The Relationship Between The Presence of Red Light Cameras and Traffic Incidents at Intersections in Toronto

Shon Verch and Leon Kalish

January 8, 2020

Abstract

This study examines whether there is a considerable relationship between red light cameras, specifically their presence and density, and the number of traffic incidents. To accurately determine the effect of red light camera, the city is split into an equally-sized grid and then each 'chunk' (sub unit of the grid) is analysed as a single unit. A weighted regression model ($R^2 = 0.671$ and $R = 0.825$, taken as an average across regression models developed using 6 years of data: 2014-2019) demonstrated that there is a positive relationship between the number of red light cameras and the number of traffic charges in Toronto. While more research is desired to identify stronger and more accurate trends (primarily that of conducted on larger samples sizes with more data available), there is reasonable evidence to suggest that red light cameras are effective as law enforcement devices; however, it is inconclusive whether red light cameras are effective in deterring drivers from red light running.

CONTENTS

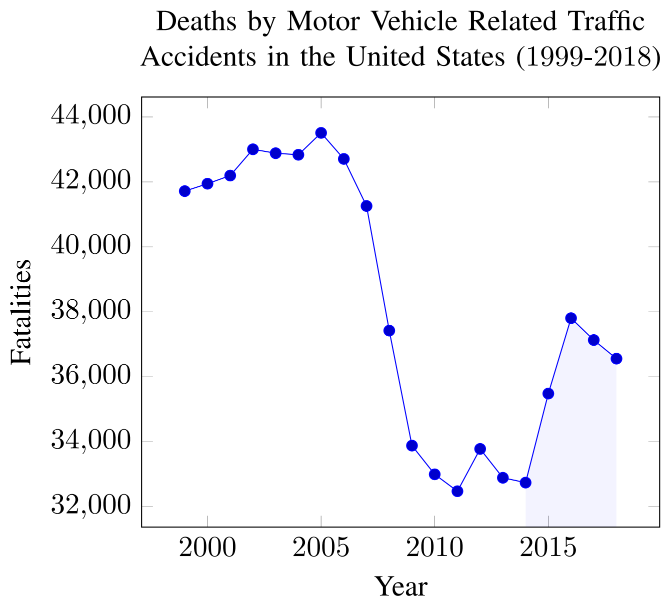

FIGURE 1.1. The number of deaths by motor vehicle traffic collisions in the United States from 1999 to 2018 [9]. Notice that after 2007 there was a rapid decline; however, since 2015, the trend has been reversing, with a rapid increase between 2015 and 2016; this trend is represented by the blue shaded area on the graph.

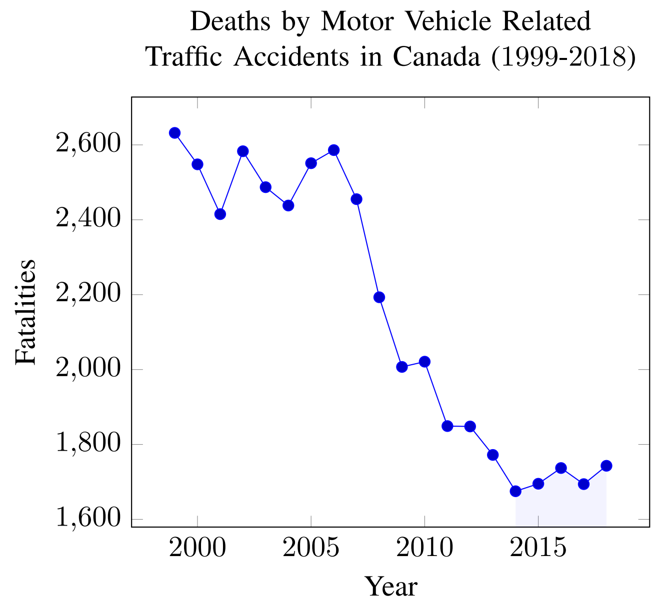

FIGURE 1.2. The number of deaths by motor vehicle traffic collisions in Canada from 1999 to 2018 [3]. Notice that after 2007 there was a rapid decline; however, since 2014, this trend has been slowly reversing; this trend is represented by the blue shaded area on the graph.

I. Introduction

In North America, deaths by motor vehicle traffic collision are one of the leading causes of death among individuals aged 5-34 [31]. In the United States, $36,560$ people were killed in a motor vehicle related accident in 2018 [28]. While the number of deaths and deaths compared to the overall population of the United States has decreased over the past two decades, the trend reversed in 2015 and has continued to rise in recent years. Figure 1.1 illustrates the deaths by motor vehicle related traffic accidents in the United States from 1999 to 2018. In a similar fashion, fatality rates as a result of motor vehicle traffic collisions have been increasing in Canada since 2015 as well (refer to Figure 1.2).

The quality of roads has major effects on crash risk as their condition can significantly impact how drivers perceive their environment; factors such as road surface, lane width, street markings, and flow can markedly influence the speed of a vehicle [5, 8]. And in traffic, these factors can severely bottleneck road systems causing increased congestion and crash risk. Poorly designed or maintained roads may further influence the frequency and severity of traffic accidents [18]. Intersections are most prone to these effects however, as they are, after all, the point at which two or more roads cross each other. This means that accidents which occur at intersections are not only affected by the condition of the road which the driver is on, but also, the conditions of all other contacting roads—and any driver action such as turning left, right, or crossing. While intersections only comprised of a small proportion of the total road network, accidents at these locations account for a significant percentage of total accidents. In 2015, approximately 5.3 million intersection-related crashes occurred in the United States, representing nearly half (47.24%) of all reported crashes and 28.30% of traffic fatalities [30].

Crashes at intersections have serious potential to induce harm on both drivers and pedestrians; nearly a third (31.05%) of all traffic signal crashes resulted in injury in 2017 [30]. Red light runners are a particular class of drivers responsible for injury and death in traffic signal collisions. Approximately 28.62% of traffic and pedestrian fatalities in 2017 occurred at intersections as a result of red light running [29]. Red light running occurs when a car enters an intersection during the red phase of the traffic signal; however, motorists who are inadvertently in an intersection when the red phase commences (such as when they are waiting to turn left) are not considered red light runners [26]. In general, red light running poses a serious threat to drivers, passengers, and pedestrians. In 2017, over

$10,000$ people were killed in intersection related accidents in the United States—of those, 809 were directly involved in collisions due to red light running [29]. Unsurprisingly, red light runners are more likely to be more reckless drivers or individuals with otherwise atypical driving behaviours. Among drivers who consumed alcohol, they were 54% more likely to run a red light than that of drivers who did not consume alcohol. Behaviour such as lack of seat belt use and driving with a suspended, revoked, or otherwise invalid license is more prevalent among red light runners [24]. Consequently, a high proportion of red light runners were found to have at least two traffic violations; 72% of those involved in fatal red light running collisions had a previous driving under the influence conviction [25]. This makes sense since reckless driving such as red light running is more likely to result in a revocation and/or suspension of license, and individuals who take less care in their driving (for example, seat belt use), are thus also more likely to take less care in waiting for the green phase of the traffic signal to commence.

Red light running violations among drivers is remarkably common-place, despite the serious danger that it poses for both the driver and others; approximately 56% of Americans acknowledge running a red light [25]. Retting et al. conducted a study at five busy intersections in Fairfax, Virginia, USA and found that, prior to the use of red light cameras, on average, a motorist ran a red light ever 20 minutes at each intersection [23]. A similar study which analysed the red light violation data from 19 intersections without red light cameras found that 3.2 motorists ran a red light per hour per intersection [10].

