Chasing Speeders: Analyzing Toronto's 2023 Speeding Trends in School Zones"*

Volume Versus Speed Limits: 0.12 km/h Declines That Add Up Across School-Traffic Zones

Karen Riani

December 4, 2024

This study investigates over-speeding in Toronto, defined as the amount by which vehicles exceed posted speed limits, using data from the city's School Safety Zone Watch Your Speed Program and Traffic Camera datasets. A multi-linear regression analysis revealed key drivers of speeding behavior: higher traffic volumes reduce over-speeding by 0.12 km/h per additional vehicle, while increasing speed limits correlate with a 0.13 km/h rise in excess speed per km/h limit increase. Geographic variability was significant, with Ward 4 showing a 4.8 km/h reduction in over-speeding compared to the reference ward. Despite these findings, the negligible impact of speed cameras highlights limitations in existing enforcement data. These insights underscore the importance of nuanced, data-driven approaches to traffic management and policy development in urban settings.

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 in 2023. 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), `modelsummary` (Arel-Bundock 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 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 measure of speeding severity, distinguishing between minor infractions and extreme violations.

- **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.

[fig-heatbar] visualizes the total volume of over-speeding incidents across various wards. Ward 10 stands out prominently with the highest number of over-speeding incidents, exceeding 1,500, which may indicate unique characteristics such as higher traffic density or insufficient

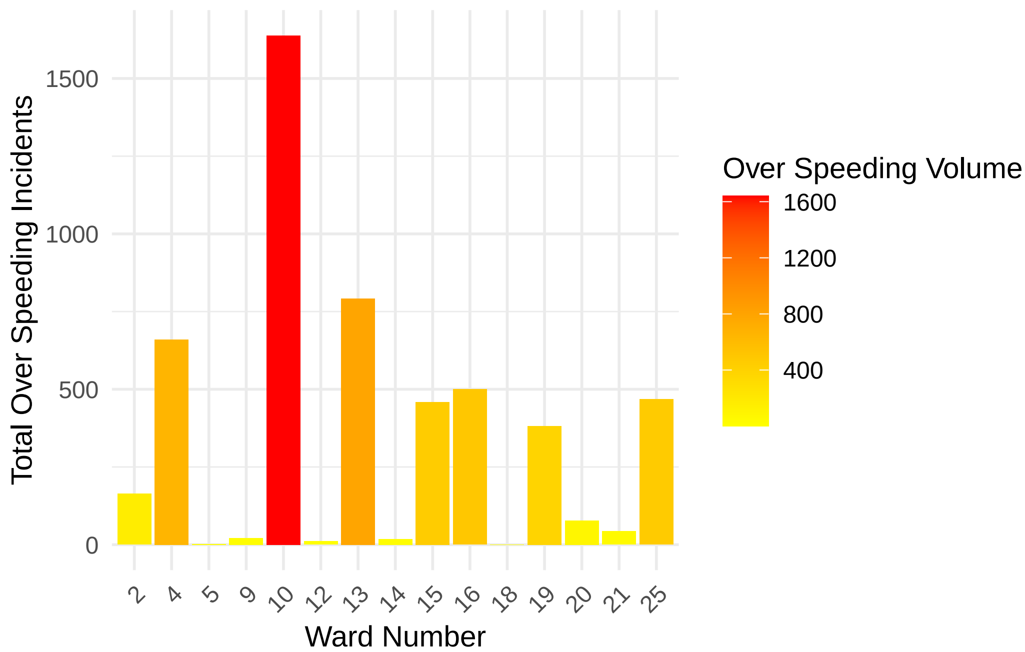

Figure 1: Number of Vehicles Exceeding Speed Limits by Ward

enforcement measures. Other wards, such as Wards 13 and 16, also show elevated levels of over-speeding but fall significantly behind Ward 10.

The scatter plot in [fig-cam-dens] explores the relationship between the number of speed cameras nearby and the average speed over the limit across various wards. The slight downward trend, indicated by the fitted regression line, suggests a weak negative association between the number of cameras and over-speeding behavior. Specifically, as the number of cameras increases, the average speed over the limit tends to decrease, albeit marginally. However, the variability in average speed over the limit across wards with the same number of cameras highlights the potential influence of other ward-specific factors, such as enforcement policies or road design.

