-01	30	65	23	182810
2022-06-01	8	37	0	58684
2022-02-01	15	47	0	195854
2020-11-01	14	47	0	87896
2024-02-01	0	28	0	3472

2.5 Extreme Speeds Dataset

Vehicles traveling at extreme speeds (over 100 km/h) represent a critical factor in understanding traffic safety and violations. Table 3 highlights the top 5 months with the highest counts of vehicles exceeding 100 km/h. This provides insights into when extreme speeds are most prevalent.

Table 3: Top 5 months with the highest counts of vehicles exceeding 100 km/h.

Month	Total Volume	Vehicles (>100 km/h)	95th Percentile Speed
2021-11-01	45735	3423	122
2023-01-01	62476	1640	85
2023-01-01	87523	1475	77
2020-06-01	273454	1391	87
2020-07-01	224624	1304	52

2.6 Proportion of Vehicles at Moderate Speeds

Another key insight from this dataset is the proportion of vehicles traveling within a moderate speed range (50–70 km/h). This measure helps understand how typical driving speeds align with safe and efficient traffic flow. Table 4 highlights monthly proportions of vehicles in this speed range.

Table 4: Monthly proportions of vehicles traveling at moderate speeds (50–70 km/h).

Month	Moderate Speed Vehicles	Total Volume	Proportion (%)
Jan	32248526	208619163	15.46
Feb	31376308	213960232	14.66
Mar	35727340	235271611	15.19
Apr	31383270	194534832	16.13
May	32859985	207592245	15.83
Jun	32981330	201948797	16.33
Jul	33826108	196693112	17.20
Aug	33837333	195652825	17.29
Sep	36352915	240764265	15.10
Oct	32154660	203667881	15.79
Nov	33186997	217011827	15.29
Dec	32651478	211458699	15.44

2.7 Extreme Speeds and Traffic Volume

Figure 1 illustrates the relationship between extreme speeds (vehicles traveling over 100 km/h) and total traffic volume, categorized by pre-2020 and post-2020 periods. The post-2020 data points show a significant increase in both traffic volume and extreme speed counts, with some months recording over 2,000 vehicles exceeding 100 km/h. This increase may reflect changes in traffic behavior following the COVID-19 pandemic and a rapid population influx into Toronto through immigration. In contrast, pre-2020 data points display lower traffic volumes and fewer extreme speed counts, rarely exceeding 2,000. The fitted lines for each period highlight this divergence, with a steeper slope for post-2020 data, suggesting that higher traffic volumes post-pandemic are associated with a broader range of extreme speed occurrences.

2.8 Trends in Key Percentiles

Figure 2 examines trends in key speed percentiles (5th, 50th, and 95th percentiles) from 2017 to 2024, along with the average speed spread (the difference between the 95th and 5th percentiles). The 95th percentile speed demonstrates a clear decline, falling below 50 km/h over the years. Similarly, the 50th percentile speed dropped significantly between 2017 and 2018, stabilizing at around 30 km/h with minor fluctuations. The 5th percentile speed remained steady until mid-2022, when it experienced a noticeable decline to 5 km/h, down from its previous 10 km/h range. Despite these shifts, the speed spread remains relatively consistent, ranging between 35 and 40 km/h, indicating that the overall variability in speeds has not significantly changed. These trends may reflect broader changes in traffic regulations, driver behavior, or urban infrastructure developments in Toronto.
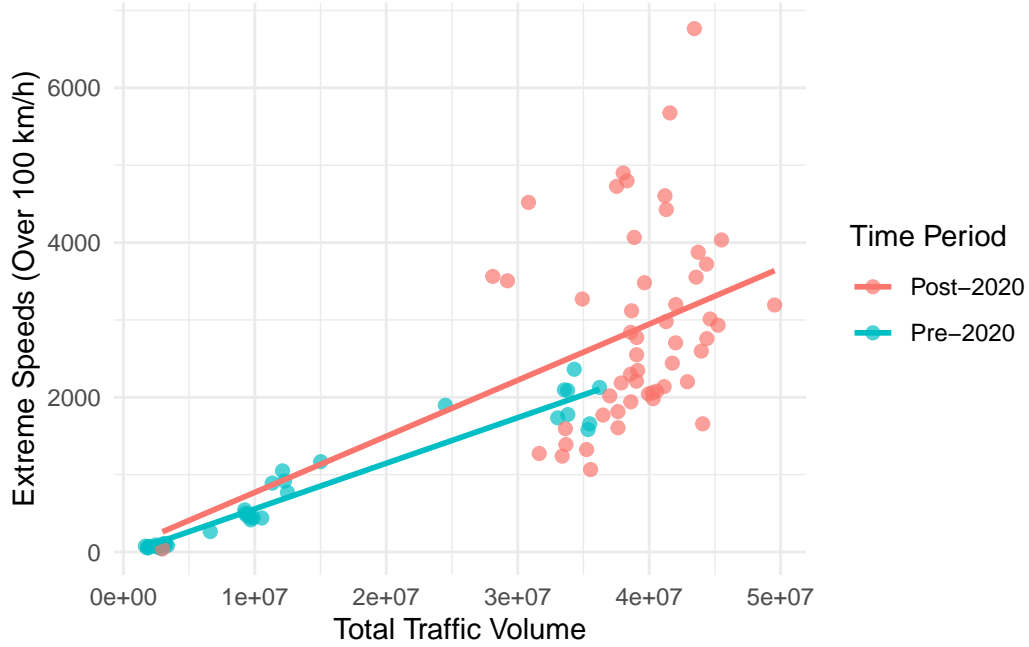
Figure 1: Relationship between extreme speeds (over 100 km/h) and total traffic volume, categorized by pre-2020 and post-2020 time periods.