Several solutions have been proposed to address red light running—to decrease the risk of red light running for drivers and pedestrians whilst also including law enforcement such as ticket issuing. One proposed solution, red light cameras, have been used internationally since the early 1970s, and have been used in Toronto since the 1990s [22, 26]. A red light camera is a traffic enforcement device that automatically captures an image of a vehicle that has entered an intersection during the red phase of a traffic signal. This photo is then used by authorities as evidence; thus tremendously short cutting the otherwise tedious and often laborious judicial process—wherein police officers must collect evidence manually [26]. In Ontario, while red light violations do not issue demerit points to the plate holder, there is a fine associated with the violation which is billed to the plate holder—the idea being that these fines will discourage people from running red lights in the future [26]. Prior to the advent of traffic light cameras in Toronto, which occurred in the late 1990s, traffic violations related to running intersections chiefly required the presence of one or more police officers [26]. The police officer(s) on duty would chase offenders through the intersection. This is both a dangerous and ineffective practice. While traffic light cameras do not replace the police officer, they greatly reduce the risk of collisions as a result of law enforcement, and simplify the process of evidence collection. However, there is still debate among traffic administrators and law enforcement regarding their effectiveness: one side suggests that it acts as law enforcement—even when law enforcement is not present—while the other suggests that it's presence worries drivers and encourages them to act in unsafe ways to avoid a potential ticket [13].

A. Background

Early studies on red light cameras had mixed conclusions on the effectiveness of the devices. A study conducted in Sydney, Australia analysed 16 red light cameras for 2 years before and after enforcement. The results found a significant reduction in injury crashes, an insignificant decrease in total crashes, and an increase in right-angle and rear-end crashes. Hiller et al. concluded that red light cameras are only effective in locations where red light running is frequent [11]; however, a review of this study noted that the results may have been biased by warning signs being placed at all intersections with red light cameras [16]. As a result, it is hypothesized that the warning signs amplified the results in intersections with red light cameras—and thus exaggerating the data [16]. A similar study conducted in Adelaide, South Australia, examined the change in the total number of crashes at 8 red light camera locations and 14 control sites 5 years before and after implementation [14]. It was found that after the introduction of red light cameras, there was a slight reduction in the number of crashes compared to that of the control sites; however, the authors acknowledged that the red light cameras were installed at intersections with high-risk, meaning that the results of the study could have been influenced; the small sample size along with the lack of variance in samples sites likely limited the accuracy of the data as it was largely unable to capture the impact of small effects [14].

Several studies regarded the lack of effectiveness of red light cameras. One such work, conducted in Greensboro, North Carolina, USA, examined 303 intersections, 18 of which had a red light camera installed [2]. A regression model was developed—adjusting for externalities such as speed limit, traffic volume, and average annual daily traffic as just a few examples. The authors noted that sites with red light cameras were associated with a 42% increase in overall crashes, a 78% increase in rear-end crashes, and a 12% decrease in right-angle crashes [2]. Though, a review by Kyrychenko and Retting stated that this study contained several limitations that invalidated the conclusions [12]. One such limitation was the controls chosen by the author: intersections without the red light cameras were chosen within the same community, which may have impacted the effects of the study. News about red light cameras being installed in a community is likely to make drivers more aware and thus also more vigilant. As a result, the results of sites without a camera were likely to have produced underestimates of actual data. A similar study, though more expansive, examined the effect of red light cameras in 6 regions of Virginia, USA based on 7 years of data (January 1998 to December 2004) [7]. Garber et al. analysed 3500 crashes at 25 sites with red light cameras in comparison to 48 intersections without red light cameras [7]. Analytical models were then developed to quantify and draw conclusions from the data. The results were largely varied across different types of crashes and region. The authors found an increase in rear-end crashes and decrease in red light running crashes [7].

However, some studies did seem to have more positive results: one such study, done by Fisher and Gallager examined the intersections of 3 major Texan cities over a 12-year period to see if the implementation of red light cameras would create safer roads. The authors found that there was an increase in total accidents; however, a reduction in serious 't-bone' accidents. Thus, they concluded that while there was no evidence of a change in total accidents or injuries, there was a change in the composition of accidents [6]. While this study is insightful, it shows little in terms of effectiveness when aiming to reduce the number of red light camera tickets. Though, it does demonstrate that non-serious accidents *take the place* of more serious ones upon the installation of red light cameras—but does not necessarily affirm that people's behaviour changes over time in the presence of a red light camera.

Another study conducted by Martinez and Porter is topically similar to our research, however, focuses on deterrents that are not red light cameras—such as changing the timing of the traffic signal phases or redesigning the actual intersection to make the indicators of red lights more clear (i.e. clearing obscuring foliage or creating warning signs of upcoming red lights) [15]. The authors found that these factors alone decreased the number of red light camera tickets up to 17%—meaning that one way to deter people is not to threaten them with fines, but to rather make it easier for them not to violate the law in the first place [15].

Research on red light cameras, and traffic intervention methods in general does not appear to be sparse. Despite of this fact however, there is still much to be desired in terms of concrete results. There is still no consensus on the effectiveness of red light cameras, which is enigmatic given their age. One prevailing issue with many studies on the topic, as discussing earlier, is not the analysis but rather the methodology and/or data—which rings especially true for studies which collected their own data over a period of several years. On the other hand, studies which analysed varied and multifaceted data such as Garber et al. [7] found more success. This means that there is still ripe potential for further research into red light cameras and their effectiveness as red light running intervention methods.

B. *Motivation*

This research area is particularly important because it concerns the safety of our roads—a topic of immense interest due to the severity of Toronto's winters and the ever growing population of the city. Notions of road safety have influenced many major streets and projects, especially the King Street Pilot Project that has shut down car lanes all together, demonstrating how safety and efficiency is the priority of the city. Finding a correlation and an explanation for our correlation will help provide insight on whether having red light cameras make an intersection safer—and whether we should continue allocating our budget towards these programs.

The goal of this research is to better understand the effect of red light cameras on Toronto's intersections. In particular, we analyse the presence of red light cameras and their relationship to the number of traffic violations (in the form of charges published by the municipal government) over the course of up to 12 years (if the data is available); as well as determining if the threat of a ticket is enough of a deterrence to drivers. Using a statistical model developed from geographical data, we can be better represent how people in different areas of Toronto drive in signalized intersections with the installation of red light cameras.

C. Research Questions and Hypotheses

Based on the results of previous research and similar studies, there is reasonable evidence to suggest that red light cameras impact the crash rate at signalized intersections; however, there is still much to be desired. With motor vehicle fatality rates only growing in recent, the importance of safe road practices and traffic control technology cannot be understated. Research into these technologies, such as red light cameras, is extremely important for administrators, legislators, and lawmakers alike so that they can better allocate resources and budgets into only the most promising solutions.

The main research question of this paper is whether there exists a relationship between the number of red light cameras and the number of charges laid. Does the density of red light cameras affect their utility? Do they deter people from red light running? We hypothesize that in general there is a positive relationship between.the number of red light cameras and number of tickets laid—that an increase in red light cameras will also follow in an increase in the number of traffic ticket charges. It is important to note that we are not measuring the number of violations but rather, we are measuring how many charges occur as a measure of the devices effectiveness as a law enforcement device. We also hypothesize that, over time, the drivers in the area will learn which intersections have red light cameras, thereby decreasing the aggregate number of tickets laid—due to an overall increase in awareness of red light cameras.

II. METHODOLOGY

We first describe the search strategy and collection of the datasets which are used for our analysis—focused on factors that may affect red light camera effectiveness, along with general statistics regarding intersections in Toronto. Then, we introduce how we cleanse the data so that it can be used for analysis. Finally, we present two techniques for analysing the intersection data from a geographical interpretation.

A. Search Strategy

During December 2019, datasets and pertaining to red light cameras, traffic intervention methods, signalized intersections, and motor vehicle collisions were retrieved from the *Toronto Open Data Portal* [4], *Toronto Police Services Public Safety Data Portal* [27], and *Statistics Canada* [17]. The following free-text queries were used in the "All fields" category to search for relevant datasets, organized by database:

- **Toronto Open Data Portal (TO)**: `traffic, traffic camera, red light camera, red light running, camera, intersections, congestion, volume, pedestrian volume`; ordered by relevancy.
- **Toronto Police Services Open Data Portal (TPS)**: `traffic, traffic camera, red light, red light camera, camera, crash, vehicle, pedestrian, collisions, traffic tickets`; ordered by relevancy.
- **National Collision Database (NCDB)**: `collision, traffic, traffic ticket, red light running`; organized by relevancy.

