Understanding Traffic Patterns and Median Speeds: Insights from Toronto's Speed Display Program*

Northbound and Southbound Traffic Show Higher Speeds; High-Speed Vehicles Raise Median Speeds

Xingjie Yao

December 1, 2024

This study investigates how traffic direction and high-speed vehicles affect median speeds at locations with speed display signs in Toronto. Northbound and Southbound traffic showed higher median speeds than Eastbound, and more high-speed vehicles raised the median speed. These findings highlight how road design and driver behavior influence traffic patterns. The results inform strategies to improve road safety by addressing high-speed driving and targeting high-risk areas.

1 Introduction

Speeding behavior has long been a focus of traffic safety research, with studies highlighting the psychological and situational factors influencing drivers' speed choices (Fildes, Rumbold, and Leening 1991). Drivers' attitudes toward risk, control, and perceived enforcement often play a critical role in determining whether they comply with speed limits or exceed them (Fildes, Rumbold, and Leening 1991). Understanding these attitudes is essential for designing effective interventions that target both behavioral and environmental contributors to speeding (Fildes, Rumbold, and Leening 1991). By situating these insights within the context of Toronto's WYSP, this study explores how structural factors, such as traffic direction, and behavioral tendencies, such as high-speed driving, shape observed speed patterns (Fildes, Rumbold, and Leening 1991). Speed-related traffic accidents are a major concern in urban areas, prompting cities to explore interventions such as speed display signs to encourage safer driving behavior. The Mobile Watch Your Speed Program (WYSP) in Toronto rotates radar-equipped signs across various locations to remind drivers of their speeds and promote compliance with posted

*Code and data are available at: https://github.com/Stella41603/Toronto-Speeding-Analysis.git.

limits. Despite the program's widespread use, the impact of these installations on traffic speed patterns has not been thoroughly analyzed. This paper investigates the relationship between traffic direction, high-speed vehicle prevalence, and median driving speeds at WYSP locations. By leveraging a Bayesian linear regression model, we aim to uncover patterns that can inform traffic management policies and improve road safety.

The primary estimand of interest is the conditional average of median speed observed at each sign installation given two predictors: traffic direction (categorized as Eastbound, Northbound, Southbound, and Westbound) and the count of vehicles exceeding 100 km/h at the same location. The median speed serves as a summary measure of central driving tendencies, while traffic direction and high-speed vehicle counts provide explanatory variables to capture both structural and behavioral factors influencing speed. The analysis seeks to quantify how these factors shape typical driving behavior at monitored locations.

The analysis reveals that traffic direction significantly influences median speed. Compared to Eastbound traffic, Northbound and Southbound directions exhibit higher median speeds, suggesting potential differences in road characteristics, flow, or traffic patterns. In contrast, Westbound traffic shows a slight reduction in median speed relative to Eastbound traffic. Additionally, the count of vehicles exceeding 100 km/h is positively associated with median speed, with each additional high-speed vehicle contributing incrementally to the overall median. The Bayesian model, using weakly informative priors, provides robust estimates that align with observed patterns, explaining approximately 12.9% of the variation in median speeds. These findings highlight the combined effects of structural (directional) and behavioral (high-speed driving) factors on traffic dynamics.

Understanding how traffic direction and high-speed behavior affect median speed has direct implications for traffic safety and urban planning. These findings can guide policymakers in targeting specific road segments with higher observed speeds for interventions, such as additional signage or enforcement. Moreover, the results underscore the importance of addressing high-speed driving, as even a small number of speeding vehicles can substantially influence overall traffic flow. By identifying patterns and predictors of median speed, this study contributes to a growing body of evidence that supports data-driven approaches to improving road safety. The insights gained from this research not only enhance the evaluation of programs like WYSP but also provide actionable recommendations for broader traffic management strategies.

The remainder of this paper is organized as follows. Section 2 provides an overview of the data sources and key variables used in our analysis, offering a detailed description of the Mobile Watch Your Speed Program (WYSP) dataset. Section 3 outlines the Bayesian linear regression framework used to explore the relationship between these variables. Section 4 presents the findings, highlighting the factors that influence median vehicle speeds and exploring how traffic direction and the frequency of speeding contribute to observed patterns. Section 5 interprets the broader implications of the findings, particularly in the context of traffic safety and speed compliance.

2 Data

2.1 Overview

For this analysis, we used the R programming language (R Core Team 2023) to explore data from the Mobile Watch Your Speed Program (WYSP) – Speed Summary in Toronto. The dataset, sourced from Toronto's Open Data Portal (Gelfand 2022; Toronto 2024), provides a summary of vehicle speeds observed at temporary speed display sign installations across the city. These signs are part of an initiative to remind motorists to obey speed limits by displaying their real-time speeds. Key aspects analyzed include the median speed observed during each installation, the count of vehicles exceeding 100 km/h, and the cardinal direction of traffic flow.

A variety of R packages were employed to streamline data manipulation, modeling, and visualization. The `tidyverse` suite served as the backbone for data processing, enabling efficient and reproducible workflows (Wickham et al. 2019). File management was handled with `here`, ensuring consistent and reliable dataset access (Müller 2020). Data cleaning was facilitated by `janitor`, which prepared the dataset for analysis (Firke 2023), while `testthat` validated the simulated and processed datasets, ensuring workflow integrity (Wickham 2011). Bayesian modeling, conducted using `rstanarm`, provided robust inference and clear probabilistic interpretations (Brilleman et al. 2018). Model summaries were generated with `modelsummary` for concise, intuitive reporting (Arel-Bundock 2022). Additionally, `arrow` ensured fast and memory-efficient handling of the dataset, critical for its scale and complexity (Richardson et al. 2024). Best practices in workflow organization, as outlined in Alexander (2023), ensured transparency and reproducibility, integrating data cleaning, modeling, and visualization into a unified pipeline. This structured approach allowed for meaningful insights while supporting validation and future research extensions.

