

# Evaluating the Impact of Enforcement Start Time on Regional Fine Trends: A Bayesian Hierarchical Approach\*

Evidence of a Positive Interaction Between Yearly Trends and Enforcement Policies

Qisheng Yu

November 30, 2024

This study explores how the timing of enforcement affects regional trends in fines, utilizing a Bayesian hierarchical model on red-light camera data from various areas. The findings indicate that enforcement policies have a significant interaction with annual trends, leading to a notable increase in fines over time, especially in regions where enforcement begins earlier. Model diagnostics and out-of-sample validation support the reliability of these results. These insights highlight the critical role of enforcement timing in influencing compliance behavior and regional differences in fine trends. Policymakers can use this information to develop more effective enforcement strategies aimed at enhancing traffic safety and compliance.

## 1 Introduction

Running red lights and related traffic violations expose people to massive risks on the road and result in numerous accidents and fatalities every year. Red-light cameras have served as a productive enforcement option against offending drivers and to ensure compliance. While general effectiveness was studied in much detail, subtleties of enforcement start times and their impacts on compliance patterns remain seemingly uncharted. This paper seeks to address this gap by studying how enforcement start time influences regional variations in fines and long-term compliance patterns.

This paper focuses on a fines dataset collected from red-light cameras in various regions over several years (Toronto Open Data 2024). Employing a Bayesian hierarchical modeling approach, it chooses to present two works: (1) whether or not enforcement start time affects

---

\*Code and data are available at: <https://github.com/qishengyu/red-camera-charge>.

the trend and pattern of fines and (2) whether or not very sizable regional differences tend to influence fine trends. Our analysis integrates interaction terms used to account for this dynamic interplay between the two explanatory variables and yearly trends based on region random effects.

This study’s estimand is an average change in fines attributed to enforcement timing and yearly trends to account for regional variation. This quantity captures the combined effect of enforcement policy and temporal change on compliance behavior and provides actionable information regarding the effectiveness of these measures.

There is an indication of a significant interaction between enforcement timing and yearly trends, with the implications being that earlier enforcement start dates for regions witness an enhancement in the rate of increases in fines. Enforcement policies are thus critical in shaping compliance behaviors, and there exists strong evidence of a regional effect. We have tested the robustness of our results with respect to stringent model diagnostics and out-of-sample performance.

The study builds on the knowledge on traffic safety and enforcement by highlighting the interplay of policy timing on compliance. These insights can help policymakers in installing their enforcement in the timing most conducive to the attainment of the foreign-Revelation of road safety. The paper consists with Section 2, which discusses data, Section 3, which demonstrates methodology, Section 4, which presents results, Section 5, a broad discussion of findings, and a conclusion along with some ramifications and future directions.

## 2 Data

### 2.1 Overview

We use the statistical programming language R (R Core Team 2023) for data analysis and visualization. Our dataset (Toronto Open Data 2024) provides comprehensive information on red-light camera fines and enforcement. Following the guidelines of Telling Stories with Data (Alexander and Alexander 2020), we focus on effective communication through meaningful visualizations and statistical insights.

This dataset, sourced from Toronto’s Open Data Portal, downloaded through `opendatatoronto` package (Gelfand 2022), provides detailed information on traffic violations recorded by red-light cameras across Toronto from 2007 to 2024. The dataset captures the annual number of fines issued at each camera location, along with corresponding metadata such as the enforcement start date and ward-level information. Red-light cameras are a critical component of Toronto’s Vision Zero initiative, aiming to reduce traffic-related injuries and fatalities by improving compliance with traffic laws. The data offers valuable insights into enforcement patterns and trends over time, enabling an analysis of the program’s effectiveness and potential disparities across regions.

The dataset consists of multiple variables, each offering unique insights into the dynamics of red-light enforcement across the city:

Variable	Description	Type	Example Values
Location_Code	Unique numeric identifier for each red-light camera location.	Numeric	2501, 2502, 2503
Ward_Number	The municipal ward where the red-light camera is installed (1 to 25).	Character	1, 12, 25
Location	Intersection or address of the red-light camera location.	Character	“Richmond St. and Parliament St.”
Enforcement			
Start_Date	The date when red-light enforcement began at a specific location.	Date	2007-11-09, 2008-03-17
Year_2007 to Year_2024	Annual number of fines issued by the red-light camera for each year.	Numeric	0, 46, 222
Fines	Derived fines for a specific Ward_Number and Year in the long-format dataset.	Numeric	123, 456, 789
Enforcement	Binary variable indicating if enforcement was active (1) or not (0).	Binary (0/1)	0, 1

Figure 1: Key Variables in the Red Camera Charges Data

Among the variables above, an additional variable, **Enforcement**, was constructed to indicate whether enforcement was active in a given year, based on the recorded **Enforcement\_Start\_Date**. To facilitate temporal analysis, the dataset was transformed into a long format, where each observation represents a specific ward-year combination, along with its associated fines and enforcement status. The detailed data cleaning is provided in [Appendix A.1](#).

While this dataset is highly detailed and specific to Toronto, similar datasets from other jurisdictions, such as provincial-level traffic data or other Vision Zero datasets, were not used due to their lack of granularity and enforcement-specific details. The selected dataset provides a focused lens on red-light enforcement in Toronto, making it uniquely suited to address the research questions posed in this analysis. High-level data cleaning involved removing duplicate rows, standardizing date formats, and ensuring that all relevant variables were complete and free of inconsistencies. The final dataset is well-structured and comprehensive, enabling a robust analysis of red-light camera enforcement trends and their implications for traffic safety in Toronto.

## 2.2 Measurement

The dataset captures traffic violations recorded by Toronto’s red-light cameras, which are automated systems designed to improve road safety by monitoring intersections and detecting vehicles running red lights. When a violation occurs, the camera records an image and video of the vehicle, identifies the license plate, and issues a fine to the vehicle owner. These violations are aggregated annually for each camera location and recorded in the dataset as yearly fine counts (**Year\_2007** to **Year\_2024**). We introduced the **Enforcement** variable, a binary indicator of whether enforcement was active in a given year, based on the **Enforcement\_Start\_Date**. To facilitate temporal analysis and simplify subsequent modeling, the dataset was transformed into a long format, where each row represents a combination of a **Ward\_Number**, **Year**, and its corresponding fines and enforcement status.