In the case of international data collection, such as data regarding the United States of America, datasets were gathered from their respective administration portals. In specific, the *National Center for Statistics and Analysis (NCSA) Motor Vehicle Traffic Crash Data* was used to retrieve NCSA publications and statistics [1]–using the following free-text queries in the "All fields" category: traffic safety facts, motor vehicle traffic fatalities, fatal crashes, non-fatal crashes, motor vehicle traffic crashes, red light running; organized by relevancy.

B. Data Collection

In December 2019, the following datasets were retrieved using the search terms outlined in Section 2-A, organized by database and category: "Red Light Cameras", "Charges Laid by Location and Year", "Traffic Signal Vehicle and Pedestrian Volumes", "Intersection File - City of Toronto", and "Automobile". Figure 2.1 provides a more in-depth summary of all datasets used along with their source database and other relevant metadata.

One major issue with the datasets used is that the street names are not standardised; that is, while the names themselves are *technically* the same, the row data is not identical in terms of the string data. Therefore, it becomes very difficult to programmatically cross-reference the datasets in their original form. Due to the decimal inaccuracies of geographical coordinates systems (such as the provided longitude and latitude), we are required to use street names as keys to cross-reference between datasets. As a result, we first cleanse all datasets involving street names by writing a small Python program to convert columns containing street names into a standard format: all names are converted

Dataset Usage Summary			
Source	**Name**	**Description**	**Format**
TO	Red Light Cameras	This dataset identifies the intersections in Toronto where red light cameras are located.	`GeoJSON` in `WGS84` projection.
TO	Charges Laid by Location and Year	This dataset identifies the charges at intersections in Toronto by year.	`CSV`
TO	Traffic Signal Vehicle and Pedestrian Volumes	This dataset contains the most recent 8 peak hour vehicle and pedestrian volume counts collected at intersections where there are traffic signals.	`XLXS`
TO	Intersection File - City of Toronto	A geospatial file containing all of the street intersections within the City of Toronto.	`SHP`
TPS	Automobile	This dataset is a subset of the Killed and Seriously Injured (KSI) dataset from 2007-2018. It includes all in traffic collisions events where an occupant of an Automobile is involved.	`CSV`
NCDB	Motor Vehicle Collision Statistics	This dataset contains motor vehicle collison statistics for the following variables: "Year", "Collision Conf", "Persons-Coll.", "Vehicles-Coll.", "Pedestrians", "People in Veh.", and "Injury Severity".	`CSV`

FIGURE 2.1. An in-depth breakdown of the datasets used. The sources are given us abbreviations for layout purposes; "TO" refers to the Toronto Open Data Portal, "TPS" refers to the Toronto Police Service Public Safety Data Portal, and "NCDB" refers to the National Collision Database.

into uppercase (i.e. `Steeles Ave` becomes `STEELES AVE`), all abbreviations are expanded (i.e. `YONGE ST` becomes `YONGE STREET`, `SHEPPARD AVE W` becomes `SHEPPARD AVENUE WEST`, and `HWY 27` becomes `HIGHWAY 27`), and only alphanumeric characters are allowed—all other characters are omitted and converted to spaces; this includes removing any whitespace from either end of the string (which was a frequent issue with the names in the "Traffic Signal Vehicle and Pedestrian Volumes" dataset from the Toronto Open Data).

Cross-referencing the data is done using a Python script which iterates through the source dataset, finds the relevant intersection via street name, and then appends the relevant data to the destination dataset.

In order to map the geographic data (such as the latitude and longitude coordinates in the "Red Light Cameras" dataset and "Automobile" crash dataset), we first need the road network data as a shapefile for the relevant regions. We extract the data from the *OpenStreetMap* (OSM) dataset [19], an open-source world map database, using the *Overpass API* and *QGIS* tool [20]. To only extract the desired features from OSM, an Overpass query which filters for only roads of type `motorway`, `primary`, `secondary`, and `tertiary` is used. This query is executed using the *QGIS* tool with the *QuickOSM* plugin [21] and each road feature is exported separately in individual shapefiles.

C. Data Analysis

Before the data can be analysed with a statistical model, it must be quantified into suitable measures which can be correlated with one another. Analysing the number of red light cameras at a specific intersection is not very useful, at least in terms of a two variable analysis, as most intersections have only one or, at most, two, red light cameras—meaning that it is unfeasible to analyze the red light cameras on an individual basis. We instead propose a technique which splits the geographic data into a grid. If we split the road system into equally sized regions (called 'chunks'), and from there, compare the cumulative red light cameras to the cumulative charges in each region, we can approach a good estimate for the overall effect of red light cameras on the traffic charges.

A simple spatial partitioning algorithm is used to compute the position of the chunk for any given coordinates of the region. For some set of points P, we define $G(P, \vec{D})$ as the spatial partition of P such that there are $D_x \times D_y$ chunks, where $\vec{D} \in \mathbb{N}^2$ is a vector whose components describe the size, in terms of chunks, of the spatial partition (i.e. D_x is the number of chunks horizontally and D_y is the number of chunks vertically). The spatial partition has the property that it contains all points P, regardless of the choice of $\vec{D}$. We compute $G(P, \vec{D})$ by iterating through each point of P, suppose u, and finding the two-dimensional index of the chunk which contains u; the index of a chunk is the zero-based row and column of the grid square (starting form the top-left)—refer to Figure 2.2. Given a point u contained in the chunk and such that where $u \in P$, the index of the chunk, (i, j), is given by

$$(i, j) = \left(\left\lfloor \frac{u_x - \min(P_x)}{\Delta x} \right\rfloor, \left\lfloor \frac{\max(P_y) - u_y}{\Delta y} \right\rfloor \right),$$

where $\min(P_x)$ is the minimum longitude and $\max(P_y)$ is the maximum latitude[1] and Δx and Δy represent the width and height of a single chunk respectively.

While the technique of spatially partitioning the geodata helps prepare the data for use in a regression model, it also decreases the quality of the model—and also makes it more prone to amplify small errors in the data. For this reason, it is paramount to consider the resolution of the spatial partition, $\vec{D}$; if the components of $\vec{D}$ are too small then the chunks encompass *too much* of the data and there is not enough separation. On the other hand, if the components of $\vec{D}$ are too large then no useful trends are preserved in the dataset since as $\vec{D}$ tends to infinity, each chunk will contain less points on average—at some point, for some value of $\vec{D}$, a chunk will either contain 0 or 1 data points which negates the initial purpose of the spatial partition. Therefore, it is important to choose a good middle-ground value which encompass *just enough* data points for a good regression. This value can be found analytically by maximizing the function which gives the average number of points in any chunk; however, for most geodata, experimentation is often a far quicker method for determining an approximation to the ideal dimension.

A Python implementation of the spatial partitioning algorithm detailed above is used to compute $G(P, \vec{D})$, generate the geographic visualizations, and to generate the processed data which we are used int the regression analysis.

The regression analysis is completed using both a simple unweighted regression to the mean and also a variance-based regression. Our goal in using the weighted model is to help compensate for measurement variance and excessive data concentration in certain regions, compared to that of other regions. The weighted model aims to control for the inaccuracies that are accompanied by the spatial partitioning technique.

[1]Note that while we refer to the horizontal and vertical coordinates as x and y, they are *not* referring to a Cartesian coordinate system but rather a geographical coordinate system. The x and y coordinates refer to the longitude and latitude respectively. Since we are dealing with very small regions, we can treat these coordinates, and thereby the grid, as two-dimensional projections.

$(0, 0)$	$(1, 0)$	$(2, 0)$	$(3, 0)$
$(0, 1)$	$(1, 1)$	$(2, 1)$	$(3, 1)$
$(0, 2)$	$(1, 2)$	$(2, 2)$	$(3, 2)$
$(0, 3)$	$(1, 3)$	$(2, 3)$	$(3, 3)$

FIGURE 2.2. The layout of the chunks, with their two-dimensional index, in the spatial partitioning. The chunk indices start at the top-left corner and count from left to right, top to bottom.