3 Model

The purpose of this paper is to analyze the relationship between traffic volume, speed percentiles, and the mean speed of vehicles observed in Toronto. By utilizing a Bayesian linear regression model, we aim to investigate how various speed percentiles and total traffic volume predict the average vehicle speed (spd_mean). This model enables us to quantify the contributions of different speed measures and traffic density to overall traffic dynamics.

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

3.1 Model set-up

The Bayesian multiple linear regression model used in this paper predicts the mean speed (spd_mean) as a function of total traffic volume (volume) and speed percentiles (pct_05, pct_50, and pct_95). The model is defined as follows:
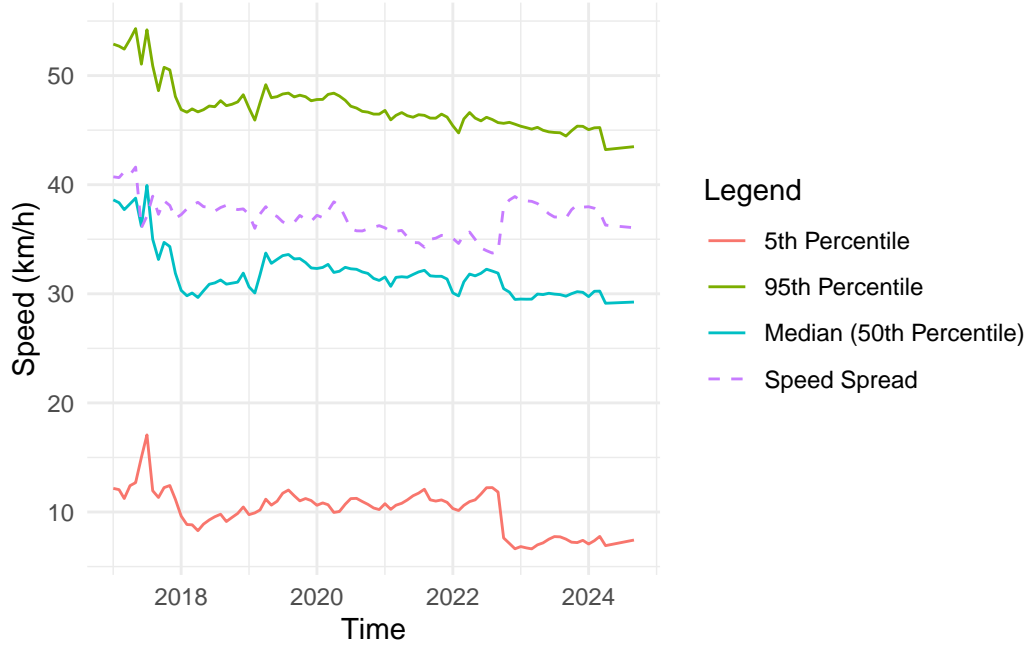
Figure 2: Trends in key speed percentiles (5th, 50th, and 95th) and the spread between the 95th and 5th percentiles over time.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \beta_0 + \beta_1 \cdot \text{volume}_i + \beta_2 \cdot \text{pct\_05}_i + \beta_3 \cdot \text{pct\_50}_i + \beta_4 \cdot \text{pct\_95}_i \tag{2}$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{6}$$

$$\beta_4 \sim \text{Normal}(0, 2.5) \tag{7}$$

$$\sigma \sim \text{Exponential}(1) \tag{8}$$

In the above model:

- μ_i is the predicted mean speed (`spd_mean`) for the i-th observation given the total traffic volume and speed percentiles.
- β_0 is the coefficient for the intercept.
- β_1 is the coefficient for the predicted change in the mean speed given a one-unit increase in traffic volume (`volume`).

9

- β_2 is the coefficient for the predicted change in the mean speed given a one-unit increase in the 5th percentile speed (`pct_05`).
- β_3 is the coefficient for the predicted change in the mean speed given a one-unit increase in the 50th percentile speed (`pct_50`).
- β_4 is the coefficient for the predicted change in the mean speed given a one-unit increase in the 95th percentile speed (`pct_95`).
- σ is the standard deviation of the residuals, representing unexplained variability in the mean speed.

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

3.1.1 Model justification

The above model was chosen to investigate the relationship between the average speed (`spd_mean`) of vehicles and key predictors, including total traffic volume (`volume`) and speed percentiles (`pct_05`, `pct_50`, and `pct_95`). These variables were selected because they provide a comprehensive view of traffic patterns and vehicle behavior across different speed thresholds.

The percentiles represent critical thresholds of speed distribution, with the 5th percentile (`pct_05`) capturing the slowest vehicles, the 50th percentile (`pct_50`) representing the median speed, and the 95th percentile (`pct_95`) reflecting the fastest vehicles. These thresholds are expected to influence the average speed significantly. Similarly, total traffic volume (`volume`) accounts for overall road congestion and is anticipated to impact the mean speed inversely, as higher volumes may reduce average speeds due to congestion.

We employ a Bayesian linear regression model implemented using the `rstanarm` package (Goodrich et al. 2022). This approach provides a probabilistic framework for parameter estimation and allows the incorporation of prior information. Default priors were used for the regression coefficients (`Normal(0, 2.5)`) and residual standard deviation (`Exponential(1)`), as they are appropriate for general regression problems without overly constraining the estimates.

The linearity assumption is reasonable given the nature of traffic data, where incremental changes in predictors like percentiles and volume are expected to have proportional effects on the mean speed. However, this model does not capture potential nonlinearities or interactions, which could be explored in future work if evidence suggests their relevance.

We hypothesize that:

- Higher 5th, 50th, and 95th percentile speeds will correspond to higher average speeds.
- Higher traffic volumes will be associated with lower average speeds due to congestion effects. This model's simplicity balances interpretability with analytical rigor, making it a suitable choice for understanding traffic patterns and their determinants.

Table 5: Model summary of the predicted impact of traffic volume and speed percentiles on mean speed.