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

Figure 3 provides a detailed visualization of speeding behavior across Toronto's wards, showcasing the relationship between the extent of over-speeding (measured in km/h above the limit) and the total volume of vehicles recorded. Ward 10 stands out with the highest total speeding volume, concentrated in lower over-speeding categories (<40 km/h above the limit), indicating significant non-compliance with posted limits. Conversely, instances of extreme over-speeding

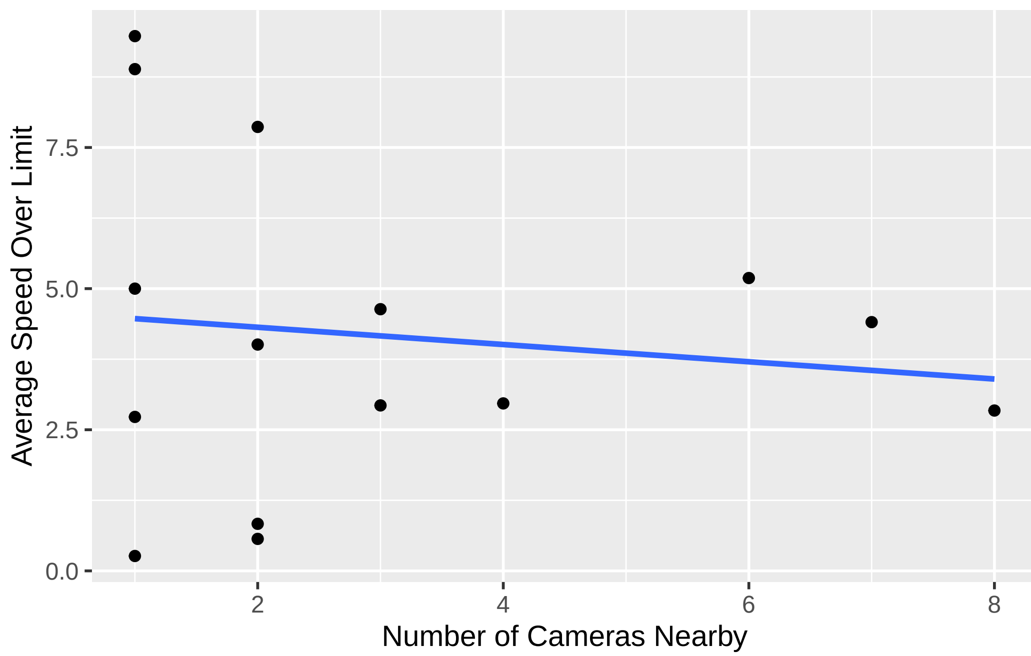

Figure 2: Camera Density vs. Speeding Behavior

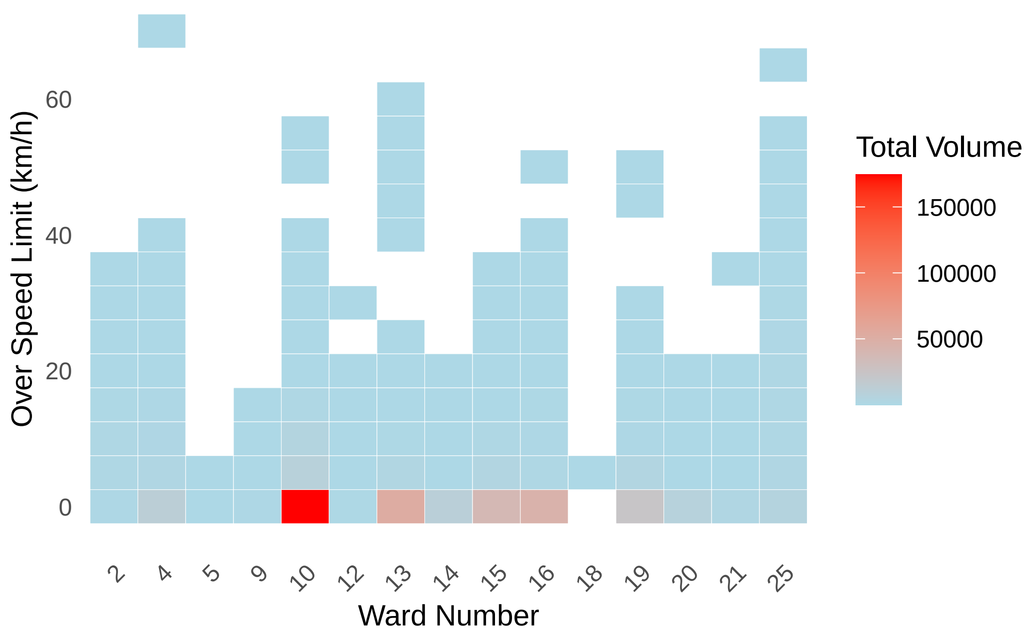

Figure 3: Speeding Volume by Ward and Over Speed Limit

(>50 km/h above the limit) are sparse but occur across wards 4, 13, and 25, highlighting potential hotspots for dangerous driving behavior.

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: Posted speed limit at a given location.
- Volume: 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_id) 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

$$\text{over\_speed\_limit}_i = \beta_0 + \beta_1 \times \text{speed\_limit}_i + \beta_2 \times \text{volume}_i + \sum_{j=1}^{13} \beta_{3j} \times \text{ward\_no}_{ij} + \epsilon_i \quad (1)$$

2. Model 2: Excluding Ward-Level Effects

$$\text{over\_speed\_limit}_i = \beta_0 + \beta_1 \times \text{speed\_limit}_i + \beta_2 \times \text{volume}_i + \epsilon_i \tag{2}$$

3. Model 3: Including Camera Identifiers (X_id)

$$\text{over\_speed\_limit}_i = \beta_0 + \beta_1 \times \text{speed\_limit}_i + \beta_2 \times \text{volume}_i + \sum_{j=1}^{13} \beta_{3j} \times \text{ward\_no}_{ij} + \beta_4 \times \text{X\_id}_i + \epsilon_i \tag{3}$$

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 validation metrics. By including ward-level effects, it effectively captured geographic differences in speeding behaviour, 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. Please see Figure 6 in the appendix for a detailed graph.

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

4.1 Model Summary Statistics

The analysis examined over-speeding, defined as the amount (in km/h) by which vehicles exceeded the posted speed limit. In Table 3, shows the variation in speeding effects for model 1, including the ward number, speed limit, and volume among other metrics.

