Chasing Speed Demons: Analyzing Toronto's Speeding Hotspots and Enforcement Gaps"*

My subtitle if needed

Karen Riani

December 4, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

---

*Code and data are available at: https://github.com/karenrni/Analyzing-Torontos-Speeders/.

1 Introduction

Toronto's battle against speeding takes some wild turns—literally and figuratively. A speed camera notorious for issuing tens of thousands of tickets on Parkside Drive, Toronto's most ticketed street for speeding, was cut down less than 24 hours after being reinstalled (Toronto 2023; CBC News 2023). While the city struggles to curb speeding, new research by the Canadian Automobile Association (CAA) paints a troubling picture: nearly 70% of Canadians admit to speeding in residential areas, 50% routinely speed on highways, and 20% regularly push well beyond the limit (Jobber Nation 2023).

The rise of advanced vehicle technology and the adrenaline appeal of street racing exacerbate these issues, with young male drivers (18–25 years) disproportionately represented in this trend. Though rare at **0.9% prevalence**, street racing poses a **five-fold increase in crash risk (OR = 5.23, p < .001)**, even after accounting for other risk factors (Simons-Morton et al. 2017). Despite Ontario's strict penalties, such as **$10,000 fines** and **30-day license suspensions for excessive speeding**, enforcement alone has not fully addressed these challenges (Traffic Paralegal Services 2024). Recent headlines, such as a driver clocked at **102 km/h in a 50 km/h school zone** and a Toronto speed camera being destroyed less than 24 hours after reinstallation, underscore the disconnect between enforcement measures and public behavior (CityNews Staff 2024; CBC News 2023).

This paper focuses on understanding the estimand **over_speed_limit**, the amount by which speeds in each observed speed bin exceed the posted speed limit. This dependent variable is calculated as the difference between the lower bound of the observed speed bin and the speed limit, ensuring a straightforward measure of speeding severity. By quantifying how far drivers exceed the limit, over_speed_limit allows for identifying patterns of high-risk speeding behaviour. Using open data from Toronto's School Safety Zone Watch Your Speed Program, Watch Your Speed Detailed Counts, and Traffic Camera datasets, including vehicle counts (volume) in each speed bin, speed limits, and ward-level identifiers, this study investigates how traffic patterns, enforcement mechanisms, and geographic factors contribute to variations in speeding behaviour across the city.

To address this, the study employs multi-linear regression models to examine the relationship between excess speed and predictors such as traffic volume, speed limits, and ward characteristics throughout 2023. Results indicate that higher traffic volume reduces excess speed by **0.12 km/h per additional vehicle**, while speed limits positively influence speeding by **0.13 km/h per unit increase**. Wards exhibit substantial variability, with some, such as Ward 4, showing significant decreases in speeding. However, the impact of individual speed cameras was negligible, suggesting limitations in the data and methodology for evaluating camera-specific effects.

The rest of this paper is structured as follows: Section 2 presents the data sources and methodology, Section 3 presents the forecasting model, then Section 4 followed by Section 5, which

discusses the results and their implications. Section 6 concludes with a discussion of the limitations of the study and suggestions for future research, followed by Appendix Section A with a budgeted idealized survey and methodology.

2 Data

2.1 Overview

The data for this analysis is sourced from Open Data Toronto (Gelfand 2022) and comprises three datasets: School Safety Zone Watch Your Speed Program – Locations (Dumas 2024b), School Safety Zone Watch Your Speed Program – Detailed Speed Counts in 2023 (Dumas 2024a), and the Traffic Camera dataset (Toronto Open Data 2024). These datasets provide detailed insights into speeding behavior, traffic enforcement mechanisms, and geographic characteristics within Toronto. The traffic camera and speed sign data were merged through a spacial join with a range of 500m using geographical coordinates. This was joined with the detailed counts of speeds through Toronto ward number (25 ward structure).