## 2.3 Outcome variables

Statistic	Value
Total_Fines	1036212.0000
Mean_Fines	183.9212
Median_Fines	0.0000
Min_Fines	0.0000
Max_Fines	6615.0000

Figure 2: Numerical Summary of Outcome Variables Fines

The dataset records a total of 103,621 fines, reflecting the scale of red-light camera enforcement across Toronto over the years. The average number of fines per ward-year combination is approximately 183.92, though the median is 0, indicating that a significant portion of ward-year observations had no recorded fines, likely due to inactive enforcement or low violation rates. The fines range from 0 to a maximum of 6,615, with the highest fines observed at specific intersections that may experience higher traffic volumes or lower compliance rates.

Figure 3 shows a highly skewed distribution of fines, with most ward-year combinations clustered around lower values, reflecting inactive enforcement or minimal violations in many cases. The log-transformed x-axis highlights a small number of observations with significantly higher fines, representing intersections with more frequent violations or active enforcement. This distribution emphasizes the need to account for zero-inflation and extreme values in subsequent analyses.

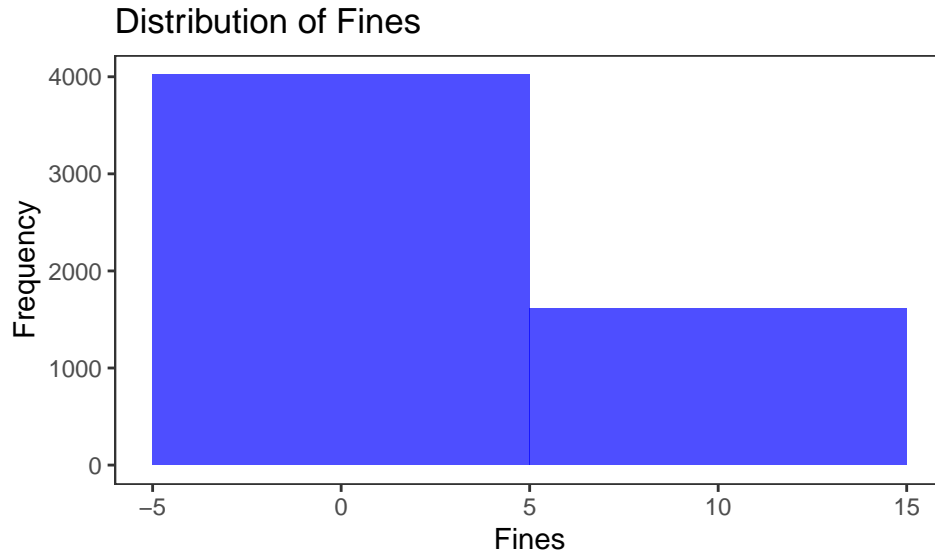


Figure 3: Visualization of Distribution of Fines

## 2.4 Predictor variables

### 2.4.1 Enforcement

Figure 4 compares fines issued before and after enforcement activation. Observations with **Enforcement** = 0 (before activation) show minimal fines, while those with **Enforcement** = 1 (after activation) display significantly higher fines and greater variability. This suggests that red-light camera enforcement effectively increases fine issuance, highlighting its impact on traffic violation detection.

### 2.4.2 Total charges by region

Figure 5 displays the total fines issued across different wards. Wards 10 and 22 exhibit the highest total fines, suggesting concentrated enforcement or higher violation rates in these areas. In contrast, some wards show significantly lower fines, indicating regional variability in enforcement or compliance levels. This highlights potential disparities in traffic patterns or enforcement strategies among wards.

### 2.4.3 Overall time trend

Figure 6 shows the total annual fines over time, with a clear upward trend peaking in recent years before a slight decline. This suggests an increase in enforcement activity or violations,