4.1.0.1 Analysis of Over-Speeding Behavior

For every 1 km/h increase in the posted speed limit, the predicted over-speeding increased by 0.13 km/h, reflecting a slight inclination for drivers to exceed limits more on roads with higher speed thresholds. Traffic volume also played a role, with each additional vehicle in the observed volume leading to a decrease in over-speeding by 0.12 km/h. This suggests that heavier traffic may act as a natural deterrent to excessive speeding.

Ward-level differences were also observed, with Ward 2 serving as the reference category. Ward 4, known for its high-speed enforcement on Parkside Drive, showed a reduction of 4.84 km/h in over-speeding compared to Ward 2. Larger reductions were observed in Ward 20 (10.84 km/h) and Ward 14 (11.22 km/h). Conversely, Ward 12 displayed an increase of 4.15 km/h, indicating that over-speeding was more prevalent in this area.

4.1.0.2 Performance Statistics

The model's residual standard error was 7.36 km/h, indicating the average deviation of predictions from observed speeding levels. This was supported by an RMSE of 7.36. A relative RMSE of 10.51% of the observed range and 32.91% of the standard deviation indicates moderate predictive accuracy. While it suggests the model captures a portion of the variability, the error may be considered substantial given the dataset's complexity. For urban traffic behavior, this level of error isn't ideal but isn't unexpected due to high variability and unobserved factors influencing over-speeding.

The (R^2) value of 0.176 and adjusted (R^2) of 0.175 show that the model explains approximately 17.6% of the variance in over-speeding. Although relatively modest, this reflects the challenge of modeling complex urban traffic behaviors with diverse contributing factors.

4.1.0.3 Statistical Fit Metrics

With an AIC of 99,722.4 and a BIC of 99,851.4, the model, these values are consistent with the dataset's size, being 14,594 observations.The model's F-statistic of 207.459 shows the statistical significance of the predictors in explaining variations in over-speeding. This shows that the inclusion of speed limits, traffic volume, and ward-level effects contributes significantly to the model's explanatory power.