III. RESULTS

Descriptive statistics for the traffic charges are provided in Figure 3.1. Notice that average number of charges issued increases with time. Also notice that the standard deviation increases with time. These increases are particularly noticeable between 2007 and 2008 where the mean increased by 2924.73% and the standard deviation increased by 1098.77%. On the other hand, the minimum charges has stayed relatively constant over time; however, the maximum number of charges has been steadily increasing; Figures 3.2 and 3.3 illustrate the descriptive statistics from 2007 to 2019.

Figure 3.4 describes the descriptive statistics for the 8 peak hour traffic volume by variable (pedestrian and vehicle). Despite a 440% change between the minimum and maximum vehicle traffic, the number of traffic charges is relatively stagnant per year; there is only a 65% change in the number of traffic charges in 2018 at the intersections. Figure 3.5 shows the number of traffic charges from 2017 (the common year that traffic charge data is first available) for the to 2019 for both

Summary of Traffic Charges by Year				
Year	**Mean ($\bar{x}$)**	**SD (s)**	**Min**	**Max**
2007	5.77	28.37	0	222
2008	171.20	340.09	0	1467
2009	307.15	398.23	0	1962
2010	447.80	438.29	0	2291
2011	368.24	374.80	0	1966
2012	464.31	433.13	0	2210
2013	422.15	362.49	0	1589
2014	362.00	308.72	3	1093
2015	388.83	299.89	16	1163
2016	476.27	453.53	16	2662
2017	421.68	645.11	0	3994
2018	606.42	919.21	0	6615
2019	286.45	518.35	0	4480

FIGURE 3.1. A summary of the descriptive statistics for the traffic charges from 2007 to 2019, from the "Charges by year and location" dataset provided by the Toronto Open Data. The standard deviation is referred to as 'SD' in the table. The values are rounded to two decimal places where necessary. Notice that the mean, standard deviation, and max roughly increase with time, until 2019.

the intersections with minimum and maximum vehicle traffic (Simcoe St./Wellington St. W and Sheppard Ave E./Bayview Ave. respectively).

The spatial partition of the red light camera data is illustrated in Figure 3.6 in the form a geospatial map spanning from Highway 427 to Scarborough. The data extracted from the spatial partition is used in our regression analysis.

A. Regression Analysis

We developed two regression models to represent the relationship between the number of red light cameras to the number of traffic charges. The first model performs an ordinary linear regression on the data; however, the issue with an ordinary linear regression is that it treats each chunk as an aggregate sum of all red light cameras in the region, with no further adjustment for external factors. Consider Figure 3.6 which depicts the geospatial partition of the red light data. In particular, notice that highly populated regions like downtown Toronto contain far more red light cameras than the regions to the north—which contain only a fraction of the red light cameras in downtown Toronto. This discrepancy of course, is not translated in the ordinary linear regression—despite the fact the each chunk has variance in its volume, population, density, they are still treated as identical units. This means that attempting a regression analysis using an ordinary linear regression model yields potentially inaccurate estimates—because there is precisely very little data to go by; while a trend is certainly evident, it is biased. Our technique to workaround the limitation of ordinary linear regression is to employ a *weighted* linear regression, which minimizes the squared errors between the approximation ($\hat{y}$) and the original data (y) with respect to the weight of each data point (w).

While looking at the number of red light cameras to the number of traffic charges in any given chunk can be helpful, one external factor is how close apart the red light cameras are—their density. Suppose that in some chunk, all red light cameras are in the same intersection, meaning that the red light cameras are very dense. In the ordinary linear regression model, each point in the region is treated equally, which might mean that the true value of the whole intersection is over counted multiple times. In reality however, since the cameras are dense, they do not cover a lot of area, implying that they are less effective compared to a chunk with the same number of cameras but whose positioning is less dense. Therefore, our density-based

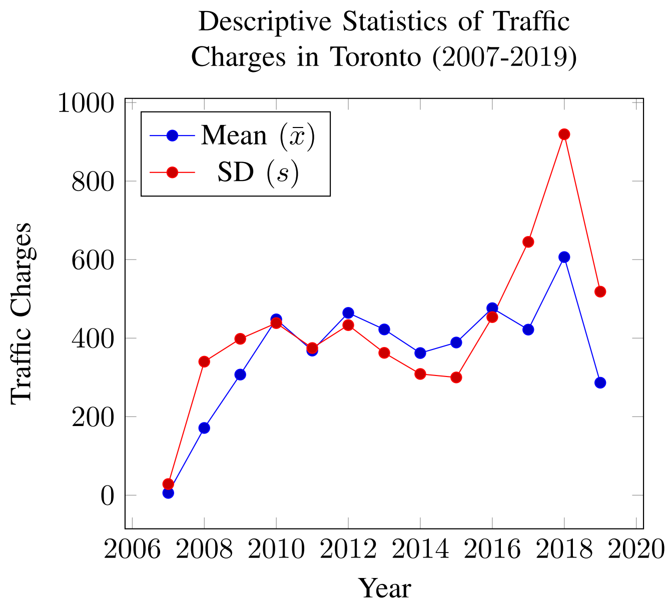

FIGURE 3.2. The average number and standard deviation of traffic charges from 2007 to 2019. Notice that both the mean and standard deviation (referred to as 'SD' in the plot) have increased steadily over time; however, in 2019, there was a sudden decline of 52.76% in the mean and a 43.61% decrease in the standard deviation. This means that until 2019, more traffic violations were being caught and enforced on average, and that the number of charges became more spread out between intersections; however, in 2019, the number of charges between intersections became more consistent, though average number of charges also decreased.

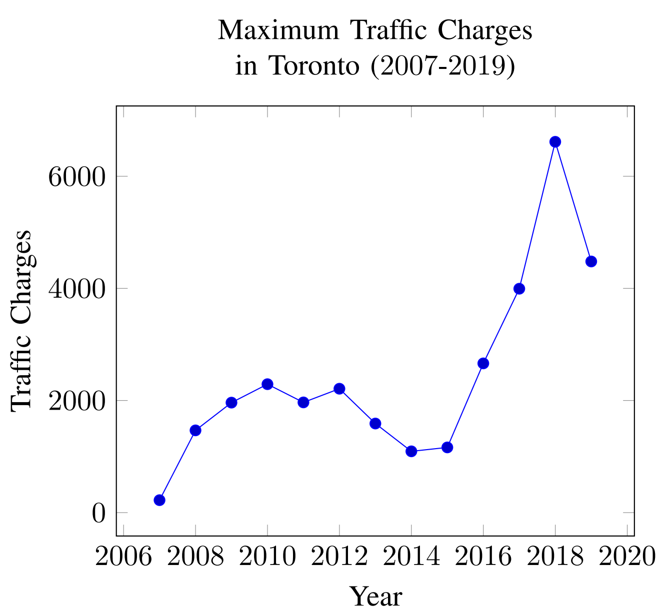

FIGURE 3.3. The maximum number of traffic charges in Toronto from 2007 to 2019. Notice that the maximum rapidly increases between 2007 and 2008, and then continues to steadily increase over time until 2019 where it drops by 32.28%.

weight model measures the density of the red light cameras in any given chunk and uses that as the weight of each chunk in the regression. The density is measured as an inverse ratio of the area of the convex hull of all the red light camera positions to the total area of the chunk. A convex hull of the points P is the smallest convex polygon which contains all points in P. The density, d, of the chunk containing the points R is given by

$$d = 1 - \frac{A_C(R)}{A_T}, \tag{3.1}$$

where $A_C(R)$ is the area of the convex hull formed by the points, R, and A_T is the total area of the chunk. In other words, the density is a representation of how much area the red light cameras span—though not necessarily how much they effectively monitor/cover. If the area of the convex hull of the red light cameras is very *close* to the area of the chunk, then they are *not* dense and thus d is a small value; on the other hand, if the area of the convex hull covers a very *small* proportion of the total chunk, then the cameras are very dense, and thus the value of d is high.

The density-based weight model weights chunks proportional to the reciprocal of their square density, which is given as

$$w_i = \frac{1}{d_i^2},$$

where w_i is the weight of the i-th chunk and d_i is the density of the i-th chunk.