The dataset, sourced from Toronto's Open Data Portal, underwent extensive cleaning to ensure its suitability for analysis. Key variables, such as the median speed (50th percentile), the count of vehicles exceeding 100 km/h, and the direction of travel, were retained, while irrelevant fields were excluded to streamline the dataset. Observations with missing or incomplete data were removed to uphold consistency and accuracy. Although alternative datasets, such as detailed hourly speed counts, were available, the WYSP summary dataset was selected for its clarity and direct relevance to evaluating driver compliance with speed limits. This focused dataset offered a concise yet comprehensive overview of observed speed patterns, supporting the analysis of program effectiveness. The rigorous cleaning process ensured that the dataset was well-prepared for modeling and interpretation, enabling a reliable assessment of how WYSP sign installations influence vehicle speeds and driver behavior.

2.2 Measurement

The dataset from the Mobile Watch Your Speed Program (WYSP) represents a systematic effort to capture real-world driving behaviors and translate them into structured data for analysis. This process begins with the installation of speed display signs, which are portable devices equipped with radar technology and LED screens. These signs are mounted on existing infrastructure, such as hydro poles or streetlights, and measure the speed of oncoming vehicles. The radar records the speed, and the LED display shows it to drivers in real time, encouraging them to adhere to the posted speed limits. The act of vehicles traveling past these signs, with their speeds being detected and displayed, forms the raw phenomenon in the real world that is subsequently recorded in the dataset.

The radar device captures each vehicle's speed as a numeric value, and these measurements are aggregated into summary statistics such as the median speed and the count of vehicles exceeding 100 km/h. Each installation is associated with a specific location, date, and direction of traffic flow (e.g., Eastbound, Northbound), creating a comprehensive record of speed observations at that site. These data points are then uploaded and organized by the city into a structured dataset. This transformation process converts individual speed measurements into summarized and standardized entries, making it possible to analyze patterns across different locations and time periods.

The dataset reflects not only the behavior of drivers but also the operational parameters of the program. For example, the mobile signs are rotated across Toronto's wards on a monthly basis, ensuring broad coverage and diverse data collection. This systematic rotation is documented in the dataset, linking speed summaries to specific sign installations and the conditions under which they were observed. The inclusion of these operational details adds context to the speed data, allowing analysts to account for factors such as location-specific road conditions or seasonal variations in driver behavior.

However, there are limitations to this process that must be acknowledged. The dataset only includes data from installations where the mobile sign has been removed, meaning ongoing installations are excluded. Furthermore, the count of vehicles exceeding 100 km/h should not be interpreted as a total traffic volume count, as the dataset is limited to vehicles detected by the radar and does not include undetected traffic. Despite these constraints, the dataset provides a valuable representation of driver compliance with speed limits, offering insights into how temporary interventions like speed display signs influence behavior. This transformation from real-world phenomena to structured data highlights the importance of systematic measurement and the challenges of capturing human behavior in quantifiable terms. Additional details are given in Appendix C.

2.3 Outcome variable

The outcome variable in this analysis is the median speed observed at each location where the speed display signs were installed. This variable represents the typical speed of vehicles passing the radar-equipped signs and provides a summary measure of driver behavior at that site. By focusing on the median speed, the analysis captures a central tendency that is less influenced by extreme outliers, offering a reliable indication of how most drivers behave under the influence of speed display signs.

2.4 Predictor variables

2.4.1 Cardinal direction of travel

The direction of travel, categorized as Eastbound, Northbound, Southbound, or Westbound, is a key predictor in this analysis. Traffic direction reflects potential variations in road conditions, traffic volume, and environmental factors that may influence driver behavior. For instance, certain directions might correspond to busier thoroughfares, residential streets, or roads with differing speed limits, all of which could affect median speeds. Including direction as a predictor allows the analysis to account for these geographic and logistical variations.

2.4.2 Count of vehicles exceeding 100km/h

The count of vehicles exceeding 100 km/h serves as another critical predictor, reflecting the prevalence of excessive speeding at each location. This variable captures high-risk driving behavior, which is a major concern for road safety. By examining the relationship between the count of speeding vehicles and the median speed, the analysis investigates whether locations with frequent speeding incidents also exhibit higher overall driving speeds or distinct behavioral patterns. This variable provides valuable insight into the extent and impact of speeding on typical traffic behavior.

2.5 Relationship between variables

Figure 1 presents the median speeds of vehicles across four cardinal directions: Eastbound (EB), Northbound (NB), Southbound (SB), and Westbound (WB). The median speeds appear relatively consistent across directions, with minor variations. Northbound (NB) and Southbound (SB) directions exhibit slightly broader interquartile ranges compared to Eastbound (EB) and Westbound (WB), indicating greater variability in vehicle speeds. Outliers in all directions represent occasional instances of vehicles traveling at either extremely low or high speeds, potentially influenced by traffic conditions or local environmental factors. These findings provide insights into directional trends in traffic flow and support the evaluation of speed limit sign effectiveness in managing vehicle speeds. The analysis complements a separate examination of high-speed violations, quantified as the count of vehicles observed exceeding 100 km/h at these locations.