2.2 Packages Used

The analysis utilized R (R Core Team 2024) with the following packages: `opendatatoronto` (Gelfand 2022), `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `arrow` (Richardson et al. 2024), `readr` (Wickham, Hester, and François 2023), `sf` (Pebesma 2023), `testthat` (Wickham and Hester 2023), `ggplot2` (Wickham 2016), `corrplot` (Wei and Simko 2021), `tidymodels` (Kuhn et al. 2023), `kableExtra` (Zhu 2021) and `rstanarm` (Goodrich, Gabry, et al. 2023).

2.3 Measurement

The datasets from Open Data Toronto serve as the foundation for analyzing speeding behavior in Toronto. Each variable is derived from real-world phenomena and represents a specific aspect of traffic patterns or enforcement mechanisms:

- **Over Speed Limit**: This key dependent variable quantifies the extent to which drivers exceed posted speed limits. It is derived by calculating the difference between the midpoint of the observed speed bin (e.g., `[45,50)` translates to a midpoint of 47.5 km/h) and the posted speed limit. This method provides a nuanced measure of speeding severity, distinguishing between minor infractions and extreme violations.

4

- **Sign ID and Camera ID (X_id)**: These unique identifiers represent speed cameras and their locations, facilitating the mapping of speeding patterns to enforcement points. However, sign addresses may not always align precisely with camera placements, as discussed in the limitations section.

- **Longitude and Latitude**: Geographic coordinates capturing the locations of observed speeding behaviour, allowing for spatial analysis and the identification of high-risk zones.

- **Speed Limit**: The posted speed limits for each location, ranging from 30 km/h to 50 km/h, with an average of 38.24 km/h. This variable provides the baseline against which speeding is measured.

- **Volume**: The count of vehicles observed in a given speed bin, representing traffic density. Median volume is 13 vehicles, with counts ranging widely, up to a maximum of 209 vehicles.

- **Speed Bin**: Recorded ranges of observed speeds, such as `[45,50)`, provide context for speed categories, assuming uniform speed distributions within each bin.

- **Ward Number**: Geographic zones represented as categorical identifiers, reflecting administrative boundaries within Toronto. For instance, Ward 10 accounts for 28.8% of observations, making it the most represented ward, while Ward 19 has considerably fewer observations.

- **No Camera in Radius**: This binary variable was constructed based on spatial joins between speed observation data and camera locations. If a given location fell outside the defined radius of nearby cameras, it was flagged as `TRUE`. This captures the absence of enforcement presence and allows for an analysis of its influence on speeding behavior.

- **Camera Density**: To represent enforcement intensity across wards, the number of cameras was aggregated per ward. This variable offers insight into how concentrated enforcement might correlate with traffic patterns or speeding violations.

2.4 Data Limitations

The datasets used in this study have notable limitations that must be acknowledged. **Sign addresses** in the datasets have not been verified and may represent the address of an adjacent property rather than the actual location of the sign. The **speed limit** column reflects the most recent operating parameter, which does not necessarily align with the current or previous speed limits for each location.

The **Traffic Cameras** dataset is highly up-to-date, with a last refresh on 2024-11-26. While there are no issues with accessibility or completeness, usability is constrained by unclear column names and incomplete metadata. Conversely, the **School Safety Zone Watch Your Speed Program** – **Detailed Speed Counts** dataset is over 12 months outdated, requiring reliance

Table 1

Table 2: Summary Statistics for the Final Filtered Data

Variable	Min	X1st.Qu.	Median	Mean	X3rd.Qu.	Max
sign_id	23.00	667.00	950.00	1105.00	978.00	2622.00
X_id	11.00	146.00	298.00	247.10	362.00	423.00
longitude	-79.58	-79.40	-79.36	-79.36	-79.33	-79.19
latitude	43.64	43.65	43.65	43.68	43.71	43.79
speed_limit	30.00	40.00	40.00	38.24	40.00	50.00
volume	1.00	4.00	13.00	23.35	35.00	209.00
speed_bin_lower	5.00	15.00	30.00	30.23	40.00	100.00
over_speed_limit	0.00	0.00	0.00	3.91	5.00	70.00

Summary of Joined Analysis Data

on direct CSV links for access. This reduces its applicability for analysing current traffic trends.