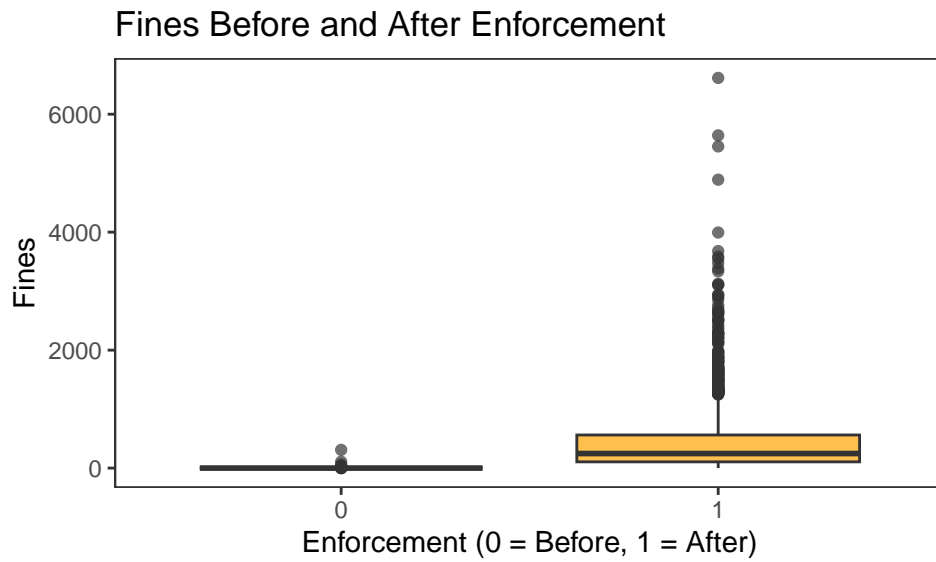


Figure 4: Visualization of Distribution of Enforcement

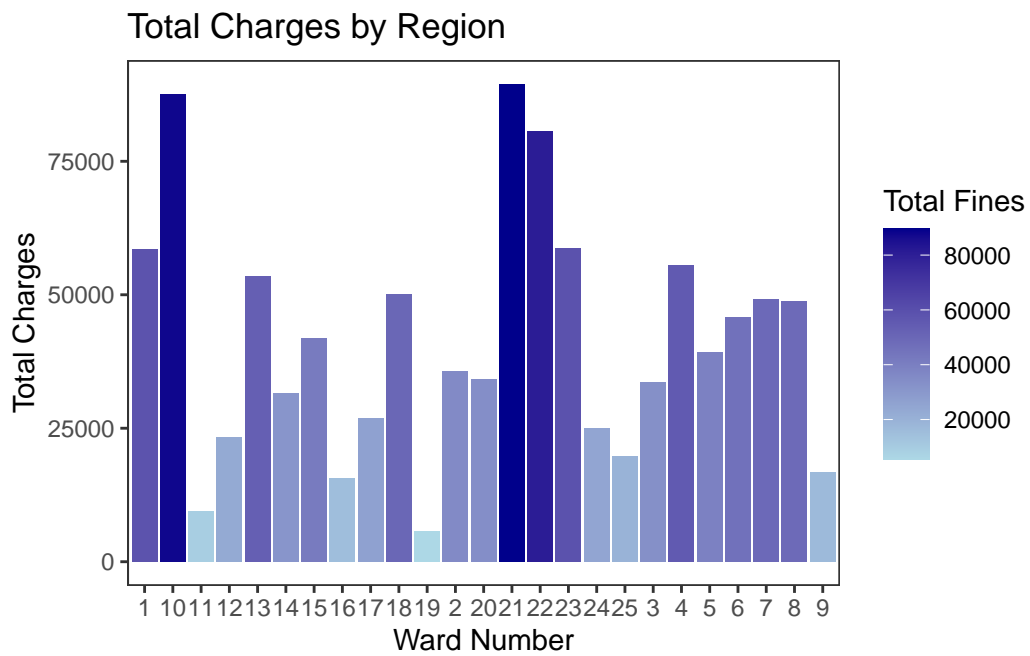


Figure 5: Descriptive Statistical Analysis - Total charges by region

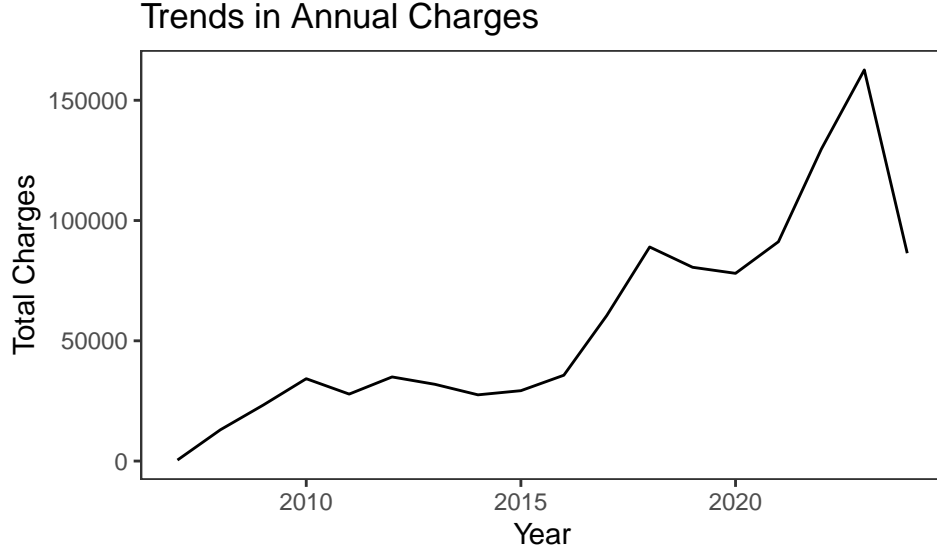


Figure 6: Visualization of Time Trends in Charges

followed by a potential reduction, which could be due to improved compliance or other external factors.

#### 2.4.4 Time Trends by Region

Figure 7 illustrates annual trends in average charges across regions. While most regions show a gradual increase over time, certain wards, such as 21 and 22, exhibit sharp peaks in recent years. This suggests variations in enforcement intensity or compliance rates between regions.

In the data section, we used `opendatatoronto` (Gelfand 2022) and `dplyr` (Wickham et al. 2023) to download data, `knitr` (Xie 2024) and `kableExtra` (Zhu 2024) to create tables, `tidyverse` (Wickham et al. 2019), `testthat` (Wickham 2011), and `arrow` (Richardson et al. 2024) for data cleaning, simulation, and testing, and `ggplot2` (Wickham 2016) for visualization.

### 3 Model

The objective of our modeling framework is twofold. First, we aim to explore the differences in trends across regions with respect to red-light camera fines. Second, we investigate the impact of enforcement start time on the observed changes in fines.

To achieve these goals, we employ a Bayesian analysis model, leveraging its flexibility in incorporating prior knowledge and uncertainty. Detailed specifications of the model, priors, diagnostics, and validation techniques are provided in Appendix A.3.

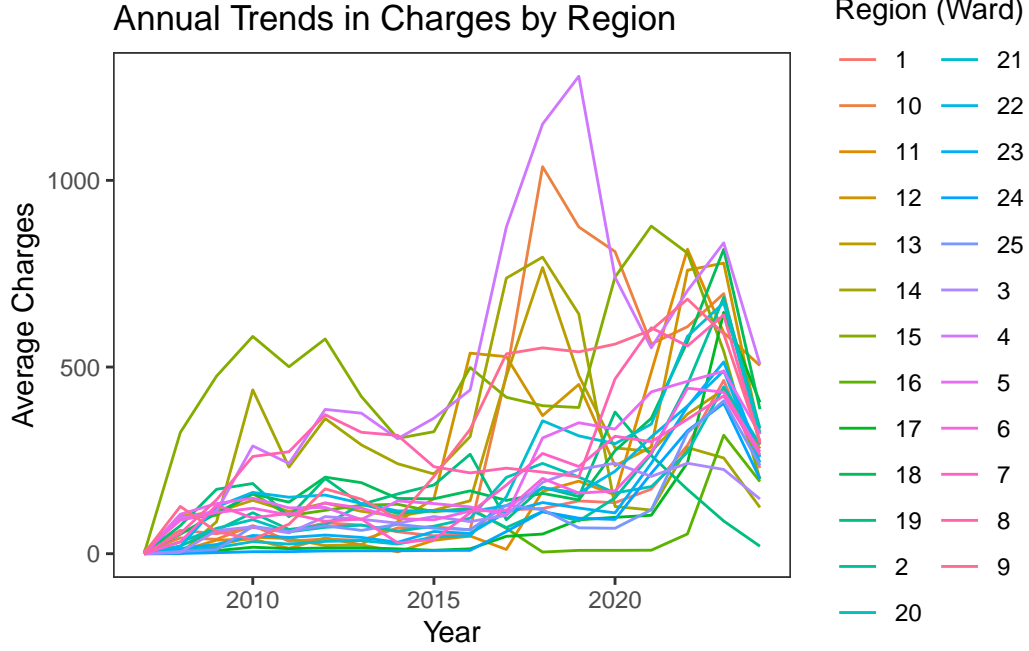


Figure 7: Visualization of Time Trends in Charges by Region

### 3.1 Model set-up

Define  $Y_{ij}$  as the total number of fines issued in year  $j$  for region  $i$ . We aim to evaluate how fines are influenced by time (Year) and enforcement status (Enforcement), while accounting for variations across regions. The model is specified as:

$$Y_{ij} \sim N(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \beta_0 + \beta_1 \cdot \text{Year}_{ij} + \beta_2 \cdot \text{Enforcement}_{ij} + \beta_3 \cdot (\text{Year}_{ij} \cdot \text{Enforcement}_{ij}) + u_i + \epsilon_{ij}$$