	(1)
(Intercept)	−0.881
volume	0.000
pct_05	0.206
pct_50	0.652
pct_95	0.184
Num.Obs.	31 990
R2	0.992
R2 Adj.	0.992
Log.Lik.	−41 414.100
ELPD	−41 436.5
ELPD s.e.	385.2
LOOIC	82 872.9
LOOIC s.e.	770.4
WAIC	82 872.9
RMSE	0.88

4 Results

4.1 Model Coefficients

The Bayesian linear regression model results, summarized in Table 5, provide insight into the relationship between mean traffic speed (`spd_mean`) and key predictors such as speed percentiles and traffic volume.

The coefficient estimates reveal the following relationships:

- Intercept (`-0.881`): This represents the baseline adjustment to the mean speed when all predictors are at their reference levels.
- Traffic Volume (`0.000`): The coefficient indicates no measurable impact of total traffic volume on mean speed, which is unexpected given traditional congestion dynamics.
- 5th Percentile Speed (`0.206`): Lower speeds positively influence the mean, showing that slower-moving vehicles contribute moderately to the overall average.
- 50th Percentile Speed (`0.652`): Median speeds have the largest positive effect on mean speed, emphasizing their central role in shaping the distribution of traffic speeds.

- 95th Percentile Speed (`0.184`): Higher speeds have a smaller but still positive influence on mean speed, reflecting their contribution to the upper end of the distribution.

The model's performance metrics underscore its robustness:

- R² and Adjusted R² (`0.992`): The model explains 99.2% of the variance in mean speed, indicating excellent fit.
- Log-Likelihood (`-41414.100`): Represents the model's likelihood given the data.
- WAIC and LOOIC (`82872.9`): These information criteria indicate model performance, balancing goodness-of-fit and model complexity.
- RMSE (`0.88`): A low root mean square error confirms the model's predictive accuracy.

These results highlight the dominant role of speed percentiles in determining mean speed while raising questions about the negligible impact of traffic volume. Further exploration into potential interactions or omitted variables could clarify this finding.

4.2 Actual vs. Predicted Mean Speeds

To evaluate the predictive performance of our Bayesian linear regression model, we compare the predicted mean speeds, derived from the model, with the actual observed mean speeds in the dataset. The comparison is visualized in Figure 3.
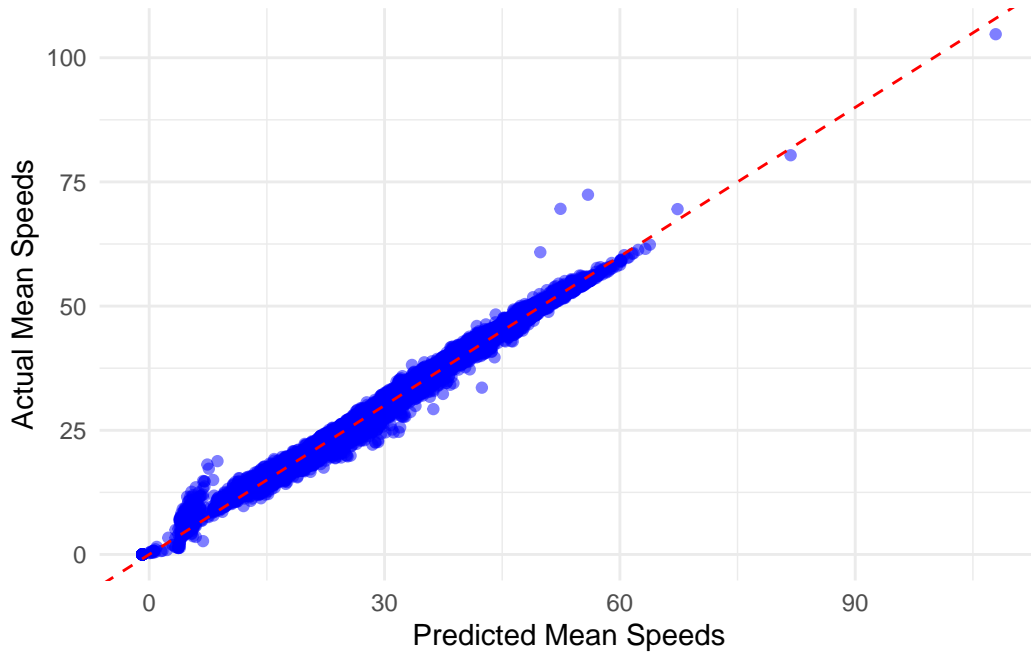

Figure 3: Comparison of actual and predicted mean speeds.

In Figure 3, each point represents an observation where the x-axis denotes the predicted mean speed, and the y-axis shows the actual mean speed. The red dashed line represents the ideal case where predictions perfectly match the actual values (i.e., a 45-degree reference line). Observations closely aligning with the dashed line indicate a strong agreement between predicted and actual values.

The model demonstrates strong predictive accuracy, as the majority of points cluster tightly along the dashed line with minimal deviation. This close alignment suggests that the model captures the relationship between the predictors (`pct_05`, `pct_50`, `pct_95`, and `volume`) and the response variable (`spd_mean`) effectively, with minimal systematic bias.

Notably, there are only a few points where slight deviations from the line are visible, mostly at higher mean speed values. This could be attributed to minor outliers or unique conditions not fully explained by the model's predictors. Overall, the results confirm the reliability of the Bayesian model in predicting mean traffic speeds, especially for the typical speed ranges observed in the dataset.

We explore the broader implications of these results, including potential improvements or limitations in the model, in Section 5.

5 Discussion

5.1 Post-pandemic trends and implications for traffic management

The analysis in Figure 1 illustrates the significant divergence in traffic patterns between the pre- and post-2020 periods. Post-pandemic traffic volumes are notably higher and are accompanied by an increase in extreme speeds (vehicles traveling over 100 km/h). This trend may reflect the rapid population growth in Toronto, driven by immigration, alongside shifts in urban mobility post-COVID-19.

