Bike Lanes and Traffic in Toronto

A Spatially Motivated Statistical Study

Kaarthick Rameshkumar

STAT 436 - Final Project

Winter 2025

1 Introduction

Biking infrastructure in Toronto has been a contentious topic for the province in recent times. The Ontario government announced that it plans to remove three major bike lanes in the city and greatly reduce the city's ability to place new bike lanes without provincial approval. The government claims that removing bike lanes will reduce congestion and improve safety. Some small business owners claim that bike lanes hurt their businesses. However, the Ministry of Transportation of Ontario (MTO) admits that it does not have adequate data to decide whether bike lanes should be removed. Furthermore, advocates suggest that bike lanes improve safety and the economy (Harrison, 2025).

The goal of this report is to understand the impact that bike lanes have on traffic. In particular, the neighbourhoods of Toronto were investigated with respect to traffic, how many bike lanes run through them, and various demographic information.

2 Data and Preliminary Exploration

The city of Toronto collects data relevant to this study and its open data catalog was used extensively for this study. Four major datasets from the catalog were used:

- `Traffic Volumes - Multimodal Intersection Turning Movement Counts` - This dataset included counts for cars, bikes, and pedestrians arriving at a number of intersections in Toronto over 14 or 8 hour periods. The counts were binned into 15 minute intervals from 6:00 AM to 8:00 PM for 14 hour observations or non-continuous routines for 8 hour observations. The counts were also sorted by the direction of the vehicles' approach to an intersection, with possibilities of North, East, South, or West. A summary dataset was available for these counts, which included the total observed vehicles and pedestrians over the entire observation duration at an intersection. The summary dataset was used in this study.

- `Neighbourhood Profiles - 2016` - This dataset included census data about the demographics of Toronto's 140 neighbourhoods. Over 2600 different characteristics were available, and a small subset of these were used for the study. Although there was a neighbourhood profile dataset from 2021, the older 2016 dataset was used. The older dataset was much better organized and labelled, which made choosing a subset of the variables much easier. Unfortunately, the 2016 dataset was created when Toronto only had 140 neighbourhoods instead of the current 158. Thus, an older neighbourhood map was also used.

- `Neighbourhood and Cycling Maps` - These two datasets contained mapping data about Toronto's neighbourhoods and cycling network respectively. The geometry from the datasets was used to place the intersections from the traffic data into neighbourhoods and calculate the number of bike lanes in each neighbourhood.

These three datasets were combined together to form the analysis of this report. In order to model traffic, the number of cars passing through each observed intersection was used as a proxy. In particular, the total number of cars observed at each intersection during the observation period from the summary dataset was used. The observed counts were totaled for each of the 140 neighbourhoods of Toronto. The number of hours that each intersection was observed for was also totaled for each neighbourhood.

The intersections and their respective counts of cars is shown in the first map of Figure 1. This maps shows that the larger counts of cars occur along neighbourhood boundaries and other lines running through the map. These lines correspond to the major roads of Toronto, and so it makes sense that the larger traffic counts were viewed here. Furthermore, we see that some neighbourhoods have many more observed intersections than others. This was likely because the City of Toronto was more interested in certain intersections and neighbourhoods that others. It may also be attributed to the fact that certain neighbourhoods near the city's center had more intersections. The second map in Figure 1 shows the aggregated car cous for each neighbourhood divided by the time that the neighbourhood was observed for. As expected, many of the neighbourhoods with larger observed cars per hour are located near the city center. Furthermore, neighbourhoods with higher rates are generally group together. There are few exceptions to this. A notable one is the Woburn neighbourhood (found close to the east end of the map), which has one of the lowest cars per hour. This neighbourhood is close to neighbourhoods with larger rates. However, this second map largely supports the idea that there is spatial dependence on traffic between neighbouring neighbourhoods of Toronto. This idea will be used to create traffic models later in this report.

Histograms for the number of cars observed in each neighbourhood, the time spent observing intersections, and their ratio are pictured in Figure 2. The distribution of cars and time observed is havily right skewed. This indicates that certain neighbourhoods had more observed intersections in them than others. This was also observed in the maps of Figure 1. The mean number of cars observed in a neighbourhood was 652521.1 with a standard deviation of 506581.3. The mean number of hours that were spent in each neighbourhood was 483.0143 with a standard deviation of 352.9226. The large standard deviations are indicative of the skewness in the first two histograms. The third histogram is a histogram of the average number of cars per hour in each neighbourhood. This histogram is not skewed and the rate seems to be normally distributed. The mean cars per hour was 1468.406 with a standard deviation of 379.464. The models which were fit for this study use the counts for each neighbourhood adjusted for the time of observation to account for the differences among neighbourhoods. In addition to adjusting by the observation time, the study aimed to adjust