Figures 3.8-3.13 illustrate the ordinary linear regression models for the last 6 years (2014-2019). Figure 3.7 contains a summary of all the R and R^2 values for both the unweighted and weighted regression models for the last 6 years of data. The weighted regression models more accurately model the data in all 6 years of regression model. In specific, the density-based weight function is able to more closely model the correlation, along with better minimize the square errors of the residuals, due to its assignment of weights based on density.

IV. DISCUSSION

The purpose of this paper was to examine the relationship between red light cameras and traffic charges in Toronto intersections. The results indicate that there is indeed a positive relationship between number of red light cameras and the number of traffic charges, confirming the first hypothesis (outlined in Section 1-C).

Summary of Traffic Volume (8 Peak Hour)				
Variable	**Mean ($\bar{x}$)**	**SD (s)**	**Min**	**Max**
Pedestrian	2176.73	2953.20	70	18152
Vehicle	22803.18	7506.65	9034	48835
Variable	**Intersection Name**			**Traffic Charges (2019)**
Min Pedestrian	Steeles Ave. W and Highway 27			55
Max Pedestrian	University Ave. and Wellington St. W			2201
Min Vehicle	Simcoe St. and Wellington St. W			407
Max Vehicle	Sheppard Ave. E and Bayview Ave.			246

FIGURE 3.4. A summary of the descriptive statistics for the 8 peak hour traffic volume by variable (pedestrian and vehicle) from the "Traffic Signal Vehicle and Pedestrian Volumes" dataset provided by the Toronto Open Data. The standard deviation is referred to as 'SD' in the table. The values are rounded to two decimal places where necessary. The names and most recent traffic charge data of the intersections with the maximum and minimum value for each variable is also provided.

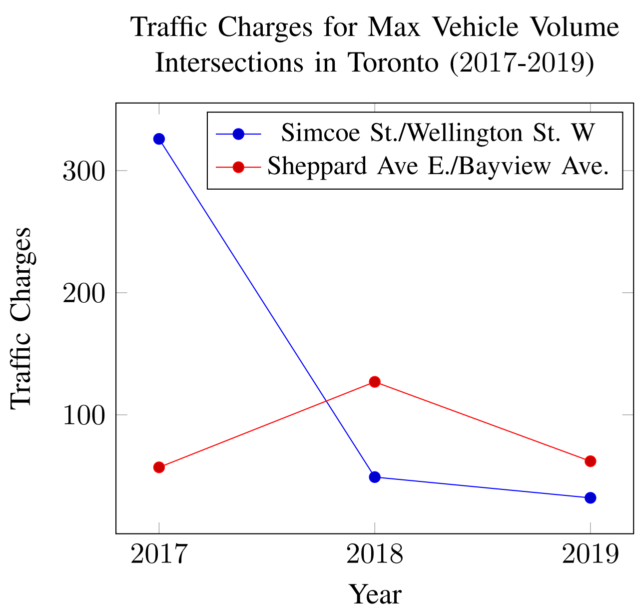

FIGURE 3.5. The number of traffic charges from 2017 to 2019 for both the intersections with minimum and maximum vehicle traffic (Simcoe St./Wellington St. W and Sheppard Ave E./Bayview Ave. respectively).

A. *Outliers and Bias*

Overall, in terms of the red light camera and traffic charges data, there were minimal outliers present. However, in the initial hypothesis, we believed that drivers would become more accustomed to the presence of red light cameras, and as such, the total charges per year would decrease steadily. But, the traffic charge data indicates an opposite trend (refer to Figure 3.14 for historical data—red light camera vs. traffic charges—for the chunk containing downtown Toronto). In 2014, 2015 and 2016, the number of traffic charges approached nearly 5000 in densest region of Toronto (the downtown core), meaning that 2016 had an overall larger influence on the trend than that of 2015. Similarly, in 2017 the number of charges in the densest chunk spiked to 12500, which was nearly triple the amounts between 2014 and 2016. The following year spiked to 17000 charges; however, we do not consider this an outlier because it is both within a tolerable range relative to the historical data and since it follows a similar trend found in many other intersections and chunks: that the data is slightly volatile from year-to-year; we see spikes in certain years and declines in others. This has to do with the relatively unpredictable nature of the variable.

In Section 2-C, we presented an algorithm for spatially partitioning the geodata into chunks which can then be used in a regression analysis. This is a necessary step because on their own, the effect of red light cameras on traffic charges cannot be analysed locally (relative to a single intersection)—traffic law enforcement, and more importantly red light running, have an effect on the neighbourhood of intersections around any given point. Road networks are not isolated units, and a change in one part of the system, can have drastic effects on another separate part of the system. In other words, to analyse trend of traffic charges at one intersection, we have to analyse the trend of charges of all the intersections in the neighbourhood. This is achieved by measuring the trend of traffic charges, relative to the number of red light cameras, in a given *region* of the road network; however, this is only an approximation—

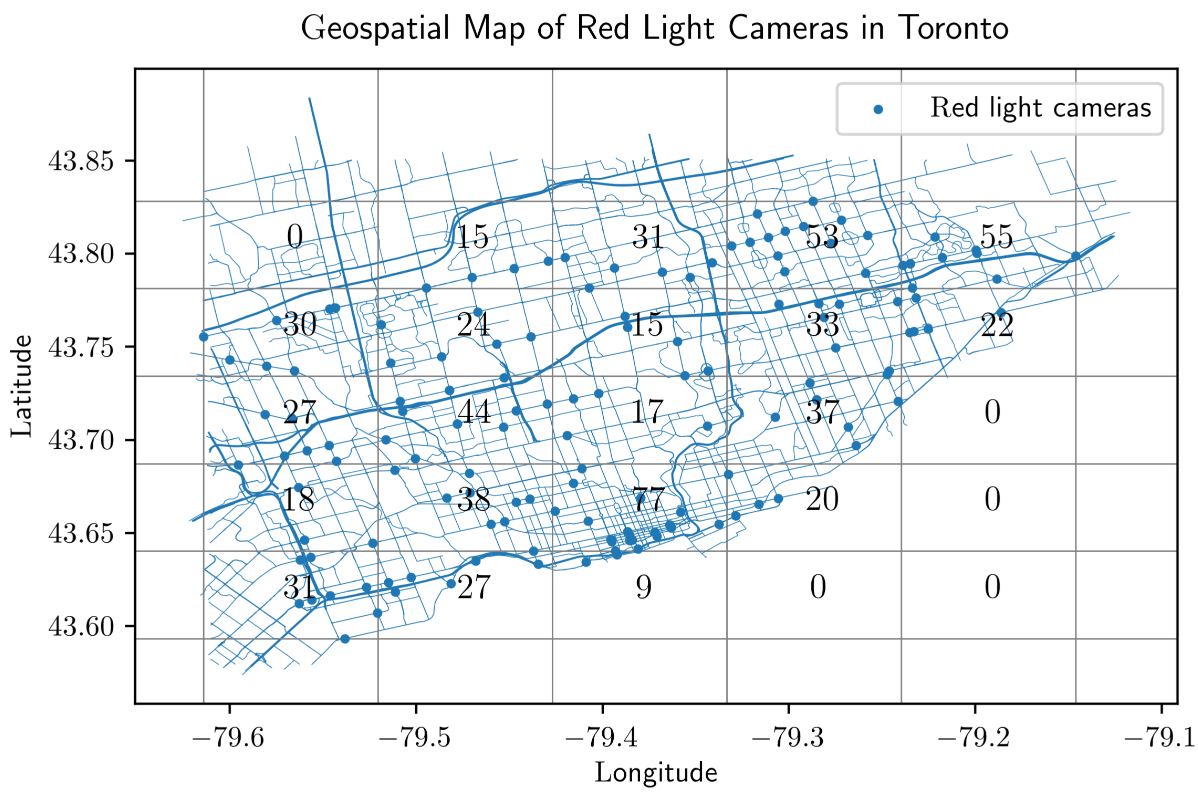

FIGURE 3.6. The geospatial map (computed using the algorithm outlined in Section 2-C) of the red light cameras, with the number of red light cameras in each chunk overlaid on top of the geospatial map. The dimensions of the spatial partition, $\vec{D}$, is 5×5 chunks. Notice that downtown Toronto, which is contained in the chunk $(2, 3)$, is the most concentrated in terms of red light cameras.