The steeper slope for post-2020 data in Figure 1 suggests that extreme speed counts have become more sensitive to increases in traffic volume. One potential explanation for this is the stress placed on infrastructure as more drivers navigate an increasingly dense urban network. These findings highlight the importance of re-evaluating existing traffic management strategies and prioritizing policies that address the dangers posed by extreme speeds in high-traffic environments.

Targeted interventions, such as extending safety zones, introducing stricter enforcement of speed limits, or implementing dynamic speed regulation systems, could help mitigate the risks associated with this upward trend in extreme speeds. Additionally, these findings underscore the importance of accounting for macro-level disruptions, like the pandemic, in transportation policy design to better anticipate and address evolving urban mobility challenges.

5.2 Variability in speed percentiles and safety implications

The trends in speed percentiles visualized in Figure 2 reveal nuanced changes in traffic behavior over time. While the 95th percentile speed has shown a steady decline to below 50 km/h, suggesting progress in limiting higher speeds, the sharp drop in the 5th percentile speed since mid-2022 warrants attention. A decline from approximately 10 km/h to 5 km/h may indicate increased congestion, construction zones, or other road conditions causing slower-moving traffic. This could create a wider speed differential among vehicles, potentially leading to more frequent and severe accidents.

The stability of the 50th percentile speed at approximately 30 km/h reflects consistency in typical driving conditions, while the steady spread between the 5th and 95th percentiles (35–40 km/h) points to minimal change in speed variability. However, the importance of addressing the outliers at the extremes cannot be overstated, as they have disproportionate safety implications. This analysis supports the need for targeted measures to minimize variability in speeds, including enhanced road design and enforcement strategies to reduce abrupt accelerations or decelerations.

5.3 Limitations in modeling and data interpretation

While the Bayesian linear regression model employed in this paper provides meaningful insights into the relationship between key predictors and mean vehicle speed, several limitations warrant discussion. As highlighted in Section 3, the model does not account for potential interactions among predictors, such as the interplay between traffic volume and percentile speeds. For instance, higher traffic volumes may disproportionately affect the 5th percentile speed compared to the 95th percentile, but this relationship remains unexplored due to the model's linear structure.

Additionally, as noted in Section 2, the dataset's reliance on radar-equipped safety zone signs introduces biases related to location and driver behavior. Vehicles may reduce speed upon approaching these signs, potentially underestimating typical speed distributions for the broader road network. Moreover, the dataset aggregates observations monthly, which may obscure short-term fluctuations in traffic patterns.

Future modeling efforts could incorporate interaction terms and hierarchical structures to better capture the complexity of traffic dynamics. Similarly, integrating data from diverse monitoring systems (e.g., highway cameras or GPS datasets) could provide a more comprehensive view of Toronto's traffic behavior across different contexts.

5.4 Challenges and Opportunities in Traffic Enforcement

Automated enforcement systems present both significant opportunities and challenges for traffic management in Toronto. Evidence from cities like Washington, D.C., and New York City

highlights the potential of these systems to address traffic violations effectively. For example, Washington, D.C.'s bus-mounted cameras issued over 147,000 tickets to drivers illegally using bus-only lanes within months of implementation, showcasing their efficiency in catching violations that might otherwise go unnoticed. Similarly, New York's Automated Bus Lane Enforcement (ABLE) program has been expanded to address issues like double parking, further demonstrating the adaptability of automated systems (Elliott 2024).

In Toronto, while automated speed enforcement (ASE) cameras have successfully reduced speeding, the city's adoption of similar technology for other traffic violations, such as blocking intersections ("blocking the box") or driving in transit lanes, has been limited. This is partly due to legislative constraints at the provincial level, which prevent issuing fines for certain violations if the vehicle owner was not the driver at the time. Additionally, despite the proven efficacy of ASE cameras in reducing vehicle speeds, as noted by the significant decrease in speeding incidents in school zones, the rollout of new cameras has been slow.

These delays underscore broader challenges in traffic enforcement, including bureaucratic hurdles, resource constraints, and reliance on traditional enforcement methods. Moreover, the cultural perspective within traffic policing, which often deprioritizes ticketing for non-criminal violations, exacerbates these issues. However, these challenges also present opportunities for leveraging technology to complement traditional enforcement and create a more comprehensive traffic management framework.

5.5 Policy recommendations and future directions

The findings in this study offer a foundation for improving traffic safety and efficiency in Toronto. A central takeaway is the value of automated enforcement in addressing critical challenges, including extreme speeds, variability in traffic patterns, and enforcement gaps. Automated speed enforcement (ASE) cameras have already demonstrated significant success in Toronto, reducing both the prevalence of speeding and overall vehicle speeds in monitored areas. Expanding the use of these cameras and integrating new functionalities, such as detecting transit lane violations or intersection blockages, can further enhance their impact. Learning from other cities like Washington, D.C., and New York, Toronto can adopt similar innovations to target a broader range of traffic offenses (Elliott 2024).

Urban road design remains another critical area for intervention. Consistent median speeds and limited speed variability are essential for maintaining safe and efficient traffic flow. To achieve this, urban planning efforts should focus on designing roads that naturally discourage dangerous speeds, while avoiding features that create bottlenecks or increase risks for vulnerable road users. For instance, implementing physical design elements such as raised crosswalks or lane-narrowing strategies can complement automated enforcement measures.

Adaptive traffic policies also hold promise for addressing the dynamic nature of urban mobility. Introducing technologies that adjust speed limits in real time based on traffic conditions, weather, or specific safety concerns could help balance safety with mobility. By integrating

ASE devices into such adaptive frameworks, the city can enhance their effectiveness in high-risk or high-traffic areas.