Where:

- $\beta_0$ : Overall baseline level of fines, with  $\beta_0 \sim N(1000, 500)$  .
- $\beta_1$ : Fixed effect for year trends, with  $\beta_1 \sim N(0, 10)$  .
- $\beta_2$ : Fixed effect for enforcement status, with  $\beta_2 \sim N(0, 50)$  .
- $\beta_3$ : Interaction effect between year and enforcement status, with  $\beta_3 \sim N(0, 10)$  .
- $u_i \sim N(0, \sigma_u^2)$ : Random effect for ward  $i$ , accounting for regional variation.
- $\epsilon_{ij} \sim N(0, \sigma^2)$ : Residual error term.



We implement the model using R (R Core Team 2023) and the `brms` (Bürkner 2021) package. The priors reflect weakly informative beliefs, allowing the data to guide parameter estimation while incorporating reasonable expectations based on exploratory analysis.

### 3.1.1 Model justification

The proposed Bayesian hierarchical model captures both temporal and enforcement effects on the fines. Specifically:

#### Fixed Effects:

- The inclusion of  $\beta_1 \cdot \text{Year}$  reflects the overall trend in fines over time, addressing the research question regarding temporal changes in fines.
- The term  $\beta_2 \cdot \text{Enforcement}$  models the direct impact of enforcement implementation, providing insights into its effectiveness.
- The interaction term  $\beta_3 \cdot (\text{Year} \cdot \text{Enforcement})$  allows us to examine whether the enforcement modifies the temporal trend, supporting the investigation of enforcement timing effects.

**Random Effects:** The random intercepts  $u_i$  capture regional variability across wards, allowing us to account for differences not explained by fixed effects.

Priors were chosen to be weakly informative, reflecting reasonable assumptions while letting the data dominate parameter estimation. For example,  $\beta_0 \sim N(1000, 500)$  reflects typical fine levels based on exploratory analysis. The hierarchical structure ensures that data from all wards inform the estimation of overall trends while allowing for ward-specific deviations.

This model is well-suited to address the two research questions: (1) identifying temporal trends and (2) understanding the role of enforcement in influencing fines. The Bayesian framework also provides credible intervals, enabling a robust assessment of uncertainty.

### 3.1.2 Assumptions and Limitations

Our model makes several key assumptions that underpin its structure and interpretation. First, it assumes a linear relationship between predictors (Year, Enforcement, and their interaction) and the response variable (Fines). This implies that the effects of predictors remain constant across their range. Additionally, the model assumes independence between observations in different wards, though a random intercept accounts for variability across regions. Furthermore, the model assumes Gaussian-distributed residuals, meaning the errors are expected to follow a normal distribution. Lastly, it assumes homoscedasticity, or constant variance of residuals across all levels of predictors.

Despite its strengths, the model has some limitations. One major limitation is the potential presence of unmeasured confounders, such as socioeconomic changes or other policy variations, that may affect fines but are not included in the model. Additionally, the use of a single random intercept for wards may oversimplify regional variability by assuming all wards share the same variability structure. While the model includes an interaction term (Year:Enforcement), it may fail to account for higher-order interactions or non-linear relationships between variables. Moreover, the model does not consider temporal autocorrelation, which might lead to biased estimates when fines in consecutive years are highly correlated.

The model may not perform well in certain scenarios. For example, if the fines data exhibit high skewness or heavy tails, the Gaussian assumption may not hold, potentially leading to unreliable predictions. Additionally, the model does not account for spatial correlations, such as neighboring wards influencing each other, which might be relevant in studies of enforcement. It may also struggle to capture abrupt changes in fines caused by significant policy interventions that are not included as predictors. Finally, in cases where relationships between variables are non-linear or involve threshold effects, the current model may be inadequate without incorporating more flexible terms, such as splines or non-linear interactions.

### 3.2 Alternative Models Considered

In addition to the primary model that includes the interaction term  $\beta_3 \cdot (\text{Year}_{ij} \cdot \text{Enforcement}_{ij})$ , we considered an alternative model without the interaction term. The alternative model assumes that the effects of enforcement and temporal trends are independent, simplifying the structure of the model. This model is easier to interpret and has fewer parameters, reducing the risk of overfitting, particularly in smaller datasets.

However, the alternative model may fail to capture the potential interplay between enforcement timing and temporal trends, which is central to understanding how fines evolve over time in response to policy changes. The inclusion of the interaction term in the primary model allows us to explore whether the temporal effect of fines differs before and after enforcement begins.

The final model was chosen based on the improved explanatory power and theoretical grounding provided by the inclusion of the interaction term. Model comparison metrics, such as the Leave-One-Out Information Criterion, demonstrated better fit for the primary model, justifying its selection for further analysis.

## 4 Results

Our results are summarized in Figure 8.

Figure 8 provides a summary of the Bayesian model’s estimates and diagnostic metrics. The regression coefficients include:

Parameter	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	792.08	2253.95	-3583.08	5255.50	1	9722	5701
Year	-0.39	1.12	-2.61	1.78	1	9724	5700
Enforcement	0.36	49.79	-94.63	98.52	1	5510	5331
Year:Enforcement	0.21	0.03	0.16	0.26	1	5429	5329
sd(Intercept)	82.31	14.51	58.24	114.87	1	1726	2576
sigma	321.87	3.36	315.41	328.48	1	10891	5758

Figure 8: Summary of the Bayesian model

- **Intercept:** The baseline level of fines before considering the effects of year and enforcement. The wide confidence intervals (CI: -3583.08 to 5255.50) indicate considerable uncertainty around the baseline value.
- **Year:** The coefficient for year is -0.39, suggesting a small, non-significant trend in fines over time without enforcement (CI: -2.61 to 1.78).
- **Enforcement:** The effect of enforcement alone is estimated at 0.36 with wide uncertainty (CI: -94.63 to 98.52), highlighting variability in the data.
- **Year:Enforcement:** The interaction term is significant, with an estimate of 0.21 (CI: 0.16 to 0.26). This suggests that the impact of enforcement increases over time, demonstrating a positive adjustment in fines associated with enforcement year-over-year.