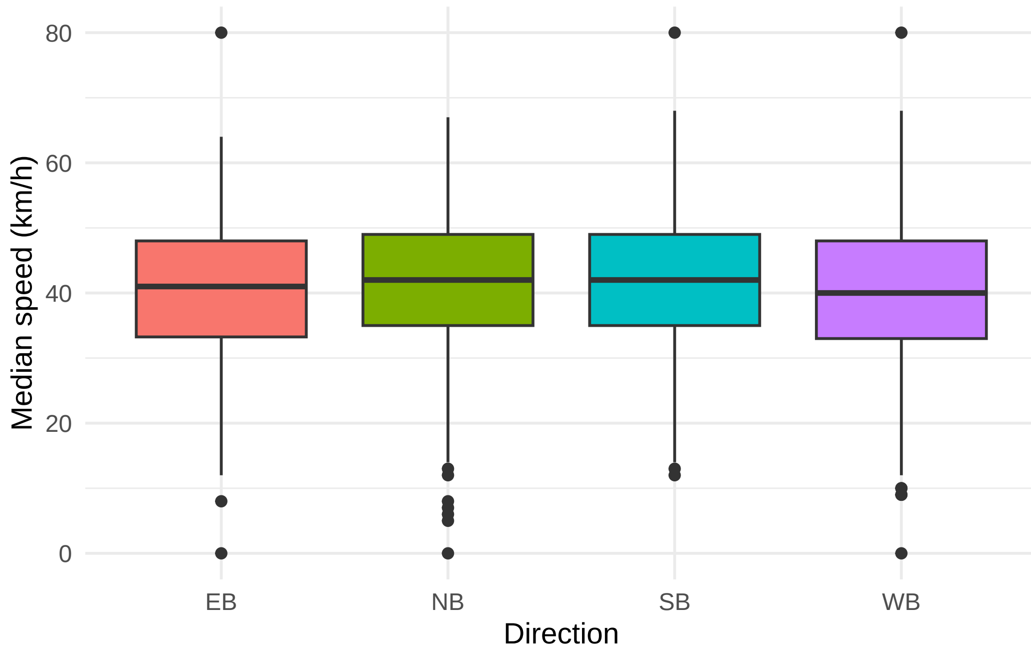


Figure 1: The distribution of the median speeds, measured in kilometers per hour (km/h), for vehicles traveling in each cardinal direction Eastbound (EB), Northbound (NB), Southbound (SB), and Westbound (WB), at monitored locations during the installation of traffic signs.

Figure 2 shows how the count of vehicles exceeding 100 km/h correlates with the median speed observed at various locations, grouped by traffic direction. The majority of locations cluster around moderate counts of speeding vehicles (below 100) and median speeds between 40 and 60 km/h. However, a few outliers with higher counts of speeding vehicles (above 300) or extreme median speeds suggest localized conditions, such as high-speed corridors or unique traffic environments. The directional grouping reveals no obvious separation in trends across Eastbound (EB), Northbound (NB), Southbound (SB), and Westbound (WB), suggesting that speeding patterns are relatively uniform across directions. This analysis highlights the interplay between speeding behaviors and overall traffic flow characteristics, providing critical information for targeting interventions to improve traffic safety.

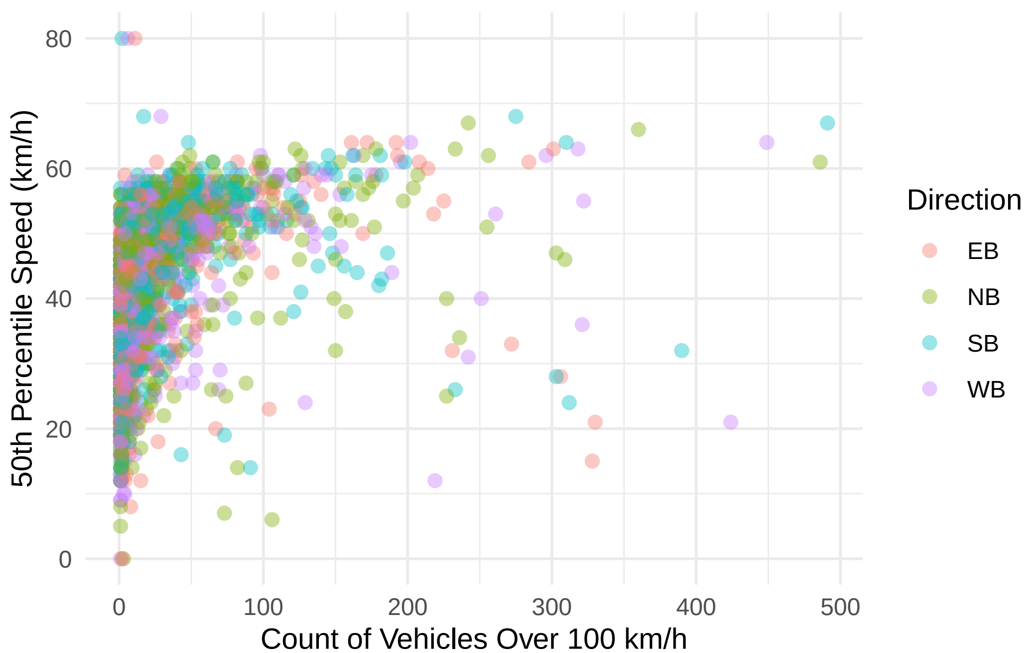


Figure 2: Illustrating the relationship between the count of vehicles observed exceeding 100 km/h and the median speed, measured in kilometers per hour, for each cardinal direction, Eastbound (EB), Northbound (NB), Southbound (SB), and Westbound (WB).