To address gaps, **spatial joins** were employed to map camera locations to speed observation points, identifying enforcement presence, and **calculated variables**, such as over speed limit and camera density, were derived to facilitate analysis. However, these transformations have limitations. Speed bin ranges assume a uniform speed distribution, which may not capture the variability in driver behaviours. Furthermore, camera placements are concentrated in high-traffic or high-risk areas, potentially underrepresenting quieter zones. Together, these factors highlight the need for caution when interpreting results, as the datasets provide a snapshot rather than a comprehensive depiction of traffic behaviour.

2.5 Descriptive Statistics

Descriptive statistics in Table 1 highlight critical patterns within the data:

- The **mean over speed limit** is 3.91 km/h, with extreme violations reaching up to 70 km/h.

- The **median speed bin midpoint** is 30 km/h, while the mean traffic volume is 23.35 vehicles.

- Geographic variables such as **longitude** and **latitude** center around Toronto's urban core.
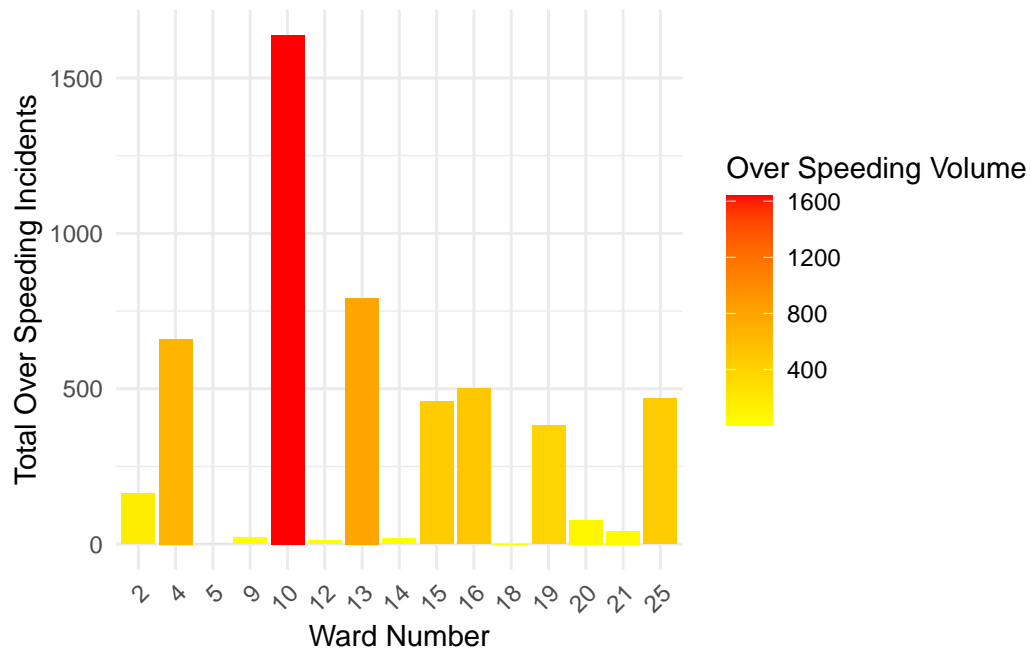
`geom_smooth()` using formula = 'y ~ x'