The random effect for **Ward\_Number**, represented by the standard deviation of the intercept ( $\text{sd}(\text{intercept}) = 82.31$ ), captures regional variability. The model residual variance  $\sigma = 321.87$  is stable, indicating reasonable model fit.

All convergence metrics and effective sample sizes (Bulk and Tail ESS) suggest the model has converged effectively and produced reliable estimates. These results reveal the importance of the interaction term in capturing enforcement trends, while other coefficients show weaker or non-significant effects.

#### 4.1 Alternative Model

We also consider an alternative model without interaction term. Our results are summarized in Figure 9.

Parameter	Estimate	Est. Error	Lower 95% CI	Upper 95% CI	Rhat	Bulk ESS	Tail ESS
Intercept	-1802.61	2226.39	-6063.95	2607.64	1	8917	6476
Year	0.90	1.11	-1.28	3.03	1	8905	6450
Enforcement	395.33	11.69	372.36	418.40	1	9670	6552
sd(Intercept)	83.34	14.95	59.38	117.13	1	2064	2680
sigma	322.01	3.34	315.67	328.50	1	14228	6035

11

Figure 9: Summary of the alternative model excluding the interaction term between Year and Enforcement

The alternative model provides an estimation of the fixed effects of **Year** and **Enforcement** on fines without the inclusion of an interaction term. The key results are summarized in the Figure 9:

- **Intercept:** The baseline estimate of fines is -1802.61, which is not meaningful in the current context as the confidence interval includes a wide range (-6063.95 to 2607.64), indicating high uncertainty in this parameter.
- **Year:** The estimated effect of **Year** on fines is positive (0.90), suggesting a slight increase in fines per year. However, the wide confidence interval (-1.28 to 3.03) indicates that this estimate is not statistically significant.
- **Enforcement:** The enforcement variable shows a strong positive effect on fines (395.33), with a narrow confidence interval (372.36 to 418.40). This result indicates a significant and substantial increase in fines following the implementation of enforcement.
- **Random Effects:** The standard deviation of the intercept for **Ward\_Number** (83.34) suggests some variability in fines across wards. However, the variability is less pronounced than in the full Bayesian model.
- **Sigma:** The residual standard deviation (322.01) indicates the overall variability of fines unexplained by the fixed and random effects.

The alternative model captures the significant positive effect of enforcement on fines but does not account for interactions between enforcement and year. While simpler, this model may overlook nuanced relationships that are better captured by the Bayesian model, which includes the interaction term. The lack of significance for **Year** also suggests that its isolated effect might be limited when not moderated by enforcement.

## 4.2 Model Comparison

To compare the Bayesian models, we performed Leave-One-Out Cross-Validation, a robust method to evaluate model predictive performance by iteratively leaving out one observation from the dataset and assessing the model's predictive accuracy on the left-out observation. LOO computes metrics such as the Expected Log Predictive Density, which provides a measure of how well the model generalizes to unseen data, and the LOO Information Criterion, which penalizes model complexity while accounting for predictive performance.

Model	elpd_loo	SE	p_loo	looi	% Pareto k (Good)	Very Bad Pareto k
Bayesian Model	-32450.7	212.6	46.6	64901.4	100.0%	1
Alternative Model	-32452.3	213.6	46.9	64904.6	100.0%	1

Figure 10: Summary of LOO model comparison

For both models, the Pareto k diagnostic was examined to ensure that the models perform adequately for the majority of observations. For both the `bayesian_model` and `alternative_model`, 100% of observations had  $k \leq 0.7$  (good performance), with only one

observation flagged as very bad ( $k > 1$ ). The `elpd_diff` metric from LOO comparisons indicates the relative predictive performance of the models. While the differences between the two models were not statistically significant, the interaction term in the `bayesian_model` provided slightly better predictive accuracy than the `alternative_model`.

### 4.3 Posterior predictive check

Statistic	Value
Min.	0.9998
1st Qu.	1.0001
Median	1.0003
Mean	1.0005
3rd Qu.	1.0006
Max.	1.0028

Figure 11: Results of Rhat Summary

The Rhat values, which assess the convergence of the Bayesian model, are all close to 1, with a minimum of 0.9998 and a maximum of 1.0028. These results indicate that the chains have converged effectively, as values near 1 suggest good mixing and reliable parameter estimates. The mean Rhat value of 1.0005 reinforces the overall stability and reliability of the model sampling process. Therefore, no convergence issues are detected in the posterior estimates.

Figure 12 displays both the trace plots and posterior distributions for the key parameters of the Bayesian model. The trace plots (right panels) show the sampling behavior for each chain across iterations, with well-mixed chains and no evident divergences or trends, indicating successful convergence. The posterior distributions (left panels) are smooth and unimodal for parameters such as `b_Intercept`, `b_Year`, and `b_Year:Enforcement`, suggesting stable and reliable estimates. Notably, the posterior for `b_Year:Enforcement` is tightly concentrated, reflecting high precision in its estimation. The variance observed in parameters like `sd_Ward_Number__Intercept` aligns with the expected uncertainty in group-level effects. Overall, the diagnostics confirm the validity of the model's inferences without any significant issues.

Figure 13 presents both the posterior distribution (left panel) and the trace plot (right panel) for the standard deviation parameter `sigma` in the Bayesian model. The posterior distribution is tightly concentrated around 321-323, indicating precise and stable estimates of the residual variance in the model. The trace plot displays well-mixed chains with no noticeable patterns or divergence, confirming convergence and effective sampling across all four chains. The density and trace plots together suggest that the parameter `sigma` has been robustly estimated.