4.2 Relationship Between predicted and observed over-speeding values

```
knitr::include_graphics(here("other", "graphs", "predObsv.png"))
```

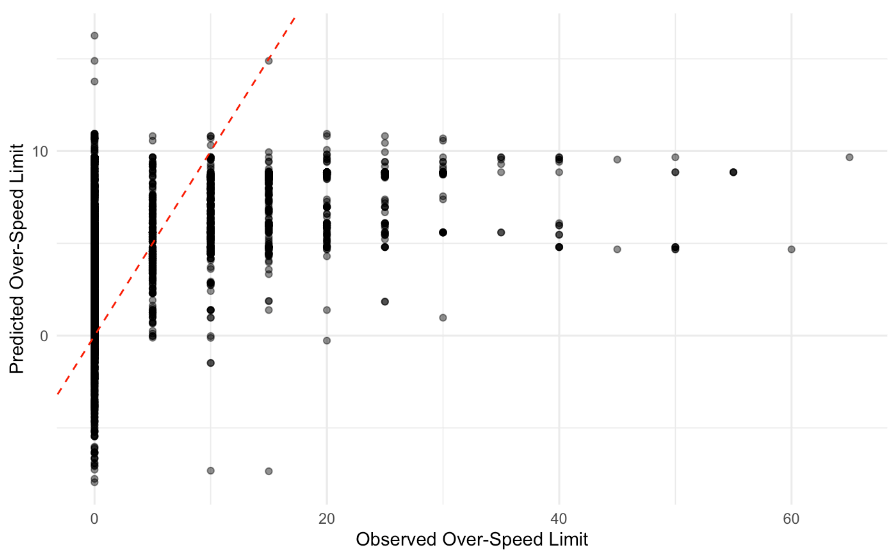

Figure 4: Observed vs. Predicted Over-Speeding: Model Accuracy Visualization

Figure 4 shows the relationship between predicted and observed over-speeding values. Predicted values align reasonably well with observed values at lower ranges (0–10 km/h). This indicates that the model performs adequately in capturing moderate over-speeding behaviors, where most of the data points fall close to the red dashed line representing perfect predictions.

Table 3: Summary of Regression Estimates for Model 1

	Model 1
(Intercept)	7.191
	(1.582)
speed_limit	0.129
	(0.050)
volume	−0.124
	(0.003)
ward_no4	−4.845
	(0.528)
ward_no5	−9.533
	(2.648)
ward_no9	−6.710
	(1.253)
ward_no10	−3.374
	(0.702)
ward_no12	4.152
	(2.087)
ward_no13	−7.430
	(0.708)
ward_no14	−11.222
	(1.164)
ward_no15	−2.134
	(0.535)
ward_no16	−6.638
	(0.715)
ward_no19	−5.263
	(0.731)
ward_no20	−10.844
	(0.741)
ward_no21	−9.098
	(1.280)
ward_no25	−1.274
	(0.540)

4.2.1 Underestimation of Higher Values

For observed over-speeding values exceeding 20 km/h, the model consistently underestimates. Predicted values do not reach the observed extremes, suggesting that the model has difficulty accounting for the variability or extremity of high over-speeding incidents. This limitation is evident from the lack of points in the higher range of the dashed line.

4.2.2 Clustering at Lower Speeds

There is a noticeable clustering of predictions around 0–10 km/h, even for observed over-speeding values as high as 40 km/h. This clustering suggests the model's predictions are conservative, failing to differentiate effectively between moderate and extreme over-speeding cases.

While the model provides reasonable predictions for average over-speeding behaviors, its performance diminishes at higher observed values, indicating potential limitations in its ability to capture outliers or extreme behaviors.