3 Model

The goal of our modeling strategy is twofold. Firstly, we aim to understand the factors influencing the median speed of vehicles at locations where speed display signs were installed. Secondly, we seek to quantify the relationships between median speed and key predictors such as traffic direction and the count of vehicles exceeding 100 km/h. Here, we describe the Bayesian linear regression model used to investigate these relationships. Background details, including diagnostic checks, are provided in Appendix B.

3.1 Model set-up

Define y_i as the median speed (in km/h) observed at location i. The predictors are x_{1i}, the count of vehicles exceeding 100 km/h, and x_{2i}, the traffic direction. Traffic direction is included as a set of dummy variables with Eastbound (EB) as the reference group. The model is specified as:

$$
\begin{aligned}
y_i \mid \mu_i, \sigma &\sim \text{Normal}\left(\mu_i, \sigma\right) \\
\mu_i &= \alpha + \beta_1 x_{1i} + \gamma_1 \text{NB}_i + \gamma_2 \text{SB}_i + \gamma_3 \text{WB}_i \\
\alpha &\sim \text{Normal}(0, 2.5) \\
\beta_1 &\sim \text{Normal}(0, 2.5) \\
\gamma_1, \gamma_2, \gamma_3 &\sim \text{Normal}(0, 2.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

Here, α represents the intercept (the expected median speed for Eastbound traffic when $x_{1i} = 0$ and the count of vehicles exceeding 100 km/h us 0), β_1 is the effect of the count of vehicles exceeding 100 km/h, and $\gamma_1, \gamma_2, \gamma_3$ are the effects of Northbound (NB), Southbound (SB), and Westbound (WB) directions compared to Eastbound (EB). The residual standard deviation σ models the variability in the median speed unexplained by the predictors. These priors are weakly informative, centered at zero with modest variance, allowing the data to dominate the posterior estimates while discouraging extreme values.

We implemented the model in `R` (R Core Team 2023) using the rstanarm package (Brilleman et al. 2018). The default priors from `rstanarm` were used for all parameters, as described above, to ensure consistency and interpretability.

3.1.1 Model justification

The chosen predictors, traffic direction and the count of vehicles exceeding 100 km/h, are included based on their theoretical and practical relevance to understanding speed behavior.

Traffic direction serves as a proxy for structural and environmental differences between roadways, such as varying traffic density, road design, or signalization, which can influence driving patterns. By using dummy variables for Northbound, Southbound, and Westbound directions with Eastbound as the reference, the model captures these directional variations in median speed, allowing for comparisons across different contexts. The count of vehicles exceeding 100 km/h is included as a continuous predictor to address behavioral influences on median speed. Locations with a higher frequency of speeding vehicles are hypothesized to exhibit elevated central speed tendencies due to social and psychological factors like peer influence or reduced perceived enforcement. Together, these predictors provide a comprehensive framework to disentangle structural and behavioral contributors to speed dynamics, ensuring that the analysis reflects both the physical and social dimensions of driving behavior.

Bayesian linear regression is well-suited for analyzing the continuous outcome of median speed, especially given its flexibility to incorporate prior information and quantify uncertainty in parameter estimates. Traffic direction is modeled using dummy variables to allow for direct comparison between the baseline (Eastbound) and other directions. This approach accounts for potential differences in road conditions, traffic patterns, or design features that vary by direction. The count of vehicles exceeding 100 km/h is included as a continuous predictor, hypothesized to positively correlate with median speed as locations with more speeding vehicles may also exhibit higher central tendencies in speed. The use of weakly informative priors ensures that the model is guided primarily by the observed data, providing robust and interpretable results while mitigating overfitting or extreme parameter estimates. This framework allows for a nuanced understanding of the factors shaping driver behavior across various locations and conditions.

3.1.2 Consideration of alternative models

To compare the prediction accuracy of the Bayesian and Frequentist regression models, the dataset was divided into a training set (80% of the data) and a testing set (20% of the data). Both models were trained on the training set using the same predictors: traffic direction and the count of vehicles exceeding 100 km/h. The Bayesian model incorporated prior beliefs about the parameter estimates, while the Frequentist model relied solely on the observed data for parameter estimation. After training, predictions were generated for the testing set. For the Bayesian model, predictions were derived by averaging the posterior predictive distributions for each observation, while for the Frequentist model, point predictions were directly calculated from the fitted parameters. Prediction accuracy was assessed using two metrics: Root Mean Squared Error (RMSE), which quantifies the average squared difference between predicted and observed values, and Mean Absolute Error (MAE), which measures the average absolute difference.

The results indicate that both models performed almost identically in predicting median speeds. The Bayesian model achieved an RMSE of 9.603 and an MAE of 7.526, while the Frequentist model had an RMSE of 9.605 and an MAE of 7.528. These results suggest that both approaches captured the key patterns in the data with similar levels of accuracy. The slightly lower RMSE and MAE for the Bayesian model indicate a marginal advantage in prediction, potentially due to its ability to incorporate prior uncertainty into the modeling process. However, the differences between the two models are negligible, implying that either method could be effectively used for this analysis.