Summary of Regression Models

Year	**Unweighted R^2**	**Unweighted R**	**Weighted R^2**	**Weighted R**
2014	0.537	0.733	0.553	0.753
2015	0.816	0.904	0.850	0.924
2016	0.714	0.845	0.767	0.878
2017	0.608	0.780	0.649	0.822
2018	0.587	0.766	0.616	0.795
2019	0.563	0.751	0.589	0.778
Average	0.638	0.797	0.671	0.825

FIGURE 3.7. A summary of the unweighted and weighted coefficient of determination (R^2) values and coefficient of correlation (R) values for the last 6 years of data.

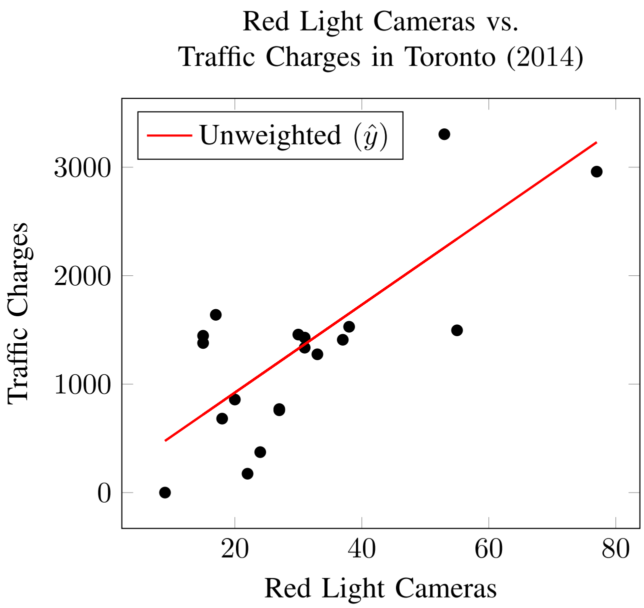

FIGURE 3.8. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2014. The line $\hat{y}$ represents an unweighted regression ($R = 0.733$ and $R^2 = 0.537$).

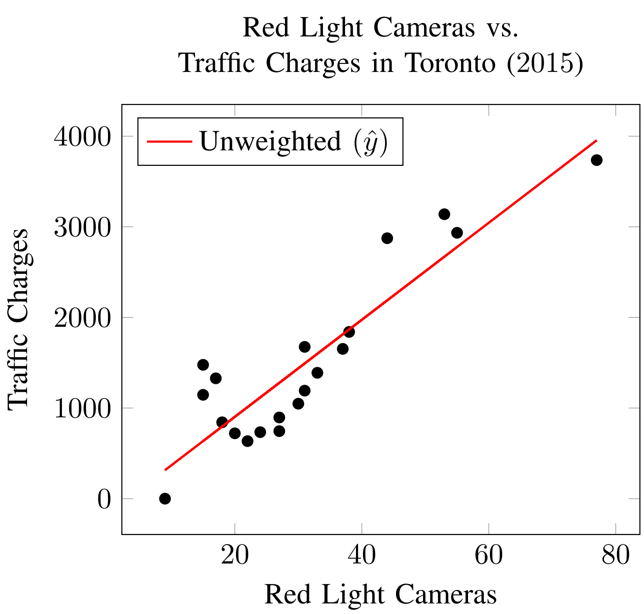

FIGURE 3.9. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2015. The line $\hat{y}$ represents an unweighted regression ($R = 0.904$ and $R^2 = 0.816$).

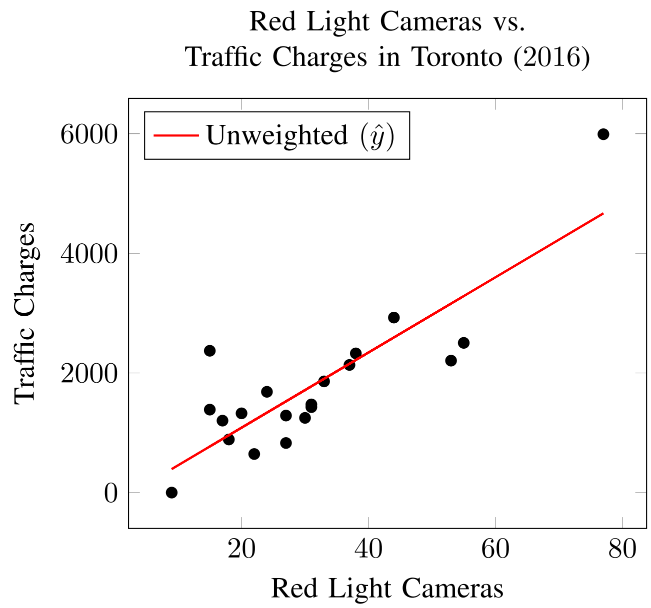

FIGURE 3.10. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2016. The line $\hat{y}$ represents an unweighted regression ($R = 0.845$ and $R^2 = 0.714$).

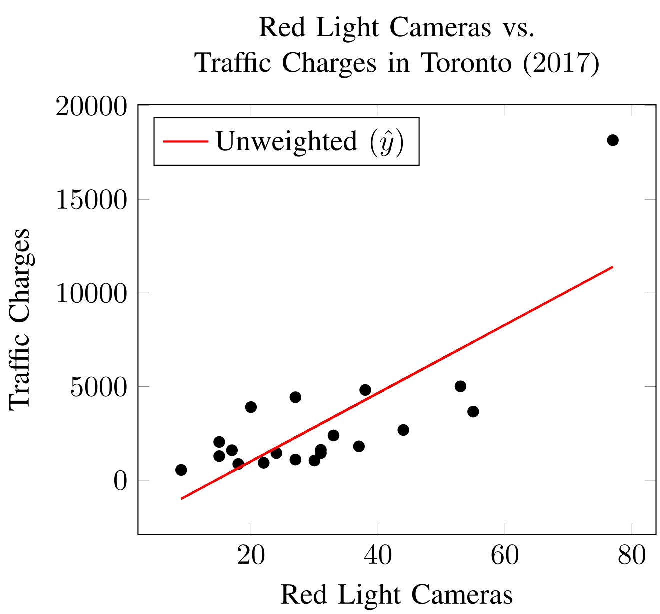

FIGURE 3.11. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2017. The line $\hat{y}$ represents an unweighted regression ($R = 0.780$ and $R^2 = 0.608$).

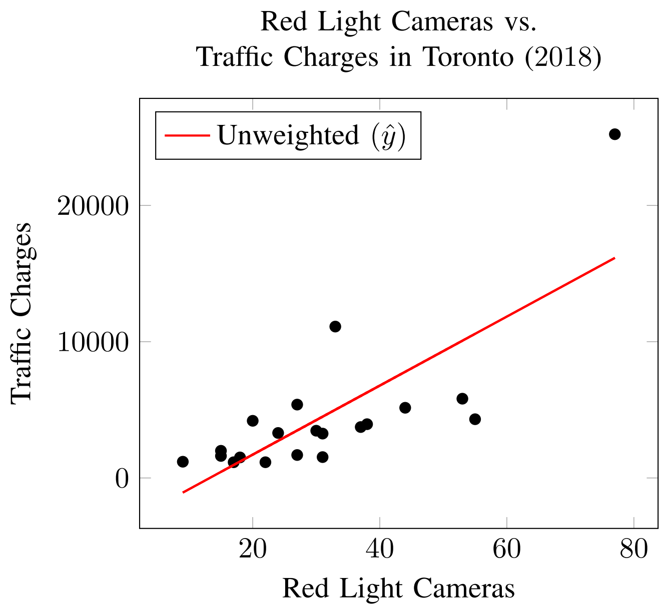

FIGURE 3.12. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2018. The line $\hat{y}$ represents an unweighted regression ($R = 0.766$ and $R^2 = 0.587$).