4.3 Average Predicted Over-Speeding by Ward

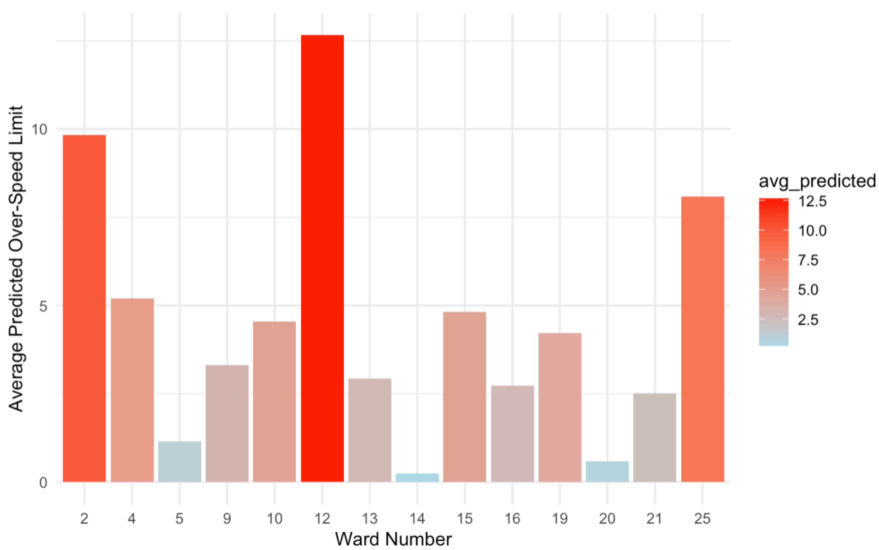

Figure 5: Predicted Over-Speed Limit by Ward

Figure 5 shows the average predicted over-speeding across Toronto wards, measured in kilometers per hour.

4.3.1 Distribution of Predictions Across Wards

Ward 12 exhibits the highest predicted over-speeding value, with an average of approximately 12.5 km/h over the speed limit. This value is notably higher compared to other wards, highlighting Ward 12 as a potential outlier. Similarly, Ward 25 follows closely with an average predicted over-speeding of about 11 km/h, suggesting consistently elevated predictions in these areas.

4.3.2 Wards with Lower Predictions

In contrast, Wards 5, 21, and 20 show substantially lower predicted over-speeding values, averaging below 3 km/h. These results suggest that predicted over-speeding behaviors are less prominent in these wards, potentially influenced by traffic conditions, enforcement presence, or other unobserved factors.

4.3.2.1 Implications of Predictions

While the predictions align with observed patterns in some wards, such as higher over-speeding in Ward 12, the variability across wards suggests that spatial characteristics may play a significant role. These predictions provide a baseline understanding of over-speeding behavior at a ward level.

5 Discussion

5.1 Ward-Level Disparities in Over-Speeding

The results show geographical disparities in over-speeding across Toronto. For instance, Ward 12 emerges as a statistical outlier, with an average predicted over-speeding of 12.5 km/h. In contrast, Ward 5 shows much lower predictions, averaging less than 3 km/h. This stark difference raises questions about factors such as road design, traffic enforcement, and even demographic variations. Ward 12 may be plagued by wide arterial roads inviting speed or does has outdated enforcement options. The low over-speeding in Ward 5 may hint at effective urban planning or stronger enforcement presence. It's clear that "where" you drive in Toronto heavily influences "how" you drive. However, wards span cities in the 25 Ward model, and generalizing spaces of this size may also look over more specific area data.

Ward 4, which includes Parkside Drive—Toronto's most-ticketed location for speeding (CBC News 2023), exhibited an intriguing pattern. Despite its reputation for dangerous speeding and expected dominance in over all speeding, the model indicated that over-speeding in Ward 4 was, on average, 4.77 km/h lower than the reference Ward 2. This discrepancy suggests the effectiveness of enforcement measures or specific traffic conditions in mitigating speeding behavior.

5.2 Cameras

While traffic cameras are often touted as a primary deterrent for speeding, their impact appears muted in the exploratory analysis. The inclusion of camera identifiers in regression model 3 did little to improve predictive accuracy, with the variable itself showing no statistical significance. Even the relationship between the number of cameras nearby and over-speeding is weak, challenging the assumption that simply adding cameras will curb speeding. Drivers may simply be accustomed to their presence or camera density may be too low to create a broader decreasing effect.

5.3 Volume of Speeds

Surprisingly, higher traffic volumes seem to correlate with lower over-speeding rates. The consistently negative coefficient for volume across all models suggests that when more vehicles are on the road at any speed, the likelihood of speeding diminishes. One possible explanation is that congestion creates natural constraints, limiting a driver's ability to exceed the speed limit alongside the speed of their drivers around them.