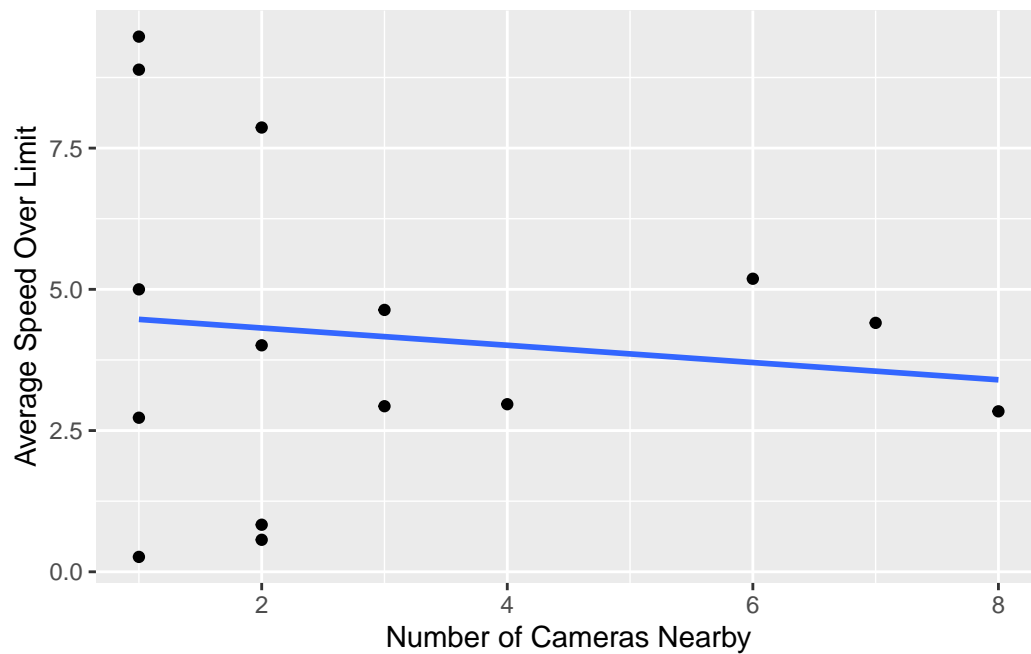Figure 1: Number of Vehicles Exceeding Speed Limits by Ward

Figure 2: Camera Density vs. Speeding Behavior

7

```
`summarise()` has grouped output by 'ward_no'. You can override using the
`.groups` argument.
```
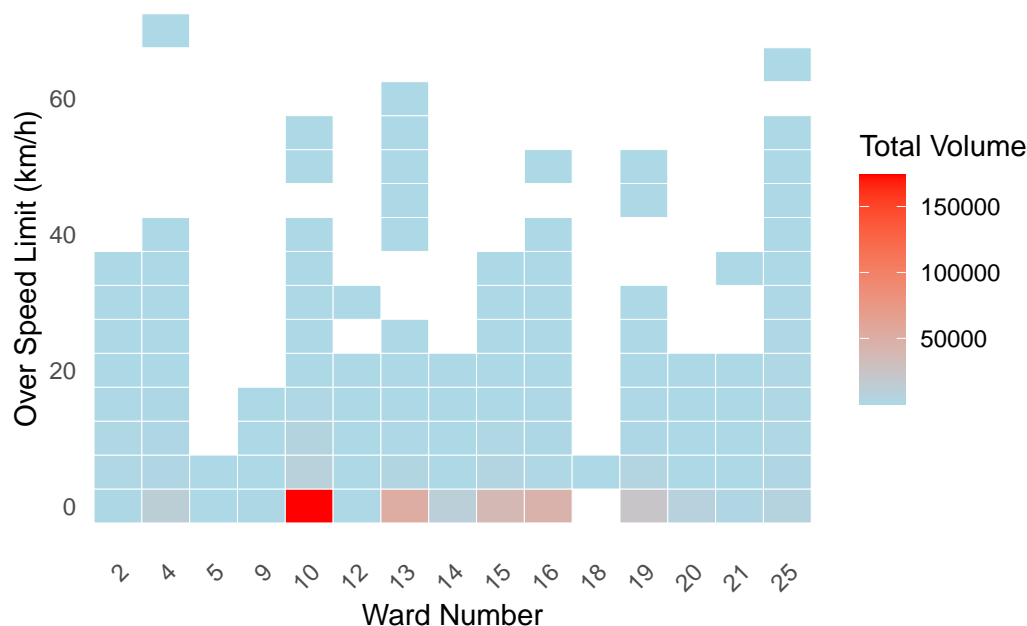

Figure 3: Speeding Volume by Ward and Over Speed Limit

3 Model

3.1 Overview

This section outlines the modeling process used to analyze over-speeding behavior based on data from Toronto's School Safety Zone Watch Your Speed Program and Traffic Cameras datasets. The goal was to predict the extent to which observed speeds exceeded posted limits (over_speed_limit) by evaluating various predictors. Three models were developed and assessed using performance metrics, including RMSE, MAE, and (R^2), to determine the most appropriate model. The final model selection was based on its balance between simplicity, accuracy, and interpretability.