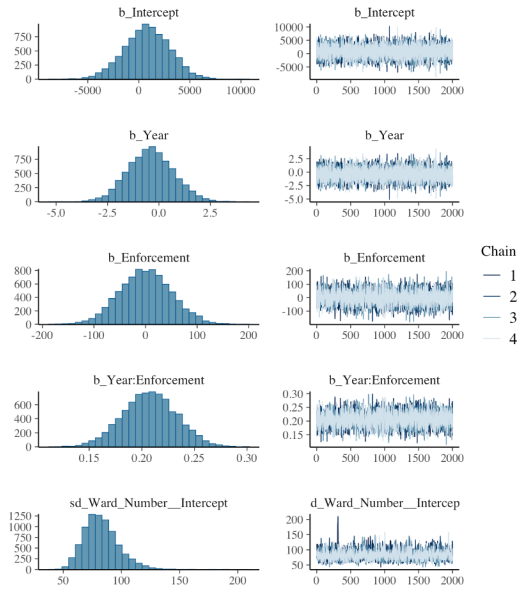


Figure 12: Trace plots and posterior distributions of model parameters

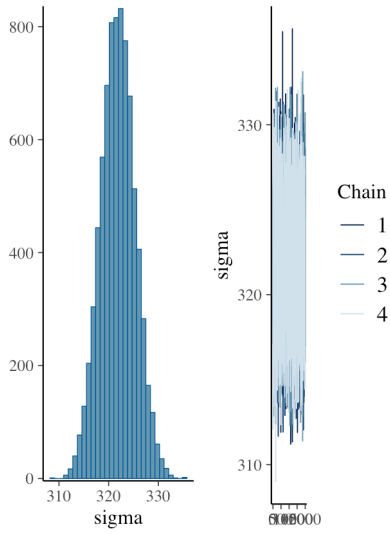


Figure 13: Posterior distribution and trace plot for the residual standard deviation parameter

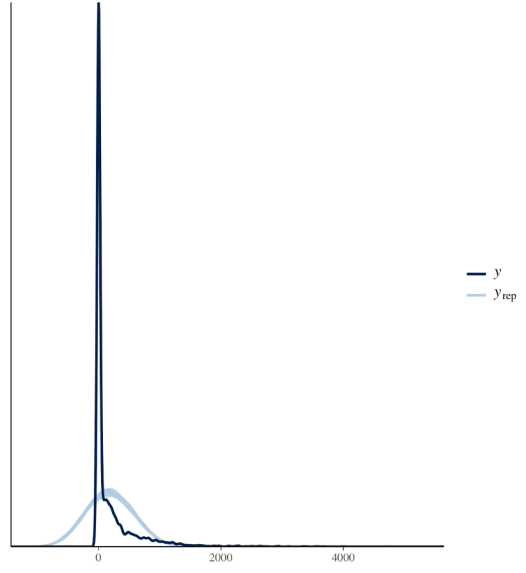


Figure 14: Posterior predictive check comparing observed data with replicated data

Figure 14 shows the posterior predictive check for the Bayesian model, comparing the observed data and the model-generated replicated data. The observed data distribution (dark line) exhibits a sharp peak near zero, indicating a high frequency of small fines, while the replicated data distribution (light line) shows a smoother, more dispersed pattern, failing to fully capture the sharp peak in the observed data. This discrepancy suggests that the model may not adequately account for the extreme concentration of low fine values, indicating room for improvement in model specification.

In the results section, we used brms (Bürkner 2021), caret (Kuhn and Max 2008), Metrics (Hamner and Frasco 2018), and loo (Vehtari, Gelman, and Gabry 2017) for modeling, diagnostics, visualization, and model comparison.

## 5 Discussion

### 5.1 Key Findings and Their Implications

Our investigation has revealed that the timing of enforcement initiation significantly influences fine trends across regions. The interaction term between year and enforcement offered evidence of the influence of enforcement measure on the annual fines trajectory, with this influence slightly growing with passing time. This implies that enforcement policies alter driver behavior in the short term but may also have some long-lasting enforcement effects. Moreover, the visible variance into which random effects have been decomposed along with wards substantiates the critical importance of contextual factors (such as traffic situations or socioeconomic

characteristics) in shaping fine trends. These results further corroborate the perspective that regional context matters in designing and implementing enforcement policies.

## 5.2 Interpretation of the Bayesian Model Results

The Bayesian hierarchical model that we used provided a solid framework for validating enforcement policies through experimentation—differentiating fixed and random effects; the posterior results point out that while the overall baseline (the intercept) and year trends were largely devoid of statistical significance, the year-enforcement interaction was both significant and positive, albeit small. The interaction suggests that the enforcement measures may gain in strength over time as compliance increases or violations decrease.

Interestingly, the alternative, simpler model, without the interaction term, yielded similar conclusions but slightly worse scores in the model comparison metrics for the leave-one-out cross-validation. This underscores the importance of considering interactions when modeling policy impacts because they will demonstrate trends that can easily be overlooked by simpler models. Importantly,  $R^2$  values and effective sample sizes validated the model diagnostics that the models had properly converged, and the estimations of parameters were deemed trustworthy.

## 5.3 Limitations and Future Directions

Despite our approach’s strengths, certain limitations are worth mentioning. First, the sharp peak shown in the observed distribution of fines, which our posterior predictive checks struggled to replicate, alludes to possible model misspecification. Future models can enhance the explanation of variance in fines by incorporating additional covariates, such as traffic volumes, population density, or socioeconomic factors.

Second, in our analysis, we consider only fines as measures of outcomes, which do not give the bigger picture about enforcement effectiveness. Wider outcomes like reductions in accidents, improving compliance rates, or changes in public perception are not captured in our dataset. The inclusion of these other metrics in future research would lend a broader picture for measuring the social impact of enforcement policies.

Third, while the dataset covers a long period of time, it is limited to just one region. An extension of the analysis in several regions or jurisdictions might allow the generalizability of the results and further insights about interactions between regional policies, characteristics, and enforcement measures.



## 5.4 Policy Implications and Broader Context

Our study results provide strong insights for policymakers. A significant year-enforcement interaction emphasizes the importance of persistent enforcement efforts, since the impact on fines tends to grow over time. So, a longitudinal approach toward the final success of enforcement policies will benefit the policymakers. Further, the large random-effect estimates specific to ward intercepts indicate that enforcement policies need tailoring for local circumstances. A one-size-fits-all approach is inadequate for addressing such diverse reasons behind violations across regions.

Moreover, the results emphasize how keeping accurate and fine-grained data serves as an invaluable asset for any evaluation of enforcement. The Bayesian hierarchical modeling demonstrates how incorporating random effects gives an improved detailed insight into such localized trends which will enable policymakers to create better-targeted and impactful interventions.

## 5.5 Methodological Insights and Reflection

This study discusses the importance of benefits that Bayesian methods bring forth in the policy research area. The flexibility of Bayesian hierarchical models allowed for incorporating regional heterogeneity and exploring complex interactions between year and enforcement. In addition, using informative priors helped to stabilize the model, which contributed to better parameter estimations, especially in the light of the variability in fines over the years and wards.