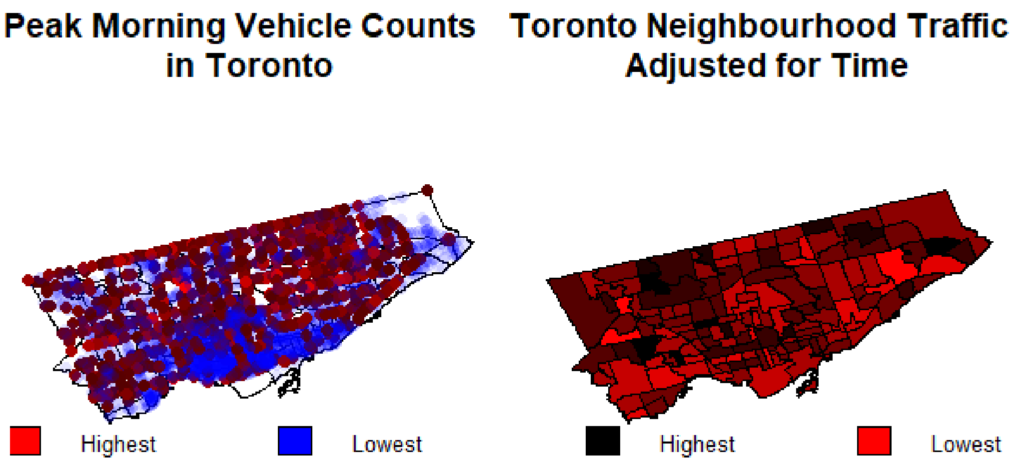

Figure 1: Maps of Traffic at Intersections and Rates of Traffic in Neighbourhoods

for demographic variables of each neighbourhood. The neighbourhood profiles dataset was used for

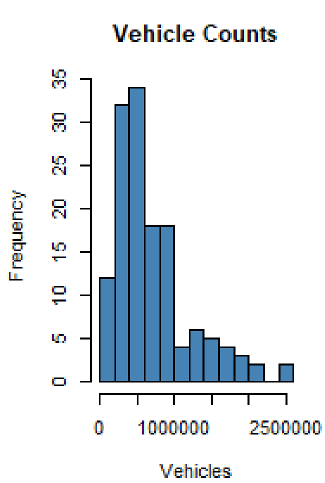

Figure 2: Histograms of Neighbourhood Car Counts, Observation Times, and Rates

this purpose. In order to reduce the amount of covariates used for regression models, only certain categories of demographic information were used. These include the population and dwellings, age characteristics, income of individuals in 2025, highest certificate, diploma or degree, labour force status, main mode of commuting, and commuting duration categories of the census data. There were a total of 122 covariates in these categories. In order to further reduce the list of covariates used, a LASSO regression model was fit with all these covariates. The response variable for this regression was the average number of cars passing through intersections per hour in each neighbourhood. The penalization parameter for the lasso regression model was chosen by cross-validation available from the `glmnet` package in `R`. The final LASSO model had a non-zero coefficient for the following covariates:

- Population Change (%) between 2011 and 2016
- Number of Males aged 85 to 89
- Number of Males aged 100 plus
- Percent of Female Participation in the Workforce
- Number of People who Commuted by Bike
- Number of People who Commuted by Means other than Walking, Transit, Biking, Or Driving
- Number of College Degree Holders
- Number of Medicine or Adjacent Degree Holders

Due to the large number of covariates, the LASSO regression model likely dropped many arbitrarily. As a result, some simplifications were made to this list. Instead of the number of males in each neighbourhood, the overall senior population was used. Instead of just the female participation in the labour force, the full participation rate was used. Furthermore, the number of bike and other commuters was not used because these variables were too similar to the number of pedestrians.

Aside from the demographic information of each neighbourhood, the number of bike lanes present in each neighbourhood was of interest. This was calculated by using the neighbourhood and bike lane geometries provided by the city of Toronto. The vertices which formed the bike lanes and paths were checked for which neighbourhood they were in, and the count for each respective neighbourhood was aggregated. However, some distinction were made between types of bike lanes. For each neighbourhood, a count of major and minor bike lanes was created. The major bike lanes included the cycle tracks (separated paths on roads with buffer space between traffic) and the bike lanes (separated paths on roads). The minor bike lanes included all other types of bike paths, such as park trails and shared roads.

The final quantity of interest was the distance of each neighbourhood to the city center. From the earlier preliminary analysis, it was clear that neighbourhoods closer to Downtown Toronto had higher rates of cars at their intersections. Therefore, a neighbourhood's proximity to downtown was important to consider for any regression models. The distance from a neighbourhood's centroid to the centroid of the Kensington-Chinatown neighbourhood was used as a proxy for this quantity. The Kensington-Chinatown neighbourhood is one of the neighbourhoods in Downtown Toronto, and so it made sense to use this neighbourhood as the fixed downtown location.