3.2 Model Structure

Common Variables Across All Models:

- Speed Limit (x): Posted speed limit at a given location.

- Volume (x): Number of vehicles observed in a speed bin.

Model-Specific Variables:

- Model 1: Includes ward-level categorical effects (ward_no) to account for geographic differences in speeding behavior.

- Model 2: Omits ward-level effects to focus solely on speed limits and vehicle volumes.

- Model 3: Adds a camera identifier (X d) to examine the influence of specific camera locations on speeding patterns.

3.2.0.1 Model Formulations

1. Model 1: Full Model with Ward-Level Effects [y = _0 + *1 x_1 +* 2 x_2 + *{j=1}^{13}* {3j} ward_no_j +]

2. Model 2: Excluding Ward-Level Effects [y = _0 + _1 x_1 + _2 x_2 +]

3. Model 3: Including Camera Identifiers [y = _0 + *1 x_1 +* 2 x_2 + *{j=1}^{13}* {3j} ward_no_j + *4 X*{id} +]

3.3 Variable Selection and Final Model

The selection process aimed to balance complexity and explanatory power. Ward-level effects were included to capture geographic variations, while Camera Identifiers were tested to evaluate their influence. Redundant variables were excluded to maintain simplicity. Model 1 was selected as the final model because it demonstrated superior performance across key metrics. By including ward-level effects, it effectively captured geographic differences in speeding behavior, making it the most suitable choice.

3.4 Model Validation and Diagnostics

3.4.1 Performance Metrics

The RMSE for Model 1 was 7.01, indicating that, on average, predictions deviated from observed values by this amount. MAE was 4.77, highlighting its precision in predicting over-speeding behavior. The (R^2) value of 0.176 showed that the model explained 17.6% of the variance in over-speeding behavior, providing a reasonable level of explanatory power for the observed data.

3.4.2 Residual Diagnostics

Residual diagnostics confirmed the validity of Model 1. The residual standard error of 7.37 aligned closely with the RMSE, reinforcing the model's accuracy. Residual plots exhibited no systematic patterns, supporting the assumption of linearity. Additionally, homoscedasticity was verified, indicating consistent variance in residuals across predictions.

3.5 Diagnostic Checks and Model Justification

Variance Inflation Factors (VIFs) were calculated to evaluate multicollinearity, with all variables in Model 1 displaying acceptable values below 4. Although Model 3 included Camera Identifiers, this variable was statistically insignificant and did not enhance model performance. Model 1 effectively balanced complexity and accuracy, integrating ward-level effects to provide valuable insights into over-speeding behavior while maintaining interpretability. Consequently, Model 1 was selected as the final model.

4 Results

5 Discussion

5.1 First discussion point

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

5.2 Second discussion point

5.3 Third discussion point

6 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

If you're incorporating feedback to diversify question types and address previous gaps, here's an improved version of your appendix for Toronto's speeding analysis with a reasonable budget. I'll also ensure it reflects insights from your earlier examples and feedback while incorporating poll aggregation and methodology details.

A Idealized Survey & Methodology - $500K Budget

A.1 Overview

This appendix outlines a survey methodology designed to enrich the analysis of excessive speeding near Toronto school zones. With a $500K budget, this approach aims to gather nuanced data from drivers, pedestrians, and community members about speeding behaviors, awareness of enforcement measures, and perceived safety. By integrating observational data with survey insights, this methodology enables robust analysis and informed policy recommendations.

A.2 Budget Justification

B Overview

A $500,000 budget is proposed for analyzing speeding trends in Toronto school zones. Based on similar city-wide traffic safety initiatives, such as Vision Zero and speed enforcement studies, this budget is allocated to account for:

Comprehensive sampling: Targeting diverse demographics and road users.

Advanced recruitment and incentives: Ensuring adequate participation.