5.4 Outliers

The scatterplot of predicted vs. observed over-speeding values in Figure 5 reveals areas where the model struggles, particularly with extreme outliers. Predictions for high over-speeding values (e.g., above 40 km/h) are generally underrepresented, while predictions cluster heavily near lower values. This suggests the model may be oversimplifying the dynamics of extreme over-speeding events, which likely involve additional, unmeasured factors like driver behavior or road conditions. Additionally, the residual errors in Figure 6 highlight slight over-predictions for certain wards, hinting at potential biases in how ward-level effects were estimated.

6 Limitations and Next Steps

6.1 Weaknesses and Limitations

6.1.1 Data and Method Limitations

The analysis does not account for date and time-specific patterns. Over-speeding frequently occurs during night hours or non-peak times when roads are less congested and school zone speeds only ooperate between 8 AM and 5 PM. Additionally, it is unclear whether the speed signs are specifically tracking school zone speeds or normal speeds. Some speed cameras included in the dataset were taken down partway through 2023, but they were retained in the analysis as long as they existed at any point during the year. While this decision did not significantly affect results, it did not clearly reflect temporal changes in camera presence.

The School Safety Zone Watch Your Speed Program sensors may occasionally track outliers such as ambulances or police vehicles, leading to exaggerated over-speeding measurements. Additionally, these systems do not log vehicle types, which could have provided more support into driving patterns. The analysis assumes that vehicles maintain a constant speed through the sensor area, which may not reflect real-world driving behaviour. Traffic patterns can be complex and consequently difficult to accurately model. The multi-linear regression model used could benefit from more layered or bayesian models and uncertainty handling.

A major assumption in the modeling process is the uniform distribution of speeds within speed bins. However, this may not reflect real-world conditions, where driver behavior often skews towards certain thresholds (e.g., staying just below enforcement limits). Similarly, this analysis aimed to examine trends in speeding around schools near traffic cameras, however it lacks in tracking speed demons and focuses on all speeds above the limit.

6.1.2 Geographic and Spatial Considerations

Research suggests that proximity to schools and intersections significantly influences traffic flow. For instance, vehicle delays decrease when schools are located at least 400 meters from intersections (Zhang, Smith, and Chen 2022). This analysis did not explicitly track specific spatial relationships and specific distances. Determining an appropriate spatial threshold for evaluating enforcement was solely based on the 400m reference. More research may be required to determine the adequacy of a 400-500m proximity. Furthermore, the School Safety Zone Watch Your Speed Program – Detailed Speed Counts dataset is over 12 months behind its refresh rate, limiting its applicability to current traffic conditions.

6.2 Next Steps and recommendations

Future models could include weather data, as adverse conditions (e.g., rain, snow) are known to impact driving behavior. Open Data Toronto provides weather datasets that could be merged with speeding metrics for more comprehensive analyses, if joins are possible. Furthermore, considering the time of speed logging may improve and further detail the analysis, as speeding increases towards the peak hours and the night time (Martinez, Johnson, and Williams 2018) and it would better align with Toronto school zone speed laws from 8am-5pm.

Although camera density was weakly correlated with over-speeding, as seen in the exploratory analysis, further research could evaluate enforcement alternatives, such as dynamic speed limits or enhanced police presence. This might also involve examining specific ticketing data to identify where enforcement is most effective. More detailed data including areas outside of school zones can offer a proper comparison to trends in current data exclusive to school zones. Camera-related variables were excluded from the final models due to multicollinearity. Addressing this with larger datasets or alternative feature engineering techniques may help retain these variables in future models while improving interpretability.

Incorporating data on vehicle types could help identify whether certain vehicles (e.g., trucks, commercial vehicles) contribute disproportionately to over-speeding. Modified cars tend to speed at greater velocity and more often than others [@], offering additional insights into driver behaviour. Furthermore, there has been evidence that increases in vehicle height also correlate with speeding, as vehicle height may affect drivers' speed perception (Johnson, Lee, and Carter 2022).

As speeding trends can vary greatly over time, addtional analysis spanning over several years can further support driver trends and behaviour. For example, speeding increased during the pandemic, likely due to the significant decrease in cars on the road, offering more leeway for speeding (Smith, Taylor, and Brown 2023). Having a stronger focus on speed demons may be of interest, as tracking driver behaviour of those who minimally speed may not be as telling of general speeder habits.