The Bayesian model is preferred in this analysis for several reasons, despite the minimal differences in prediction accuracy. First, the Bayesian approach allows for the incorporation of prior information, which can help guide the estimation process, particularly in cases where data may be sparse or noisy. This can result in more robust and reliable parameter estimates, as the Bayesian framework explicitly accounts for uncertainty in both the data and the model parameters. Second, the Bayesian model provides a full posterior distribution for each parameter, enabling a richer interpretation of results and a clearer understanding of the uncertainty surrounding predictions. This is particularly advantageous for decision-making in contexts such as traffic safety, where understanding variability and risk is crucial. Lastly, the Bayesian model's slightly lower RMSE and MAE, while small, suggest a better fit to the data, likely due to its ability to balance the influence of extreme values and outliers. For these reasons, the Bayesian model is selected as the final model for this analysis, as it offers not only comparable prediction accuracy but also a more nuanced and interpretable framework for addressing the research questions.

4 Results

Our results are summarized in Table 1. The results of the Bayesian linear regression model provide insights into the factors influencing the median speed of vehicles at locations where speed display signs were installed. The intercept of 38.77 indicates that, on average, the median speed for Eastbound (EB) traffic is approximately 38.8 km/h when the count of vehicles exceeding 100 km/h is zero. This serves as the baseline for comparison with other directions.

The coefficients for traffic direction reveal notable differences relative to the Eastbound (EB) group. Northbound (NB) traffic has a higher median speed by 0.93 km/h, while Southbound (SB) traffic shows an even greater increase of 1.12 km/h. Conversely, Westbound (WB) traffic demonstrates a slight decrease of 0.33 km/h in the median speed compared to Eastbound traffic. These directional variations may reflect differences in road conditions, traffic flow, or environmental factors associated with specific routes.

The count of vehicles exceeding 100 km/h (spd_100_and_above) has a positive and significant association with the median speed, with an estimated coefficient of 0.09 . This indicates that for every additional vehicle exceeding 100 km/h the median speed increases by 0.09 km/h, highlighting the influence of high-speed outliers on overall traffic behavior.

The model explains approximately 12.9% of the variance in median speeds ($R^2 = 0.129$), with an adjusted R^2 of 12.9%, indicating a modest but meaningful level of explanatory power. The root mean square error (RMSE) of 9.45 km/h suggests that the model's predictions deviate from the observed values by an average of about 9.45 km/h, reflecting some residual variability not captured by the predictors.

5 Discussion

5.1 Median speed and traffic patterns

The analysis highlights the substantial influence of both traffic direction and the prevalence of high-speed vehicles on median driving speeds at locations equipped with speed display signs. Notably, Northbound and Southbound traffic demonstrate higher median speeds compared to Eastbound traffic, which serves as the reference group. These directional differences likely reflect a combination of structural and environmental factors, such as variations in road design, traffic management systems, or vehicle flow dynamics. For instance, Northbound and Southbound routes may have fewer stoplights, intersections, or other impediments, facilitating smoother and faster travel. Additionally, localized traffic dynamics, including differences in commuter patterns, terrain, or surrounding land use, may further contribute to these observed variations. These findings underscore the importance of accounting for directional traffic characteristics when analyzing driving behavior and developing targeted traffic interventions.

Moreover, the analysis identifies a strong relationship between the number of vehicles exceeding 100 km/h and the overall median speed at a location. Areas with a higher concentration of speeding vehicles tend to exhibit elevated median speeds, indicating that high-speed driving, even by a minority of vehicles, can significantly influence overall traffic flow. This pattern suggests that excessive speeding alters the typical speed profile, with broader implications for traffic safety and management. Together, these results illuminate the interaction between high-speed driving behaviors and directional traffic patterns, offering a more comprehensive perspective on median speed dynamics and their determinants across varying roadway contexts.

5.2 Impact of high-speed drivers on median speed

The pronounced association between high-speed driving behavior and median speed underscores the disproportionate influence of a relatively small number of speeding vehicles on overall traffic dynamics. Each additional vehicle exceeding 100 km/h contributes to a measurable increase in the median speed, illustrating how high-speed outliers can distort the central tendency of traffic flow. This cumulative effect becomes particularly pronounced in areas with frequent speeding incidents, where such behavior not only elevates median speeds but also exacerbates risks for all road users. These findings emphasize the broader implications of high-speed driving, reinforcing the importance of addressing excessive speeding to enhance road safety.

Table 1: Explanatory models of median speed based on direction of travel and count of vehicles over 100 km/h

	Bayesian model
(Intercept)	38.77
	(0.29)
directionNB	0.93
	(0.41)
directionSB	1.12
	(0.41)
directionWB	−0.33
	(0.42)
spd_100_and_above	0.09
	(0.00)
Num.Obs.	4262
R2	0.129
R2 Adj.	0.126
Log.Lik.	−15 619.745
ELPD	−15 627.2
ELPD s.e.	61.1
LOOIC	31 254.3
LOOIC s.e.	122.3
WAIC	31 254.3
RMSE	9.45

From a policy standpoint, these results suggest that curbing excessive speeding could yield significant benefits for traffic safety and speed compliance. Interventions such as speed display signs, enhanced enforcement efforts, and physical traffic calming measures are well-suited to mitigate high-speed driving behaviors. Complementary strategies, such as public awareness campaigns highlighting the dangers of excessive speeding, can further bolster these measures by fostering a culture of safer driving practices. By focusing on the root causes of high-speed behavior, traffic management programs can not only reduce median speeds but also minimize the likelihood of speed-related collisions, creating safer roadways for all users.