High-quality data integration: Focused on cleaning, validation, and geospatial analysis.

C Proportionality

Vision Zero allocates $250,000 to $1,000,000 for focused safety measures annually, making $500,000 appropriate for a targeted study (City of Toronto 2024).

D Survey Costs

Recruiting diverse participants typically costs $200,000 to $300,000, ensuring robust and representative data (Pew Research Center 2023).

D.1 Sampling Approach

D.1.1 Target Groups

1. **Drivers:**

   - Key demographic: Ages 18-40, with emphasis on young male drivers.
   - Behavioral patterns: Frequent commuters in school zones.

2. **Pedestrians:**

   - Parents, school staff, and local residents.

3. **Community Advocates:**

   - Stakeholders involved in traffic safety initiatives.

D.1.2 Sampling Method

Multi-Stage Sampling: 1. **Stratification by Wards:**
- Ensure geographic coverage across all Toronto wards with school zones.
2. **Cluster Sampling:**
- Focus on areas near speed cameras and school safety zones.
3. **Random Sampling:**
- Within clusters, randomly select participants for unbiased representation.

Sample Size Goal:
- **3,000 respondents**: Sufficient to identify trends at the ward level while maintaining a margin of error below $\pm 2\%$.

D.2 Recruitment Strategy

D.2.1 Primary Channels

1. **Online Ads:**
   - Platforms: TikTok, Instagram, and YouTube.
   - Geo-targeted to school zones and high-traffic areas.
   - Budget Allocation: $150,000.

2. **On-the-Ground Recruitment:**
   - Flyers distributed in school zones, community centers, and traffic courts.
   - Budget Allocation: $50,000.

3. **Partnerships with Apps:**
   - Collaboration with navigation tools like Waze to deliver in-app survey prompts.
   - Budget Allocation: $100,000.

4. **Incentives:**
   - Monetary rewards: $20 gift cards for completion, with a chance to win larger prizes.
   - Budget Allocation: $100,000.

D.3 Survey Design

D.3.1 Diversified Question Types

To address the feedback, the survey incorporates **varied question formats**:

1. **Multiple Choice:**
   - Example: "What factors most influence your speed in school zones?"
     - Time pressure.
     - Peer behavior.
     - Lack of awareness of speed limits.
     - Others.

2. **Likert Scales:**
   - Example: "Rate your agreement with: 'Speed cameras make me feel safer in school zones.'"
     - Strongly disagree to strongly agree.

3. **Open-Ended:**

   - Example: "What other measures would help reduce speeding in school zones?"

4. **Ranking Questions:**

   - Example: "Rank the following in order of their effectiveness at reducing speeding: speed cameras, visible signage, increased fines, police presence."

5. **Interactive Scenario Questions:**

   - Example: "If you saw a speed camera sign but no visible camera, how likely are you to still speed? (Very unlikely to very likely)."

D.3.2 Survey Flow

1. **Introduction and Consent:**

   - Explain the purpose, time required (5-7 minutes), and confidentiality.

2. **Demographics:**

   - Age, gender, driving habits, and proximity to school zones.

3. **Behavioral Insights:**

   - Driving patterns, awareness of speed limits, and interaction with enforcement measures.

4. **Perceived Effectiveness:**

   - Opinions on speed cameras, signage, and enforcement presence.

5. **Safety Concerns:**

   - Perceptions of safety in school zones and suggested improvements.

D.4 Data Validation

1. **Logic Checks:**

   - Identify inconsistencies (e.g., reporting no awareness of cameras but frequent encounters).

2. **Cross-Validation:**

   - Compare survey responses with observational data (e.g., traffic camera footage and speed counts).

3. **Geospatial Matching:**

   - Integrate survey responses with geospatial data from speed signs and cameras for location-specific analysis.

D.5 Poll Aggregation and Modeling

D.5.1 Aggregation Approach

A **poll-of-polls method** balances survey insights with observational data: - Weight survey responses by: - **Proximity to enforcement measures.** - **Frequency of driving in school zones.** - Integrate responses with speed count data to build predictive models of speeding behavior.