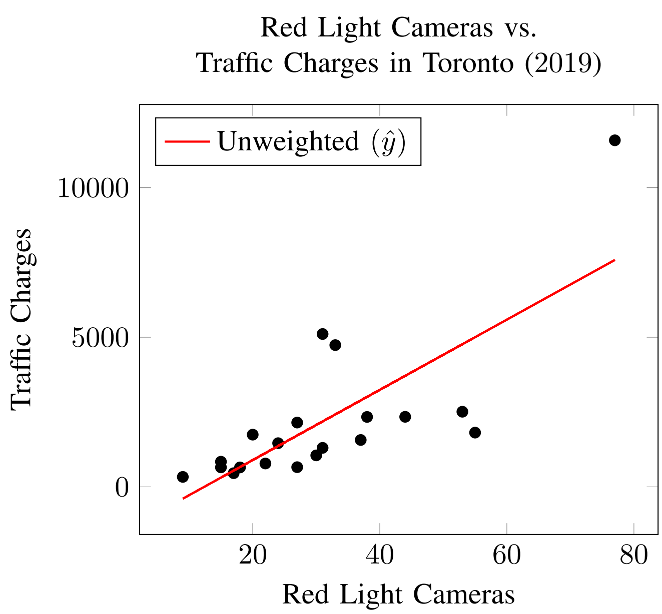

FIGURE 3.13. The regression plot for number of red light cameras to number of traffic charges in Toronto for the year 2019. The line $\hat{y}$ represents an unweighted regression ($R = 0.751$ and $R^2 = 0.563$).

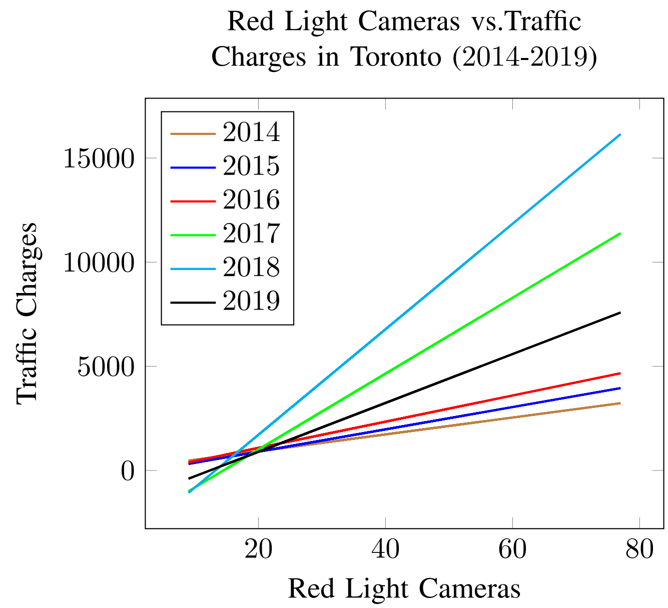

FIGURE 3.14. The unweighted regression plots for the number of red light cameras to number of traffic charges in Toronto from 2014 – 2019.

since we are effectively reducing the size of the sample population by a factor of the partition's dimensions. As a result, the chunks, while acting as a good approximation, might also overestimate in very dense regions (i.e. downtown Toronto where there are many red light cameras in very close proximity to one another). This is because in regions with dense red light cameras, more of them can fit inside the same chunk. This is addressed by increasing the number of chunks in the partition; however, this has the disadvantage of 'slicing the data thinner,' which can decrease its accuracy. In any case, no matter the dimension of the partition, there will always be some sort of bias created by outliers in the dataset. We minimize this bias by finding the optimal dimensions, but we can never remove it.

Likewise, due to the distinctly binary nature of red light camera tickets (either drivers get ticketed or not), there is little room for judgement and therefore little room for bias. But, some unintentional biases may still occur when analyzing the data, mostly due to the lack of standardization between the data sets. For example, we noted in Section 2-C that the datasets sometimes contained non-standardised naming conventions. While we did normalize the data by adopting a consistent set of rules, it is still possible that we missed data points (as we did not enumerate through each row of the dataset manually). Furthermore, from an external standpoint, the Toronto Open Data is aggregated from various sources, some international, meaning that there is potential for non-standard naming to 'corrupt' the data as well. For this reason, there are cases where intersections, which are physically the same, appear as two duplicates in the dataset due to conflicts (i.e. "Sheppard Avenue E" and "Sheppard Ave. E" appear as different rows in the dataset despite both describing the same intersection). Another issue we noted with our data was the lack of data in certain years. Either these rows were omitted or simply unavailable; however, in either case, it means that we cannot draw as accurate conclusions because there is simply not enough historical data. For example, some red light cameras were installed in 2007 and have been operating for 12 years, while a handful were only installed in 2016, which means that there are only 3 years worth of data from which we can extrapolate. Ultimately, it is difficult to arrive at meaningful generalizations about the effectively of red light cameras without knowing the conditions under which the driver ran the red light—for then one could target specifically the points that are weakest. For example, ticketing drivers that drove past a late amber as a safety precaution rather than a reckless action should not be treated as equal instances because the primary goal of red light cameras is to deter reckless drivers, not to punish the safe ones.

B. Casual Relationships

If we sort the data by alternative metrics (i.e. congestion and volume levels), we see that certain trends begin to emerge. The intersection with the most recorded vehicles in 8 hours over the entire history of volume trackers is Sheppard Ave E./Bayview Ave. (refer to Figure 3.4 for a complete summary of the descriptive statistics of the traffic volume data). When relating it back to the number of red light tickets, an interesting detail emerges where the number of tickets given is not proportional to the volume of the intersection itself. The number of traffic charges stay within the range of 57 to 127 charges for the entire year. An operating hypothesis was formed earlier where it was posited that the larger the street/intersection is, the more tickets will be laid; however, based on our data, this does not appear to be true because Sheppard Ave. E./Bayview Ave. has arguably the highest volume daily but only a few charges laid throughout the year. On the other end, the intersection with the least amount of vehicle traffic was Simcoe St./Wellington St. W. which received 3 times more tickets in 2017 and about half as many as Sheppard Ave. E./Bayview Ave. in the following 2 years (refer to Figure 3.5). This can most likely be attributed to the fact that a more congested street has reduced speeds, leaving more time for drivers to slow down and stop at a red light. Similar logic can be applied to smaller streets. Though, at some distinct point, the amount of cars daily reduces the overall tickets, leading us to the conclusion that the number of red light tickets to daily volume is representable as a quasi-bell curve, where at one golden spot there is next to no congestion, but a high volume of cars throughout the year.

Furthermore, based on Figure 3.3, the maximum number of charges given per year, as well as the mean amount of tickets given per year (Figure 3.2), are steadily increasing, growing increasingly further from the standard deviation. Contrary to the initial hypothesis of the red light tickets decreasing over time as drivers learn to drive safer, the red light tickets actually begin to increase. However, when examining the total amount of serious accidents (fatal or seriously injured), it becomes obvious that these accidents are decreasing, indicating that drivers are most likely opting to drive through the red light (or late amber light) to avoid accidents rather than just pure recklessness. However, it might also suggest that indeed the habits of drivers are changing—just not as we predicted; rather than learning that there are red light cameras and then driving very carefully through those intersections, it could be that drivers are simply changing their route, to avoid those intersections at all.

C. Reflections

The data collected is endorsed by both the Toronto Police Service and the City of Toronto government through the Toronto Open Data platform, and is continually updated as the years progress. The addendum for the red light camera database states that the cameras are used as a complimentary service to the monitoring police officers, with all the pictures taken being manually reviewed before mailing out tickets. This two-step verification systems ensures that the data is entirely correct and more importantly fair to drivers. It is also mentioned that a red light camera system (1 camera) costs $\$150,000$ to install meaning that the installation is no meager investment for the city, so the upkeep is important and its effectiveness is integral to the safety of the city. The statistics that were calculated were cross-referenced across the multiple databases available, ensuring that if a measure of central tendency was calculated, it was consistent among the numerous datasets. This clued us in on whether the data itself was consistent among the datasets and whether our calculations were accurate. Cross-referencing the data set allows our conclusions to become somewhat founded and allows us to come closer to a true conclusion (rather than a flawed method of thinking because of inconclusive data). A much longer study with more primary data would allow us to approach a true answer to the necessity of red light cameras and if the cost is warranted.