5.3 Relevance to traffic safety and policy

The findings of this study offer valuable insights for policymakers, transportation planners, and law enforcement agencies aiming to enhance road safety through data-driven interventions. The observed differences in median speeds across traffic directions highlight the necessity of tailoring interventions to the specific characteristics of each direction. For instance, Northbound and Southbound roads, which exhibit higher median speeds, may benefit from targeted measures such as additional signage, enhanced enforcement, or modifications to road infrastructure, such as lane narrowing or the installation of speed humps. These directional insights equip stakeholders with actionable strategies for improving compliance with speed limits and optimizing traffic management practices.

Furthermore, the strong association between the prevalence of high-speed vehicles and elevated median speeds underscores the critical importance of addressing excessive speeding in high-risk areas. Prioritizing enforcement and resources in locations where speeding is most frequent allows policymakers to achieve the greatest impact on reducing speeds and enhancing compliance. This approach also recognizes that high-speed driving is not merely an individual behavior but a systemic issue that influences overall traffic dynamics and safety outcomes. By leveraging these findings, stakeholders can adopt more targeted, efficient, and effective traffic safety interventions, contributing to broader efforts to create safer road environments for all users.

5.4 Weaknesses and next steps

While this analysis offers important insights into traffic speed patterns, several limitations should be acknowledged. A key limitation lies in the reliance on summary statistics for speed display sign installations, which excludes granular details such as hourly traffic variations, total traffic volume, and vehicle classifications. The absence of this information limits the ability to contextualize high-speed driving behavior within the broader traffic population, potentially overlooking critical patterns. For example, temporal factors, such as peak-hour versus off-peak traffic, could unveil dynamics that remain hidden in the aggregated data used in this study.

Additionally, the use of traffic direction as a categorical predictor, while effective for capturing broad trends, may oversimplify the complexities underlying directional differences. Factors such as road geometry, enforcement intensity, or surrounding land use could provide further clarity on why certain directions exhibit higher median speeds. Incorporating these variables in future analyses would enhance the ability to account for variations in traffic behavior. Expanding the dataset to include more detailed data, such as hourly speed distributions, total vehicle counts, or weather conditions, could enable a richer and more nuanced understanding of the factors influencing speed patterns.

Future research should consider exploring the interactions between high-speed driving and environmental factors, such as road conditions or visibility, to better understand their combined effects. Integrating qualitative methods, such as surveys capturing driver attitudes and perceptions, could complement quantitative analyses, providing a more holistic view of speeding behavior. Addressing these limitations and broadening the scope of inquiry will not only refine current findings but also support the development of more targeted, evidence-based traffic safety interventions.

Appendix

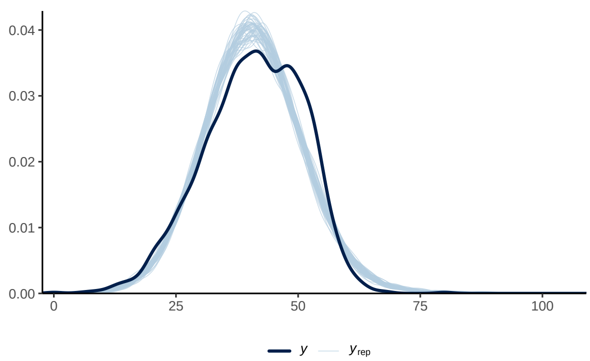
A Additional data details

B Model details

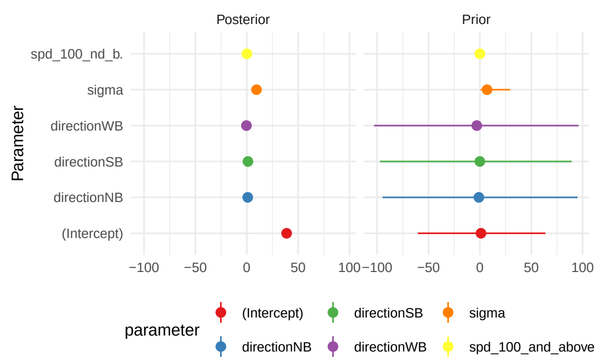
B.1 Posterior predictive check

In Figure 3a, we implement a posterior predictive check to evaluate how well the model reproduces the observed data. The observed median speed distribution is shown as a thick, dark line, while the lighter lines represent simulated predictions from the posterior distribution. The alignment between the observed and predicted distributions indicates that the model adequately captures the central tendencies and variability of the data, though some deviations in tails might suggest areas for improvement.

In Figure 3b, we compare the posterior distributions (right panel) with the prior distributions (left panel) for the model parameters. The posterior distributions for the directional effects and the count of vehicles exceeding 100 km/h are narrower and shifted compared to the priors, demonstrating that the data strongly informs these parameters. The intercept and residual variance (σ) are similarly influenced by the data, indicating meaningful updates from the prior assumptions to reflect observed patterns. This comparison highlights the balance between prior information and data-driven inference in the Bayesian framework.



(a) Posterior prediction check compares the observed data with predictions simulated from the model to evaluate the adequacy of fit.



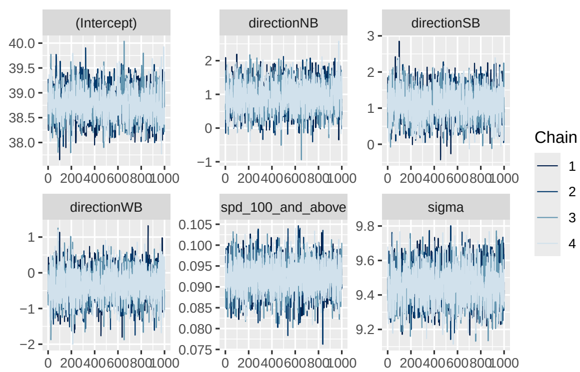
(b) Comparison of posterior distributions with prior distributions illustrates how the data informs the parameter estimates.