To support these efforts, data collection and analysis must be expanded. Toronto can benefit from leveraging complementary data sources, such as GPS datasets, time-of-day analyses, and seasonal traffic patterns. Such data can refine the deployment of automated enforcement and inform broader policy decisions. For example, cities like Washington, D.C., have used bus-mounted cameras to monitor transit lane violations effectively. Toronto could similarly implement pilot programs to gather data and address challenges such as "blocking the box" and illegal lane usage.

Finally, addressing legislative barriers is essential for unlocking the full potential of automated enforcement. While current laws limit certain applications, proactive steps—such as deploying cameras to collect data even before legislative approval for expanded ticketing—can help build the case for broader implementation. The city's leaders should take bold steps, pushing forward with camera deployment to enhance safety and traffic management, even if provincial approvals are delayed. This approach, modeled after prior successes in advocating for speed cameras, could expedite the necessary legislative changes.

By integrating automated enforcement, adaptive policies, and data-driven design into a cohesive strategy, Toronto can tackle its pressing traffic challenges and create safer, more efficient urban mobility systems.

Appendix

A Data Observation

A.1 Idealized Methodology for Traffic Data Collection

The observational data analyzed in this study stems from the City of Toronto's Safety Zone Watch Your Speed Program (WYSP), where radar-equipped speed signs monitor and aggregate vehicle speeds. While this data provides valuable insights, it also introduces limitations such as restricted spatial and temporal coverage, potential behavioral biases due to the visibility of speed signs, and an inability to capture all traffic volume accurately. This section proposes an idealized methodology to address these limitations and enhance the robustness of traffic speed data collection.

A.1.1 Goals of the Idealized Methodology

The idealized methodology aims to:

- Expand Representativeness: Ensure the data reflects traffic patterns across different locations, road types, and times of the day.
- Minimize Behavioral Biases: Reduce the influence of drivers' behavior changes caused by the visibility of monitoring devices.
- Capture Temporal Variability: Account for daily, seasonal, and situational fluctuations in traffic patterns.
- Improve Accuracy: Ensure that all traffic, including vehicles that might evade radar detection, is measured consistently.

A.1.2 Proposed Framework

- Stratified Sampling Approach

  - To improve spatial representativeness, Toronto can be divided into strata based on geographic regions (e.g., downtown, suburban, and rural areas) and road types (e.g., highways, arterial roads, local streets). The number of monitoring devices deployed in each stratum should be proportional to traffic density and road importance. This stratified approach ensures adequate representation of all areas and road conditions.

- Rotational Monitoring

- Rather than using permanent installations, monitoring devices should be deployed on a rotational basis. Devices can be moved periodically across selected sites within each stratum, ensuring broader spatial coverage. Rotation would also reduce drivers' habituation to the presence of speed signs, which might otherwise alter their driving behavior.

- Integration of Covert Monitoring

  - To minimize behavioral biases, the methodology should integrate covert monitoring techniques. Hidden radar devices or automated cameras can measure vehicle speeds without drivers being aware of their presence, thus capturing more natural driving behavior.

- Temporal Sampling

  - Data collection should account for temporal variability by monitoring during different times of day (e.g., peak vs. off-peak hours), days of the week (e.g., weekdays vs. weekends), and seasons. This approach ensures a comprehensive understanding of traffic patterns and their fluctuations.

- Technological Enhancements

  - Advanced technologies such as LiDAR and AI-based sensors can supplement traditional radar devices to increase measurement accuracy. These technologies can detect multiple vehicles simultaneously, reducing blind spots and capturing speeds across a wider field of view.

- Calibration and Standardization

  - All monitoring devices must undergo regular calibration to maintain consistency and accuracy in speed measurements. A centralized standardization protocol should be implemented to ensure uniformity across all collected data.

A.1.3 Challenges and Practical Considerations

Implementing this idealized methodology involves logistical and financial challenges:

- Cost Implications: Deploying advanced devices and rotational monitoring systems requires significant investment in equipment and personnel.
- Coordination with Stakeholders: Accessing certain road segments for device installation may require coordination with local authorities, businesses, and residents.
- Data Management: Large-scale data collection introduces challenges related to storage, integration, and processing. Ensuring data integrity and consistency is critical for effective analysis.

Despite these challenges, the proposed framework offers a robust and realistic solution for improving traffic data collection. By addressing the limitations of the existing WYSP dataset, this methodology provides a foundation for more accurate and actionable insights into urban mobility.

A.2 Simulation for Evaluating the Proposed Framework

Simulations can be an effective way to evaluate the proposed methodology before large-scale implementation. A simulation study could:

- Model Traffic Flow: Use agent-based modeling to simulate traffic flow across Toronto, incorporating variables such as road type, time of day, and vehicle characteristics.
- Test Monitoring Scenarios: Evaluate the effectiveness of rotational and covert monitoring setups in capturing accurate speed data across different strata. Assess Behavioral Impact: Simulate changes in driver behavior due to visible and hidden monitoring devices to quantify potential biases.
- Optimize Resource Allocation: Simulate various deployment strategies to determine the optimal allocation of monitoring devices for maximum coverage and cost efficiency.

By simulating these scenarios, stakeholders can refine the methodology and address potential challenges before field deployment.

A.3 Automated Speed Enforcement (ASE) Evaluation in Toronto

An evaluation of three years of data from Toronto's Automated Speed Enforcement (ASE) cameras provides valuable insights into the potential of automated systems to influence driver behavior and enhance traffic safety. Conducted by researchers from SickKids Hospital and Toronto Metropolitan University, this study highlights the effectiveness of ASE in reducing speeding across multiple speed zones (Rodrigues 2023).