D. Further Research

This study largely succeeded in finding a correlation between the number of red light cameras and the number of traffic charges in a region of intersections; however, we did not take into account many of the possible conditions that drivers could face. While we aimed to minimize bias by also tracking traffic volume and collisions, further research could investigate the reasons for the traffic charges. Furthermore, one missed opportunity was not looking at the trends of historical traffic volume over time, which would have indicated whether drivers were changing their routes as a result of red light camera installation. If the traffic volume after the installation of a red light camera decreased, this might mean that drivers are simply avoiding that intersection. Additionally, while this study was only aimed at Toronto, comparing these results to different cities, both near and far from Toronto, could provide more context and accuracy.

Were this study to be conducted once more, further research would have been conducted on habits of driving, trends of traffic volume, and comparing it to other cities with varying populations—so that we can control for more extraneous factors.

V. Conclusion

Our analysis showed that there is a moderately strong positive correlation between the number of red light cameras and the number of traffic charges in Toronto intersections ($R = 0.825$ collected as an average across regression models developed with 6 years of data. See Figure 3.7), confirming the initial hypothesis. While more extraneous data could have been used, along with tracking driver behaviour, this nonetheless shows that red light cameras are effective at ticketing red light runners. Research related to red light cameras, and in general about traffic control and intervention methods, can be used to better improve the safety of roads for both drivers and pedestrians.

References

[1] National Highway Traffic Safety Administration. *National Center for Statistics and Analysis Motor Vehicle Traffic Crash Data*. URL: https://crashstats.nhtsa.dot.gov.

[2] Mark L. Burkey and Kofi Obeng. *A detailed investigation of crash risk reduction resulting from red light cameras in small urban areas*. MPRA Paper 36261. University Library of Munich, Germany, July 2004. URL: https://ideas.repec.org/p/pra/mprapa/36261.html.

[3] Government of Canada and the Canadian Council of Motor Transport Administrators. *Canadian Motor Vehicle Traffic Collision Statistics*. data retrieved from Transport and Infrastructure, https://www.tc.gc.ca/eng/motorvehiclesafety/canadian-motor-vehicle-traffic-collision-statistics-2018.html. 2018.

[4] *City of Toronto Open Data Portal*. URL: https://open.toronto.ca.

[5] Jessica Edquist, Christina M. Rudin-Brown, and Michael Lenn. *Road Design Factors and Their Interactions With Speed and Speed Limits*. Tech. rep. 298. Monash University, 2009.

[6] Justin Gallagher and Paul J. Fisher. "Criminal Deterrence When There are Offsetting Risks: Traffic Cameras, Vehicular Accidents, and Public Safety". In: *SSRN Electronic Journal* (2017). DOI: 10.2139/ssrn.3078079. URL: http://dx.doi.org/10.2139/ssrn.3078079.

[7] N. J. Garber, J. S. Miller, R. E. Abel, S. Eslamblochi, and S. K. Korukonda. *The Impact of Red Light Cameras on Crashes in Virginia*. Charlottesville, Virginia, USA: Virginia Transportation Research Council, 2007.

[8] Francis Gichaga. "The impact of road improvements on road safety and related characteristics". In: *IATSS Research* 40 (May 2016). DOI: 10.1016/j.iatssr.2016.05.002.

[9] Insurance Institute for Highway Safety. *Fatality Facts 2018: Yearly snapshot*. 2018. URL: https://www.iihs.org/topics/fatality-statistics/detail/yearly-snapshot.

[10] Stephen E. Hill and Jay K. Lindly. *Red Light Running Prediction and Analysis*. Tech. rep. 02112. Department of Civil and Environmental Engineering. University Transportation Center for Alabama, 2003.

[11] W. Hillier, J. Ronczka, and F. Schnerring. *An Evaluation of Red Light Cameras in Sydney*. Research Note. New South Wales, Australia: Road Safety Bureau, Roads and Traffic Authority, 1993.

[12] Sergey Y. Kyrychenko and R. Retting. "Review of "A Detailed Investigation of Crash Risk Reduction Resulting From Red Light Cameras in Small Urban Areas" By M. Burkey and K. Obeng". In: 2004.

[13] Barbara Langland-Orban. "Red Light Running Cameras: Would Crashes, Injuries and Automobile Insurance Rates Increase If They Are Used in Florida?" In: *Florida Public Health Review* (2008).

[14] T. Mann, S. Brown, and C. Coxon. *Evaluation of the Effects of Installing Red Light Cameras at Selected Adelaide Intersections*. Report Series. Walkerville, SA, Australia: Office of Road Safety, South Australian Department of Transport, 1994.

[15] Kristie L. Hebert Martinez and Bryan E. Porter. "Characterizing red light runners following implementation of a photo enforcement program". In: *Accident Analysis and Prevention* 38.5 (2006), 862–870. DOI: 10.1016/j.aap.2006.02.011. URL: http://dx.doi.org/10.1016/j.aap.2006.02.011.

[16] H. W. McGee and K. A. Eccles. "Impact of Red Light Camera Enforcement on Crash Experience". In: Synthesis 310 (2003).

[17] *National Collision Database Online*. URL: https://wwwapps2.tc.gc.ca/Saf-Sec-Sur/7/NCDB-BNDC/p.aspx?l=en.

[18] Rebecca B. Naumann, Ann M. Dellinger, Eduard Zaloshnja, Bruce A. Lawrence, and Ted R. Miller. "Incidence and Total Lifetime Costs of Motor Vehicle–Related Fatal and Nonfatal Injury by Road User Type, United States, 2005". In: *Traffic Injury Prevention* 11.4 (2010).

PMID: 20730682, pp. 353–360.
DOI: 10.1080/15389588.2010.486429.
eprint: https://doi.org/10.1080/15389588.2010.486429.
URL: https://doi.org/10.1080/15389588.2010.486429.

[19] *OpenStreetMap*. URL: https://www.openstreetmap.org/.

[20] *QGIS*. URL: https://qgis.org/en/site/.

[21] QGIS Python Plugins Repository. URL: https://plugins.qgis.org/plugins/QuickOSM/.

[22] Richard A. Retting, Susan A. Ferguson, and Shalom Hakkert. "Effects of Red Light Cameras on Violations and Crashes: A Review of the International Literature". In: *Traffic injury prevention* 4 (Apr. 2003), pp. 17–23. DOI: 10.1080/15389580309858.

[23] Richard A. Retting, Allan F. Williams, Charles M. Farmer, and Amy F. Feldman. "Evaluation of Red Light Camera Enforcement in Fairfax, Va., USA". In: *ITE Journal* (1999).

[24] Robert G. Ulmer Richard A. Retting and Allan F. Williams. "Prevalence and characteristics of red light running crashes in the United States". In: *Accident Analysis and Prevention* 31.6 (1999), 687–694. DOI: 10.1016/s0001-4575(99)00029-9.

[25] Eduardo Romano, Scott Tippetts, and Robert Voas. "Fatal red light crashes: The role of race and ethnicity". In: *Accident Analysis and Prevention* 37 (June 2005), pp. 453–60. DOI: 10.1016/j.aap.2004.12.006.

[26] City of Toronto. *Red Light Cameras*. 2019. URL: https://www.toronto.ca/services-payments/streets-parking-transportation/traffic-management/pavement-markings/red-light-cameras/.

[27] *Toronto Police Service Public Safety Data Portal*. URL: http://data.torontopolice.on.ca/pages/open-data.

[28] U.S. Department of Transportation. *2018 Fatal Motor Vehicle Crashes: Overview*. Traffic Safety Facts, Research Note. DOT HS 812 826. National Highway Traffic Safety Administration, 2019.

[29] U.S. Department of Transportation. *Intersection Safety*. Federal Highway Administration. 2019. URL: https://safety.fhwa.dot.gov/intersection/conventional/signalized/rlr.

[30] U.S. Department of Transportation. *Traffic Safety Facts 2017 (Final)*. DOT HS 812 384. National Highway Traffic Safety Administration, 2016.

[31] *WISQARS (Web-based Injury Statistics Query and Reporting System)*. 2019. URL: https://www.cdc.gov/injury/wisqars.