Figure 3: Examining how the model fits, and is affected by, the data

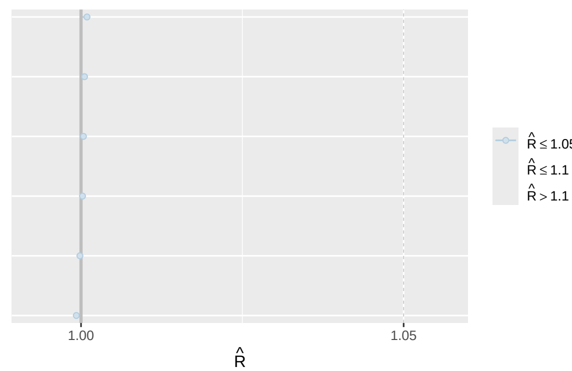
B.2 Diagnostics

Figure 4a is a trace plot. It shows the sampling chains for each parameter in the model over 1,000 iterations. Each parameter (e.g., intercept, direction, count of vehicles over 100 km/h, and residual variance) has multiple chains visualized, with different shades representing the chains. The overlapping and consistent patterns across chains suggest that the Markov Chain Monte Carlo (MCMC) algorithm has mixed well and reached convergence. The lack of significant drift or divergence indicates reliable and stable parameter estimates.

Figure 4b is an Rhat plot. It shows the Gelman-Rubin diagnostic for each parameter, which compares the variance within and between chains. All Rhat values are close to 1, confirming that the chains have converged and that the model's posterior distribution is well-sampled. This suggests that the MCMC algorithm has performed adequately and the parameter estimates are robust and reliable.



(a) Trace plot showing consistent sampling across multiple chains for all model parameters, suggesting convergence.



(b) Rhat plot indicating that all parameters have Rhat values close to 1, confirming proper mixing and convergence of the chains.

Figure 4: Checking the convergence of the MCMC algorithm

C Survey overview and methodology

Introduction

Accurate traffic speed monitoring is crucial for enhancing road safety, informing policy decisions, and optimizing urban planning. The City of Toronto's Mobile Watch Your Speed Program (Mobile WYSP) employs speed display signs equipped with radar devices and LED displays to monitor and influence vehicular speeds. This exploration delves into the sampling methodology of the Mobile WYSP, examining how observational data is collected, the strengths and limitations of the current approach, and potential enhancements to improve data reliability and representativeness.

Observational Data Collection in Mobile WYSP

The Mobile WYSP utilizes speed display signs strategically placed against existing hydro poles or streetlights. These units operate by:

- Speed measurement: Radar devices continuously monitor the speeds of oncoming vehicles.
- Real-time feedback: The measured speeds are displayed on LED signs to alert drivers, encouraging compliance with speed limits.
- Deployment strategy: Portable units are rotated monthly across different wards throughout the year.

This rotational deployment is the cornerstone of the program's sampling strategy, aiming to capture a diverse set of speed data across various geographical and temporal contexts within the city.

Sampling Design and Its Implications

1. Spatial Sampling:

- Coverage across wards: By rotating the speed display signs monthly across different wards, the program ensures that multiple areas of the city are monitored over time. This spatial distribution helps in capturing variations in traffic patterns, speed behaviors, and the effectiveness of speed interventions across diverse neighborhoods.
- Selection of installation sites: The choice of installation sites against hydro poles or streetlights is pragmatic, leveraging existing infrastructure for ease of deployment and maintenance. However, this might limit placements to areas with such infrastructure, potentially overlooking regions where these poles are scarce or absent.

2. Temporal Sampling:

- Monthly rotation: The monthly rotation allows the program to gather data across all wards within a year, accounting for seasonal variations in traffic and driving behaviors. This temporal breadth aids in understanding long-term trends and the impact of external factors (e.g., weather conditions, public events) on vehicle speeds.
- Operational parameters: Each installation's operational period is finite, meaning data is collected only during the active months. This cyclical approach might miss transient or short-term speed fluctuations that occur between rotation periods.

3. Sample Size and Frequency:

- Number of vehicles counted: While the program collects data on vehicle speeds, it explicitly states that the count of vehicles is not equivalent to a traffic volume count. This indicates that while speed data is robust, the program may not comprehensively capture traffic density, potentially limiting analyses that require both speed and volume metrics.
- Speed accuracy: With a stated accuracy of ± 1 km/hr, the radar devices provide reasonably precise speed measurements. However, this margin of error must be considered when interpreting data, especially in contexts where small speed differences are significant for safety assessments.

Strengths of the Current Sampling Methodology

- Comprehensive geographic coverage: Monthly rotation ensures that all wards are periodically monitored, providing a broad spatial dataset that reflects diverse traffic conditions across the city.
- Real-time feedback mechanism: The immediate display of vehicle speeds serves both as a data collection tool and a behavioral intervention, potentially enhancing the program's impact on reducing speeding incidents.
- Adaptability and flexibility: Portable units allow for dynamic deployment, enabling the program to respond to emerging traffic issues or shifts in urban development patterns.
- Longitudinal data collection: Over time, the repeated sampling across wards facilitates the analysis of trends, seasonal variations, and the long-term effectiveness of speed monitoring interventions.