From January 2020 to December 2022, the City of Toronto installed 50 ASE cameras, rotating them to different locations periodically, and later added 25 more cameras to increase coverage. The study reported the following significant findings:

- Reduction in Speeding: The proportion of vehicles speeding decreased by 45% in areas with an ASE device. In specific zones:

  - 30 km/h zones: Speeding dropped from 60% to 43%.
  - 40 km/h zones: Speeding dropped from 51% to 30%.
  - 50 km/h zones: Speeding dropped from 58% to 36%.

- Decrease in Operating Speeds: Vehicle operating speeds decreased by approximately 7 km/h overall in areas with ASE cameras:

- 30 km/h zones: From 44 to 37 km/h.
  - 40 km/h zones: From 50 to 44 km/h.
  - 50 km/h zones: From 63 to 60 km/h.

- Reduction in Excessive Speeds: The proportion of drivers exceeding speed limits by 20 km/h or more dropped by 87% following ASE implementation.

- Widespread Impact: Speed limit compliance improved at 80% of locations equipped with ASE devices.

This data demonstrates ASE's substantial role in curbing excessive speeding and creating safer road environments, particularly near schools and residential areas. The findings align with broader efforts to design data-driven traffic safety policies and underscore the importance of integrating automated systems into traffic management strategies.

A.3.1 Implications for Observational Data Methodology

The ASE program also reflects several elements of an idealized data collection methodology:

- Rotational Deployment: Cameras were rotated across locations to broaden coverage and minimize behavioral biases from static monitoring.
- Comprehensive Data Collection: Despite some limitations imposed by the COVID-19 pandemic, the study captured substantial data during periods of normal activity, enhancing its reliability.
- Focused Interventions: Placement of ASE cameras in high-risk areas, such as school zones, demonstrated a targeted approach to improving road safety.

This case study provides a model for integrating automated enforcement with observational data collection, highlighting its efficacy in reducing risky driving behaviors and promoting safer urban mobility.

A.4 Strengths and Limitations of Observational Data

The WYSP dataset offers valuable observational insights into Toronto's traffic patterns, but it is not without its limitations.

A.4.1 Strengths

- Ease of Collection: Radar speed signs provide a non-intrusive method for collecting large amounts of data over extended periods.
- Focus on Safety Zones: Data collection in designated Safety Zones highlights areas with specific traffic and safety concerns.

- Detailed Metrics: The dataset includes percentile-based speed metrics (e.g., 5th, 50th, and 95th percentiles), which are critical for analyzing traffic behavior.

A.4.2 Limitations

- Limited Spatial Coverage: The dataset only includes areas with installed speed signs, omitting many road segments across Toronto.
- Behavioral Biases: Visible speed signs may influence drivers to temporarily reduce their speed, resulting in data that may not reflect typical behavior.
- Temporal Gaps: Data is aggregated monthly, obscuring finer temporal variations such as hourly or daily patterns.
- Non-Representative Traffic Volume: The count of vehicles measured by speed signs is not equivalent to total traffic volume, as some vehicles may evade detection.

B Model details

B.1 Posterior predictive check

The posterior predictive check (PPC) evaluates how well the Bayesian model fits the observed data (Alexander 2023). In Figure 4, the black line represents the observed data density, while the blue lines represent the density of replicated data generated from the posterior predictive distribution. The alignment between these lines indicates the model's ability to reproduce the data. While there is a slight deviation at the peak of the density curve, the overall alignment suggests that the model provides a reasonable fit to the observed data. Minor discrepancies may warrant adjustments in the model specification or inclusion of additional predictors to improve fit.
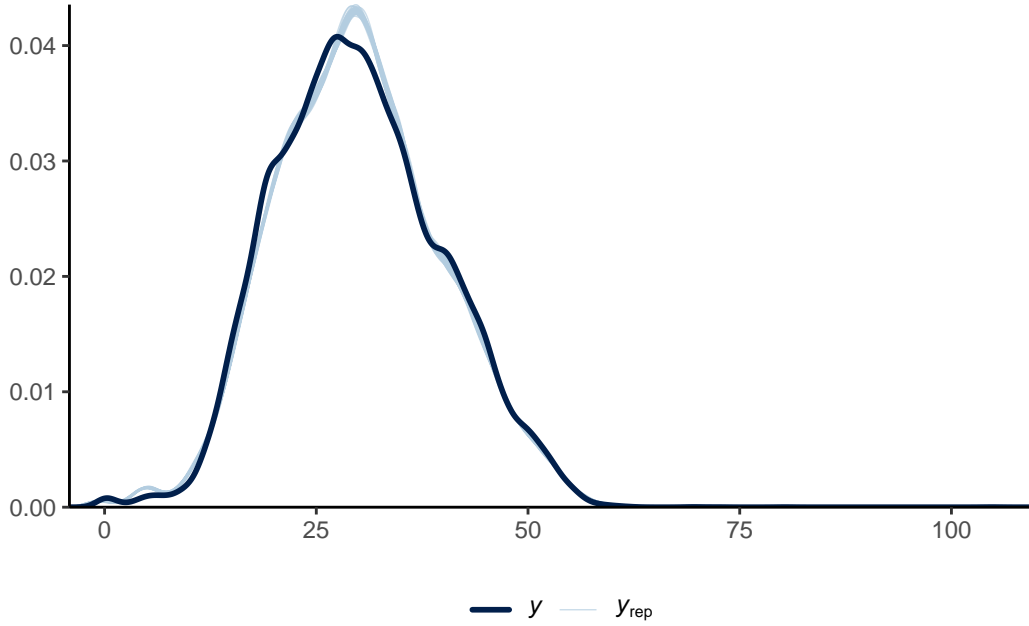
Figure 4: Examining how the model fits the data through a posterior predictive check