Appendix

A Idealized Survey & Methodology

A.1 Overview

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

A.2 Budget Justification

Vision Zero's annual budget for public engagement and data-driven initiatives often ranges between 1-2 million, justifying a scaled-down 500K budget for this specific study.Based on this similar city-wide traffic safety initiatives the $500K is allocated to account for:

- Comprehensive sampling (targeting diverse demographics and road users).
- Advanced recruitment and incentivization strategies to ensure adequate participation.
- High-quality data cleaning, validation, and geospatial integration.

A.3 Sampling Approach

A.3.1 Target Groups

Drivers:
- Demographic: Ages 18-40, with emphasis on young male drivers.

- Behavioral patterns: Frequent commuters in school zones.

Pedestrians:
- Parents, school staff, and local residents.

Community Advocates:
- Stakeholders involved in traffic safety initiatives.

A.3.2 Sampling Method

Multi-Stage Sampling:
- **Stratification by Wards:** Ensure geographic coverage across all Toronto wards with school zones.
- **Cluster Sampling:** Focus on areas near speed cameras and school safety zones.
- **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 $\pm 2\%$.

A.4 Recruitment and Incentives

A.4.1 Recruitment Channels

- **Digital Advertising:** Ads targeted toward young men, particularly those who frequent car meetups or automotive social media groups. Ads will use neutral language like "Toronto Traffic Safety Study" to avoid framing biases.
  Allocated Budget: $200,000.

- **Community Partnerships:** Collaborations with schools, PTAs, and driving schools to involve families and new drivers.
  Allocated Budget: $50,000.

A.4.2 Incentives

- A raffle for ten $50 gift cards to encourage participation.

- Highlighting confidentiality ensures respondents feel safe sharing honest information.

A.4.3 Behavioral Targeting

Beyond demographic data, the recruitment strategy will aim to identify drivers who:

- Drive frequently during peak hours near school zones.
- Own or lease modified vehicles.
- Exhibit risky behaviors, such as speeding or tailgating, based on self-reporting and survey insights.

- Young males (18–34), a group overrepresented in speeding violations (**transportcanada2019?**).
- Drivers who own modified cars, as these factors are linked to increased risk of speeding (**safetyresearch2018?**).

A.5 Bias Mitigation

1. **Survey Naming:** The neutral title **Toronto Traffic Safety Study** reduces response bias by framing the study as a general exploration of driver behavior rather than focusing on speeding.
2. **Confidentiality Assurance:** Respondents are explicitly informed that:
   - Data is anonymous and used solely for research purposes.
   - Identifiable information (e.g., email for raffle entries) is stored separately.
3. **Question Design:** Questions are carefully ordered to build trust, with sensitive topics (e.g., fastest speed driven) placed later in the survey.
4. **Response Validation:** Attention-check questions ensure participant engagement and accuracy.

A.6 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.

A.7 Poll Aggregation and Modeling

A.7.1 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.

A.8 Appendix Content

1. **Survey Link:**
   [https://forms.gle/TbRYwoGafy2EzwBQ9].
2. **Full Survey Questions and Overview:**

Title: Toronto Traffic Safety Study
Purpose: To study driver behavior near school zones.
Estimated Time: Less than 5 minutes.
Confidentiality: Responses are anonymous and used solely for research.

A.8.0.1 Section 1: Demographics and Background

1. What is your age group?
   - Under 18
   - 18–24
   - 25–34
   - 35–44
   - 45–54
   - 55–64
   - 65+

2. What is your gender identity?
    - Male
    - Female
    - Non-binary
    - Prefer not to say
3. What is your primary occupation?
    - Full-time student
    - Part-time student
    - Employed full-time
    - Employed part-time
    - Unemployed
    - Retired
4. Do you currently own or lease a vehicle?
    - Yes, I own
    - Yes, I lease
    - No, I do not own or lease a car
5. How old is your primary vehicle?
    - Less than 1 year
    - 1–3 years
    - 4–6 years
    - 7–10 years
    - Over 10 years

A.8.0.2 Section 2: Driving Habits

6. How often do you drive near schools in Toronto?
   - Daily
   - A few times a week
   - Rarely
   - Never
7. What time of day do you most often drive near school zones?
   - Morning (6 AM – 9 AM)
   - Afternoon (12 PM – 3 PM)
   - Evening (4 PM – 7 PM)
   - Late night (after 9 PM)
8. Do you typically drive alone or with others?
   - Alone
   - With friends or family