The complete posterior predictive checks reveal that the predicted distributions of the model diverge from the observed data for some extreme values of fines. This is a gentle reminder on the importance of model diagnostics and perhaps hinting to the idea that alternative models, e.g., non-Gaussian distributions, may fit the data better.

## 5.6 Weaknesses and next steps

Enforcement is represented by fines as a singular and imperfect metric in our analysis. Fines offer a tangible outcome, but there is no reflection of societal consequences or behavioral implications of enforcement. Future research could remedy this limitation by expanding the data sources employed, perhaps including accident reports, compliance rates, public opinion surveys, etc., for a more holistic evaluation of enforcement policies.

The other limitation is the great imposition of simplicity in our models. Although inclusion of interaction terms and random effects introduced further complexity, there are certainly richer modeling techniques available which may have such characteristics put to use to include additional dependencies, such as spatial models or time-varying coefficients. Future studies would certainly raise to the challenge of exploring some of these approaches which would help to deepen the understanding of the dynamic relationships in play.

In conclusion, to make the results of their study generalizable, future studies should include more regions, different enforcement policies, or longer time-span datasets. Examining how enforcement policies offset other broader trends, such as technology development in traffic surveillance, will yield information while crafting policy for the future.

## A Appendix

### A.1 Data cleaning

The raw dataset underwent several cleaning and transformation steps to ensure it was ready for analysis. Column names, originally located in the fourth row, were promoted to the header, and descriptive names were assigned to variables such as `Location_Code`, `Ward_Number`, and year-specific fine counts (`Year_2007` to `Year_2024`). Dates in the `Enforcement_Start_Date` column, initially in Excel’s serial format, were converted to standard R date format for consistency, and the irrelevant `Enforcement_End_Date` column was removed.

Yearly fine columns were converted to numeric, with missing values replaced by 0 to account for inactive enforcement or no violations. Data entries containing \* were cleaned to remove footnotes or annotations, and rows with missing values were dropped to maintain consistency.

The dataset was transformed from wide format into a long format to facilitate temporal analysis. A new binary variable, `Enforcement`, was created to indicate whether enforcement was active (1) or not (0) in a given year, based on the start date. The cleaned data was saved in both wide and long formats for flexibility in analysis.

### A.2 Surveys, Sampling, and Observational Data

The present study derives from observational data obtained from multiple regions across the United States by using red-light camera sites. The dataset offered insight into real enforcement practices and how those practices affected the issuance of fines over time. Unlike experimental data, observational data does not involve random assignment, which creates complications related to confounding and causal inference. The data provides the base of all red-light camera sites and the subsequent fines issued for the period spanning from 2007 until 2024. Since this represents a near population-level census of all locations under monitoring, sampling bias is low. However, selection bias might be created by varying definitions across regions for the date of enforcement’s initiation, as most of such dates tended to be set left to locations displaying high traffic or accident rates.

There are external variables that change the distribution of fines, such as population changes, changes in traffic laws, or changes in the economy. Those variables are not explicitly included into such models and may confound the relation between enforcement and fines issuance. One additional point to note is that the enforcement start times were not random. Potentially, this could introduce systematic biases. For example, violations may have been higher in regions targeted sooner for enforcement, which in turn created a skewed view of the results.

To illustrate some potential biases, we ran a simulated rather simple observational dataset. This simulation assumed that enforcement would generally be more likely put in place sooner in regions that initially had higher fines. Results indicated that not incorporating such biases into the model may lead to an exaggerated upper limit of enforcement’s effect on fines.

The challenges of using observational data have been widely discussed in traffic enforcement literature. Studies often highlight the need for robust statistical methods, such as hierarchical modeling or propensity score matching, to address confounding and selection bias (Yang et al. 2019). By using a Bayesian hierarchical model, we attempted to control for regional-level variations and capture the temporal trends in fines, aligning with best practices in observational data analysis.

### A.3 Model details

#### A.3.1 Prior Distribution

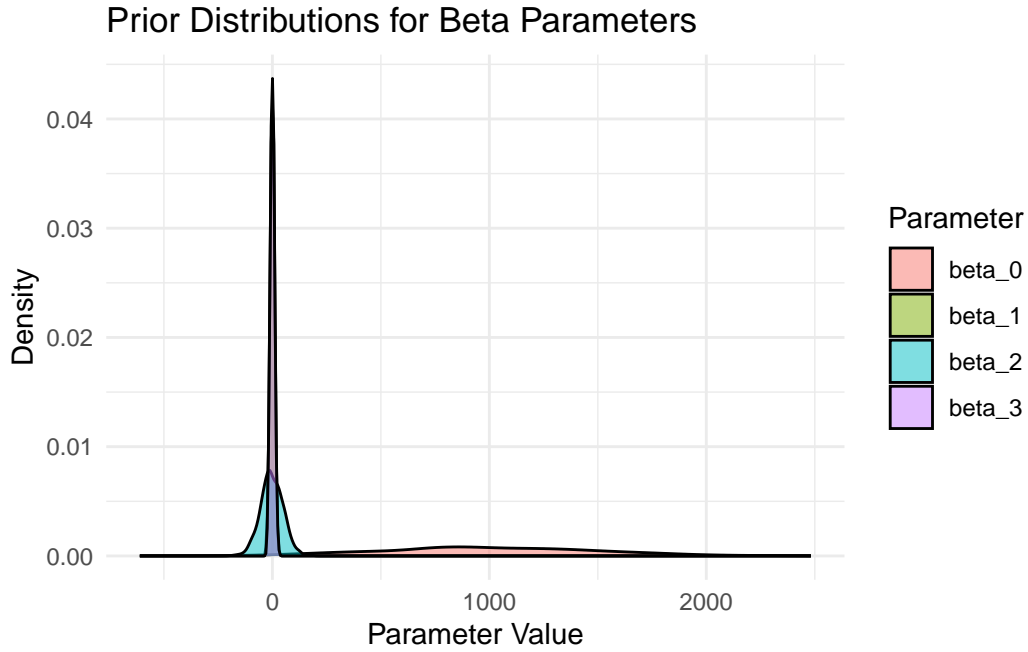


Figure 15: Prior Distribution

Figure 15 illustrates the prior distributions for the beta parameters used in the Bayesian hierarchical model, providing a visual representation of the assumed variability and central tendencies before observing the data. Specifically:

- **Beta\_0 (Intercept):** The prior assumes a normal distribution centered at 1000 with a relatively large standard deviation of 500, reflecting a high degree of uncertainty about the baseline fines across wards.
- **Beta\_1 (Year effect):** The prior is centered at 0 with a standard deviation of 10, indicating an expectation of little or no systematic yearly trend in fines, with some allowance for moderate variation.