B.2 Comparison of the Posterior and Prior

Figure 5 compares the prior and posterior distributions for the model parameters. The posterior (left) is informed by the observed data, while the prior (right) reflects initial assumptions (Alexander 2023). Notably, the posterior for all parameters, except the intercept and sigma, remains centered around zero, showing minimal change from the prior. The sigma parameter has shifted slightly to the right, reflecting a more informed estimate based on the data. The red intercept dot spans the entire horizontal range in the prior, indicating a vague prior assumption. This comparison illustrates that the data provides substantial information for refining parameter estimates.
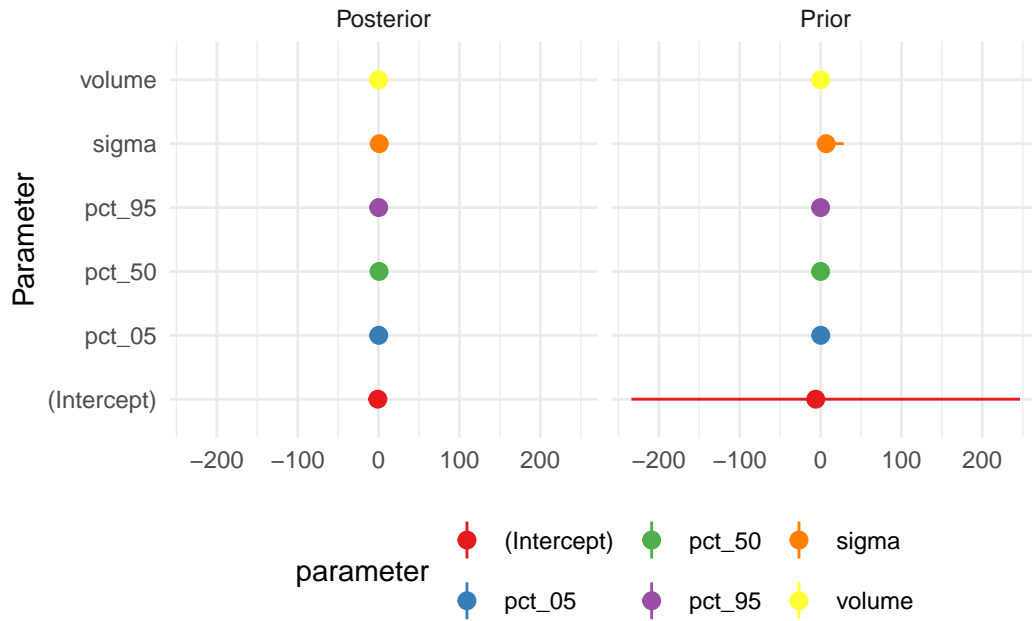
Figure 5: Comparison of posterior and prior distributions for model parameters

B.3 Markov Chain Monte Carlo Convergence

B.3.1 Trace Plot

The trace plot in Figure 6 examines the behavior of the MCMC chains for each parameter (Alexander 2023). The chains exhibit stable horizontal fluctuations with no discernible trends, indicating proper mixing and convergence. The patterns across all six parameters are consistent, suggesting that the algorithm has adequately explored the parameter space.
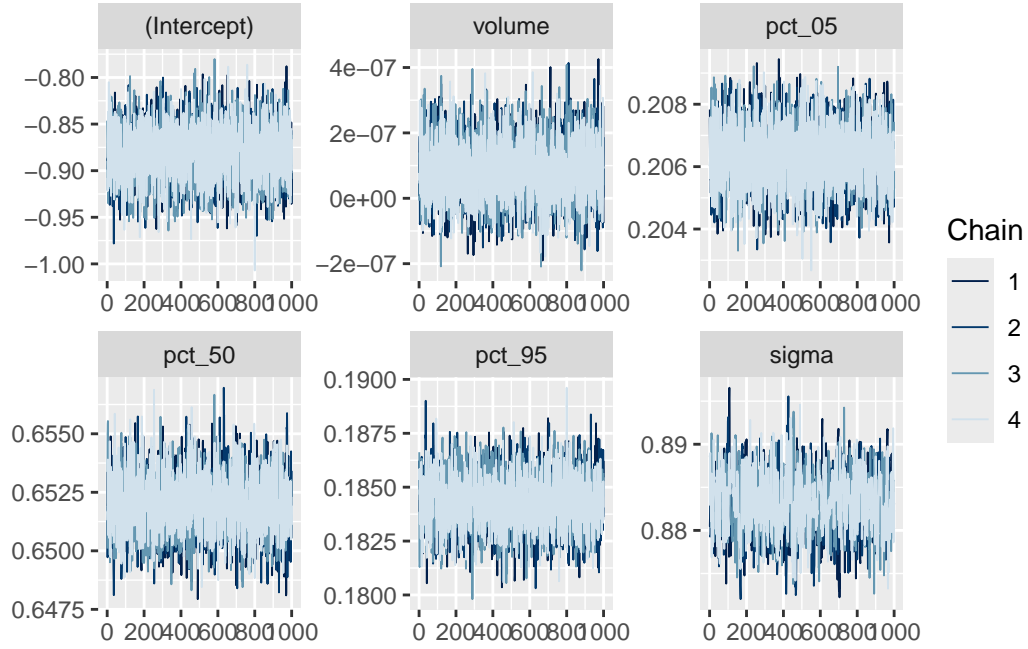
Figure 6: Trace plot for MCMC convergence assessment