D.5.2 Modeling Framework

- **Dependent Variable:** Likelihood of exceeding speed limits.

- **Predictors:**

   - Awareness of enforcement measures.

   - Demographics (age, gender).

   - Self-reported speeding behavior.

   - Proximity to cameras and speed signs.

D.5.3 Diagnostics

- Residual analysis and goodness-of-fit tests to validate model performance.
- Q-Q plots and distribution checks for robustness.

D.6 Budget Breakdown

Expense	Cost Estimate (CAD)
Social Media Recruitment	$150,000
In-App Ads and Recruitment	$100,000
On-the-Ground Recruitment	$50,000
Participant Incentives	$100,000
Data Cleaning and Validation	$75,000
Survey Design and Implementation	$25,000
Total	**$500,000**

D.7 Appendix Content

1. **Survey Link:**
   [Insert Google Forms Link Here].

2. **Copy of Survey Questions:**

   - Include all questions, organized by section.

3. **Model Diagnostics:**

   - Residual plots, fit comparisons, and distribution checks.

This improved version addresses the feedback, diversifies question types, and integrates poll aggregation and modeling. Let me know if you'd like refinements!

References

CBC News. 2023. "Toronto Street Most-Ticketed for Speeding in 2023 Branded 'Dangerous and Chaotic'." 2023. https://www.cbc.ca/news/canada/toronto/parkside-drive-dangerous-speeding-1.6851959.

City of Toronto. 2024. *Vision Zero Road Safety Plan 2024 Update.*

CityNews Staff. 2024. "Ontario Speeding Driver Charged in East Gwillimbury School Zone." 2024. https://toronto.citynews.ca/2024/09/16/ontario-speeding-driver-charged-east-gwillimbury-school-zone/.

Dumas, Raphael. 2024a. *School Safety Zone Watch Your Speed Program – Detailed Speed Counts.* https://open.toronto.ca/dataset/school-safety-zone-watch-your-speed-program-detailed-speed-counts/.

———. 2024b. *School Safety Zone Watch Your Speed Program – Locations.* https://open.toronto.ca/dataset/school-safety-zone-watch-your-speed-program-locations/.

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.

Goodrich, Ben, Jonah Gabry, et al. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm/.

Jobber Nation. 2023. "With Speeding on the Rise, What's Your Risk?" 2023. https://www.jobbernation.ca/with-speeding-on-the-rise-whats-your-risk/.

Kuhn, Max et al. 2023. *Tidymodels: A Collection of Packages for Modeling and Machine Learning.* https://www.tidymodels.org/.

Pebesma, Edzer. 2023. *Sf: Simple Features for r.* https://CRAN.R-project.org/package=sf.

Pew Research Center. 2023. *Survey Incentives and Response Rates.*

R Core Team. 2024. *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.

Simons-Morton, Bruce, Kyoung Cheon, Feng Guo, and Paul S. Albert. 2017. "Street Racing Among the Ontario Adult Population: Prevalence and Association with Collision Risk." *ResearchGate.* https://www.researchgate.net/publication/315845469_Street_racing_among_the_Ontario_adult_population_Prevalence_and_association_with_collision_risk.

Toronto, CTV News. 2023. "Disappointing: Toronto Speed Camera Cut down Less Than 24 Hours After Being Reinstalled." 2023. https://toronto.ctvnews.ca/disappointing-toronto-speed-camera-cut-down-less-than-24-hours-after-being-reinstalled-1.7129336.

Toronto Open Data, City of. 2024. *Traffic Cameras.* https://open.toronto.ca/dataset/traffic-cameras/.

Traffic Paralegal Services. 2024. *Ontario Speeding Tickets: What You Need to Know and How Traffic Paralegal Services Can Help.* https://www.trafficparalegalservices.com/speeding-tickets/.

Wei, Taiyun, and Viliam Simko. 2021. *Corrplot: Visualization of a Correlation Matrix.* https://github.com/taiyun/corrplot.

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, and Jim Hester. 2023. *Testthat: Unit Testing for r.* https://CRAN.R-project.org/package=testthat.

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

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.