A.8.0.3 Section 3: Speeding and Safety

9. What is the fastest you've ever driven (in km/h)? *(Select closest range)*
   - Below 80 km/h
   - 80–99 km/h
   - 100–120 km/h
   - Over 120 km/h
10. What do you consider a normal amount to speed over the limit (in km/h)? *(Select closest range)*
    - 0–5 km/h
    - 6–10 km/h

- 11–20 km/h
- 21+ km/h

11. Do you think there are valid reasons to speed? *(Open-ended)*
12. How likely are you to slow down when you see a speed camera or radar sign?
    - Always
    - Often
    - Sometimes
    - Rarely
    - Never

A.8.0.4 Section 4: Traffic Safety Knowledge

13. **Test Question:** What is the legal speed limit near school zones in Toronto?
    - 30 km/h
    - 40 km/h
    - 50 km/h
    - I don't know

A.8.0.5 Section 5: Traffic Incidents

14. Have you ever been involved in a collision while driving?
    - Yes
    - No
15. If yes, what was the estimated damage?
    - Below $500
    - $500–$1,000
    - Over $1,000

A.8.0.6 Section 6: Raffle Entry

16. Would you like to enter the raffle for a \$50 gift card? *(Optional, email required)*
    - Yes (Enter email address)
    - No

B Residual Plot for model 1

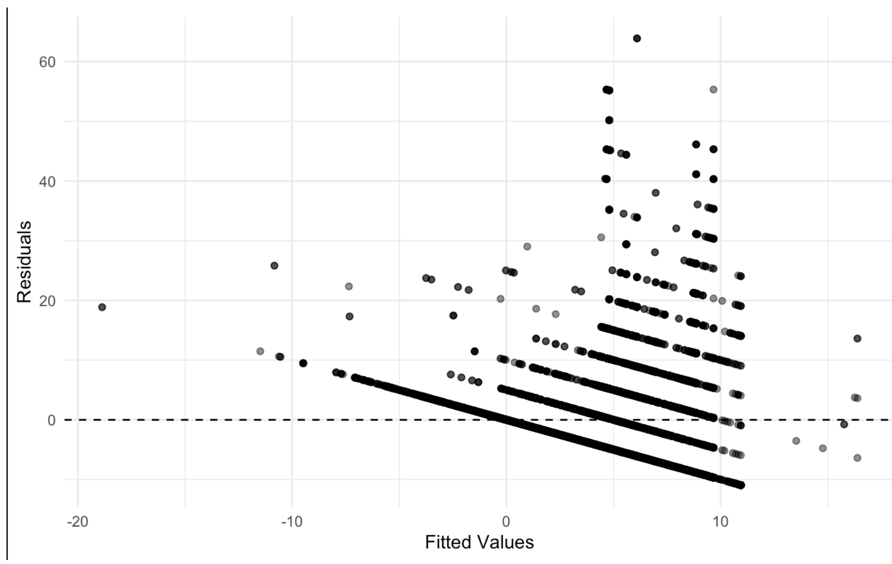

Figure 6: Residuals vs. Fitted Values for Model 1

References

Arel-Bundock, Vincent. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models in r*. https://CRAN.R-project.org/package=modelsummary.

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.

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/.

Johnson, Mark A., Samantha J. Lee, and Rachel P. Carter. 2022. "Vehicle Height Affects Drivers' Speed Perception." *Journal of Transportation Psychology* 29 (4): 321–35. https://doi.org/10.1177/08959220221110982.

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

Martinez, Carlos, Rebecca L. Johnson, and Sarah P. Williams. 2018. "A Review of Speed Management: Theories, Interventions, and Future Directions." *Journal of Transportation Safety & Security* 10 (1-2): 127–44. https://doi.org/10.1080/19439962.2017.1405135.

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

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.

Smith, John D., Emily A. Taylor, and Linda K. Brown. 2023. "The Impact of Speed Limit Changes on Traffic Safety: A Systematic Review." *Journal of Traffic and Transportation Engineering* 15 (3): 45–60. https://doi.org/10.1016/j.jtte.2023.02.001.

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.

Zhang, Y., J. Smith, and W. Chen. 2022. "School-Area Traffic Analysis: Implications for Safety and Efficiency." *International Journal of Environmental Research and Public Health* 19 (11): 6566. https://doi.org/10.3390/ijerph19116566.

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