B.3.2 Rhat Plot

The Rhat plot in Figure 7 evaluates the convergence of the Markov Chain Monte Carlo (MCMC) algorithm. The Rhat values for all parameters hover just above 1.00, indicating acceptable convergence. While `pct_05` displays the longest interval, it still falls within the acceptable range. This diagnostic suggests that the chains for all parameters have reached convergence and that the posterior samples are reliable.
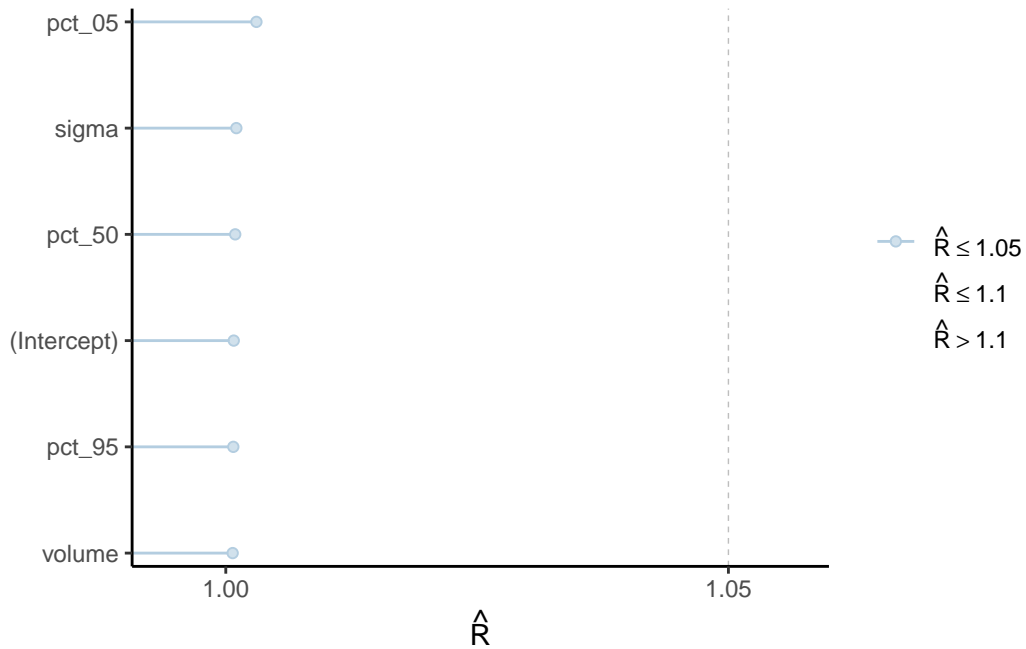
Figure 7: Rhat plot for MCMC convergence assessment

B.4 Credible Intervals

The credible intervals for the model coefficients are shown in Figure 8. The intercept is negative, suggesting a downward adjustment for baseline mean speed. The `volume` coefficient is centered at 0, indicating little influence on the predicted mean speed. Among the predictors, `pct_50` has the strongest positive effect, with a credible interval between 0.5 and 1.0, while `pct_05` and `pct_95` show moderate positive effects. The residual standard deviation (`sigma`) is near 1.0, reflecting the model's uncertainty in capturing variability.
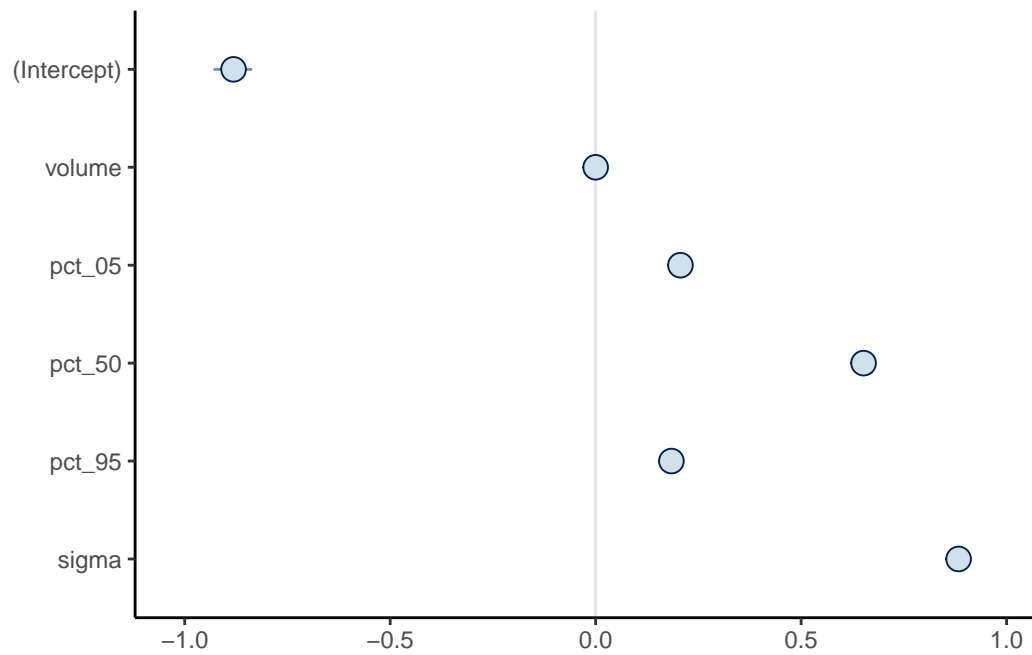
Figure 8: 95% Credible Intervals for Model Coefficients

References

Alexander, Rohan. 2023. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Dumas, Raphael. 2024. "WYSP Monthly Summary." City of Toronto, OpenData. https://open.toronto.ca/dataset/safety-zone-watch-your-speed-program-monthly-summary/.

Elliott, Matt. 2024. "Doug Ford Hasn't Let Toronto Use Cameras to Get Traffic Moving. The City Should Do It Anyway." *Toronto Star.* https://www.thestar.com/opinion/contributors/doug-ford-hasnt-let-toronto-use-cameras-to-get-traffic-moving-the-city-should-do/article_e44cd686-6ebc-11ef-ab00-ff8fb941e810.html.

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

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

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.

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

Rodrigues, Gabby. 2023. "Toronto's Speed Cameras Show a Reduction in Speeding, Report Indicates." *Global News.* Global News. https://globalnews.ca/news/9824834/toronto-speed-cameras-report-reduced-speeding/.

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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.