Limitations and Challenges

- Non-Random sampling bias: The placement of speed signs on existing hydro poles or streetlights may introduce spatial bias, as certain areas might have higher densities of such infrastructure, leading to overrepresentation of specific regions and underrepresentation of others.
- Temporal gaps between rotations: Monthly rotations may not capture short-term speed variations or transient traffic conditions, potentially overlooking critical data points that occur between active monitoring periods.
- Data completeness and volume: Since the vehicle count does not equate to traffic volume, analyses requiring both speed and volume data might face limitations. This gap restricts the ability to perform comprehensive traffic flow studies or correlate speed data with congestion levels.
- Potential for data redundancy: Repeated measurements in the same wards over time could lead to data redundancy if not managed properly, affecting storage and processing efficiency.

- Accuracy limitations: The ± 1 km/hr speed accuracy, while generally acceptable, may pose challenges in high-precision studies, especially in urban environments where speed thresholds are critical for safety.

Recommendations for Enhancing Sampling Methodology

- Randomized site selection: Incorporating random sampling techniques in the selection of installation sites within each ward can mitigate spatial biases, ensuring a more representative data collection across diverse urban contexts.
- Increasing temporal frequency: Enhancing the frequency of data collection within each ward, possibly through shorter rotation cycles or deploying multiple units simultaneously, can capture more granular temporal variations and transient traffic conditions.
- Integrating traffic volume data: Combining speed data with traffic volume metrics, either through additional sensors or complementary data sources, can provide a more holistic understanding of traffic dynamics and enhance the program's analytical capabilities.
- Leveraging data from permanent installations: Utilizing data from permanent speed monitoring installations, where available, can supplement the Mobile WYSP's data, filling gaps and providing continuity in long-term traffic monitoring.
- Advanced data validation techniques: Implementing robust data validation and calibration processes can improve speed measurement accuracy, reducing the margin of error and enhancing the reliability of the collected data.
- Engaging with the open data community: Collaborating with developers, policymakers, and civic advocates through the Open Data Portal can facilitate the creation of innovative tools and visualizations, improving data accessibility and utility for diverse stakeholders.

Idealized Survey Design and Budget Allocation

To enhance the Mobile Watch Your Speed Program's data representativeness and reliability, an idealized survey design would incorporate a stratified random sampling approach, targeting a balanced representation of various road types, traffic conditions, and demographic areas within Toronto. The budget allocation for this initiative would prioritize the procurement of additional portable speed display units to increase the frequency of data collection. Assuming the acquisition cost of each unit is approximately $10,000, an initial investment of $300,000 could fund 30 additional units, significantly expanding the program's coverage and enabling simultaneous deployments in multiple wards. Operational costs, including maintenance, calibration, and redeployment, are estimated at $500 per unit per month, totaling $180,000 annually for all units.

The recruitment process for personnel would involve hiring a dedicated team of six field technicians, each assigned to manage the deployment, maintenance, and data retrieval of units across designated zones. With an estimated annual salary of $60,000 per technician, personnel costs would amount to $360,000 per year. Additionally, a centralized data management team of three analysts, each earning $70,000 annually, would be responsible for data processing, validation, and reporting, adding another $210,000 to the budget. These analysts would utilize advanced statistical tools and geospatial software to analyze the data, identify trends, and generate actionable insights for traffic policy recommendations.

Sampling Cost and Data Collection Protocols

The stratified sampling design would focus on selecting survey locations across all wards, categorized by road type (residential, arterial, and highway), traffic density, and socioeconomic demographics. Each selected location would be monitored for a minimum of two weeks, capturing data during peak and off-peak hours to ensure comprehensive temporal coverage. Deploying 30 units simultaneously across three road types in 10 wards would yield a sample size sufficient to detect significant differences in traffic speed patterns with high statistical power. To complement these efforts, temporary counters could be installed alongside speed display units to measure traffic volume, adding an estimated $20,000 annually for rental and deployment.

Data collection protocols would involve automated downloads from speed display units via wireless networks to minimize downtime. Each unit would undergo monthly calibration checks to ensure measurement accuracy. Training sessions for field technicians and analysts would be conducted quarterly, with a one-time initial training budget of $15,000 and annual refreshers costing $5,000. An estimated $50,000 would be allocated annually for miscellaneous expenses, including software licenses, transportation costs, and contingency funds. Altogether, the program would require a projected annual budget of approximately $1.2 million, aligning resources to achieve a comprehensive, high-quality traffic monitoring system that supports data-driven urban planning and road safety initiatives.

Conclusion

The Mobile Watch Your Speed Program's sampling methodology, characterized by spatial rotation and observational speed data collection, provides valuable insights into urban traffic behaviors across Toronto's wards. While the approach offers comprehensive geographic and temporal coverage, addressing inherent limitations through randomized sampling, increased data frequency, and integration with additional data sources can significantly enhance the program's effectiveness. As the City of Toronto continues to evolve its Open Data initiatives, refining the Mobile WYSP's sampling strategies will be pivotal in leveraging data-driven solutions for safer and more efficient urban mobility.

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.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Fildes, BN, G Rumbold, and A Leening. 1991. "Speed Behaviour and Drivers' Attitude to Speeding." *Monash University Accident Research Centre, Report* 16 (186): 104–15.

Firke, Sam. 2023. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

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

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/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Toronto, City of. 2024. "Mobile Watch Your Speed Program – Speed Summary." https://open.toronto.ca/dataset/mobile-watch-your-speed-program-speed-summary/.

Wickham, Hadley. 2011. "testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

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.