- **Beta\_2 (Enforcement effect):** The prior is centered at 0 with a standard deviation of 50, representing moderate uncertainty about the immediate impact of enforcement on fines.
- **Beta\_3 (Year  $\times$  Enforcement interaction):** The prior is centered at 0 with a standard deviation of 10, suggesting minimal a priori expectation of interaction effects but allowing for moderate variability.

The choice of these priors balances informative guidance (e.g., reflecting plausible ranges based on domain knowledge) and flexibility, ensuring that the data has substantial influence on the posterior distributions. The density plots confirm the expected shapes and spread of these priors, with the intercept exhibiting the widest distribution due to its higher standard deviation.

We used `tidyr` (Wickham, Vaughan, and Girlich 2023) to convert the data frame to long format.

### A.3.2 Diagnostics

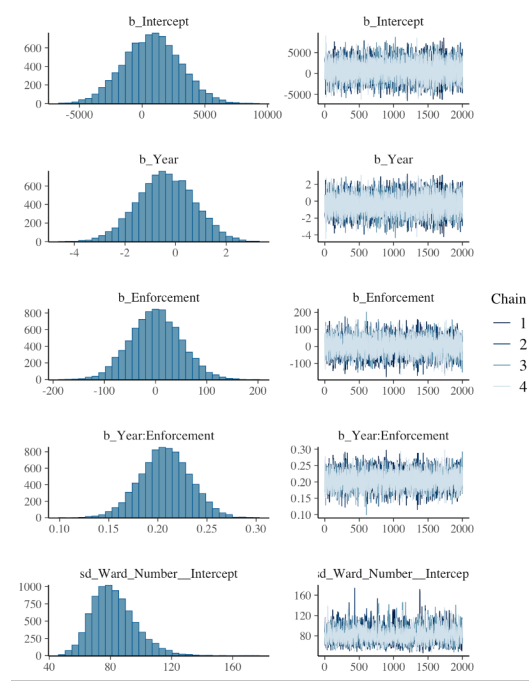


Figure 16: Diagnostic Plot (Posterior Density Plots and Trace Plots)

Figure 16 demonstrates stable and well-mixed posterior distributions for all model parameters, as evidenced by the smooth and unimodal density plots. The trace plots show good mixing across the four chains with no discernible trends or autocorrelation, indicating convergence of

the Markov Chain Monte Carlo (MCMC) sampling. These diagnostics confirm the reliability of the parameter estimates and support the validity of the model for further interpretation.

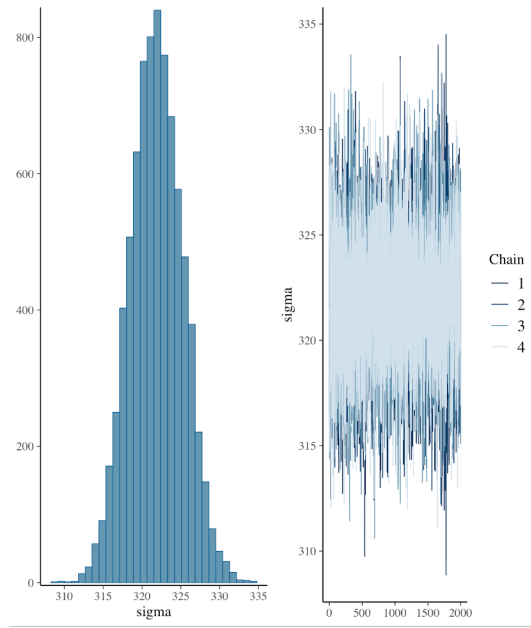


Figure 17: Diagnostic Plot (Posterior Density Plots and Trace Plots)

Figure 17 for the posterior distribution of  $\sigma$  highlights its stability and convergence. The density plot on the left shows a smooth, unimodal distribution concentrated between approximately 315 and 335, suggesting consistent sampling. The trace plot on the right confirms well-mixed chains without significant autocorrelation or trends, with all chains overlapping and exploring the parameter space effectively. These results indicate that the model's estimation of  $\sigma$  is reliable and robust.

### A.3.3 Model Validation and Performance

To evaluate the reliability and predictive capability of the model, we implemented a train-test split where 80% of the data was used for training the model and 20% was reserved for testing. This approach ensures that the model is evaluated on unseen data, reducing the risk of overfitting and providing a realistic assessment of its generalizability. The Root Mean Squared Error, calculated on the test set, was 426.51. This value provides a measure of the average prediction error, with lower values indicating better model performance. While the RMSE is reasonable given the range of fines, it suggests that there is room for improvement, potentially by including additional predictors or using more complex modeling techniques. The combination of this validation strategy and diagnostic checks as shown in Section A.3.2 strengthens the credibility of the model's results.

## References

- Alexander, Rohan, and Emily Alexander. 2020. *Telling Stories with Data*. Self-published. <https://tellingstorieswithdata.com/>.
- Bürkner, Paul-Christian. 2021. “Bayesian Item Response Modeling in R with brms and Stan.” *Journal of Statistical Software* 100 (5): 1–54. <https://doi.org/10.18637/jss.v100.i05>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Hammer, Ben, and Michael Frasco. 2018. *Metrics: Evaluation Metrics for Machine Learning*. <https://CRAN.R-project.org/package=Metrics>.
- Kuhn, and Max. 2008. “Building Predictive Models in r Using the Caret Package.” *Journal of Statistical Software* 28 (5): 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- 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, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Toronto Open Data, City of. 2024. “Red-Light Camera Annual Charges Dataset.” <https://open.toronto.ca/dataset/red-light-camera-annual-charges/>.
- Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing* 27: 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.
- 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](https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf).
- . 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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Yang, J. Y., M. Webster-Clark, J. L. Lund, R. S. Sandler, E. S. Dellon, and T. Stürmer. 2019. “Propensity Score Methods to Control for Confounding in Observational Cohort Studies: A Statistical Primer and Application to Endoscopy Research.” *Gastrointestinal Endoscopy* 90 (3): 360–69. <https://doi.org/10.1016/j.gie.2019.04.236>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.