3 Methods

In order to understand the impact of bike lanes on traffic, the study aimed to fit adequate regression models and obtain inference from their parameters. The major regression model considered was a Bayesian Spatial Linear Model. Two models in particular were considered. First, let Y_i be the number of cars per hour observed in each neighbourhood. The following model was fit:

$$Y_i = \sum_{j=1}^{7} \beta_j x_{ij} + u_i + v_i$$

where $v_i \sim N(0, \sigma_v^2)$ and the probability density function of u_i, the spatial random effect, is:

$$f(u_i) \propto (\sigma_u^2)^{-n/2} \exp\left[-\sum_{i \neq j} (u_i - u_j)^2 w_{ij} / 2\sigma_u^2 \right]$$

We let σ_u^2 and $\sigma_v^2 \sim$ Inverse Gamma$(0.001, 0.001)$ and w_{ij} to be the indicator for whether the neighbourhoods i and j are neighbours or not. Finally, the covariates $x_{i1}, \ldots, x_{i7}$ for each neighbourhood are its population change from 2011 to 2016, number of seniors, number of college degree holders, percentage of individuals in the workforce, distance from the Kensington-Chinatown neighbourhood, number of major bike lanes, and number of minor bike lanes respectively.

The second model which was fit was a Bayesian Spatial GLM. For this model, let Y_i be the number of cars arriving in total for each neighbourhood. We assume that $Y_i \sim$ Poisson(λ_i). We model the rate λ_i as follows:

$$\log(\lambda_i) = \sum_{i=j}^{3} \beta_j x_{ij} + \log(Person - Time) + u_i + v_i$$

The distribution of u_i was kept the same as the previous model, but we let σ_u^2 and $\sigma_v^2 \sim$ Inverse Gamma$(0.0005, 0.5)$. Furthermore, the $\log(Person - Time)$ was used as an offset for this Poisson GLM. It was calculated as the product of the total duration spent in each neighbourhood times the total amount of traffic (including cars, pedestrians, bikes, and other vehicles) at all the intersections observed in the neighbourhood.

These models were fit using the `rstan` package in `R`. The first model was sampled from 35000 times and the first 20000 iterations of the MCMC algorithm were discarded as burn-in. The remaining were used to get posterior means and credible intervals for the regression parameters. The second model was sampled from 4200 times and the first 2500 iterations were discarded as burn-in.

Unfortunately, both models had some convergence issues, which will be discussed later in the report. Furthermore, many of the covariates deemed important in the preliminary exploration were not used in the models here. This was due to computational complexity of running of more complicated models.

4 Results

The parameters of interest for the models fit were those which related to the number of major and minor bike lanes in each neighbourhood. The mean of each posterior distribution for the basic linear Bayesian model are found in Table 1. The posterior mean for the major and minor bike lane counts are -0.137 and -0.301 respectively. The credible interval for the major bike lane count includes zero. Hence, provided the model is correct, we may conclude that the number of major bike lanes in each neighbourhood does not significantly impact the number of cars which pass through it. The number of minor bike lanes negatively impacts traffic in a statistically significant manner. The rate of cars per hour goes down by 0.3 for each minor bike lane present. Given that there are an average of 13 minor bike lanes in each neighbourhood, this suggests that the number of cars per hour is reduced by an average of 3.9 due to minor bike lanes in each neighbourhood. The remaining parameters are all statistically significant, since none of them include zero. The posterior means of the parameters

Covariate	**Posterior Mean**	**Lower 95% CI**	**Upper 95% CI**
Population Change (2011–2016)	29.99753076	29.991282996	29.999938293
Number of Seniors	0.06088027	0.046533269	0.122916334
Number of College Degree Holders	-0.00319037	-0.006411216	-0.002336998
Workforce Participation	5.43032306	4.207788872	10.929502469
Distance from Kensington-Chinatown	19.06745369	18.730818274	19.810933787
Major Bike Lanes	-0.04775553	-0.199486639	0.025852515
Minor Bike Lanes	-0.13724403	-0.301172662	-0.072560695

Table 1: Normal Bayesian Model Parameters and Credible Intervals

and their credible intervals for the Poisson Bayesian GLM can be found in Table 2. The coefficients now have a relative rate interpretation. Interestingly, the major bike lanes appear to reduce the rate of cars, whereas the minor bike lane count appears to increase the car count. The rate ratio of car arrivals for a neighbourhood with one increased major bike lane is $\exp(0.98) = 0.49$, whereas it is $\exp(0.001) = 1.001$ for each minor bike lane. The spatial effects term for each neighbourhood varied significantly between neighbourhoods. These effects (their posterior means) are shown in the histograms in Figure 3. For the normal model, the mean posterior spatial effects are mostly centered

Variable	Posterior Mean	Lower 95% CI	Upper 95% CI
Distance from Kensington-Chinatown	-3.008381750	-3.628912114	-2.967901596
Major Bike Lanes	-0.017569406	-0.021138104	-0.016866290
Minor Bike Lanes	0.001385855	0.001179938	0.001671095

Table 2: Normal Bayesian Model Parameters and Credible Intervals

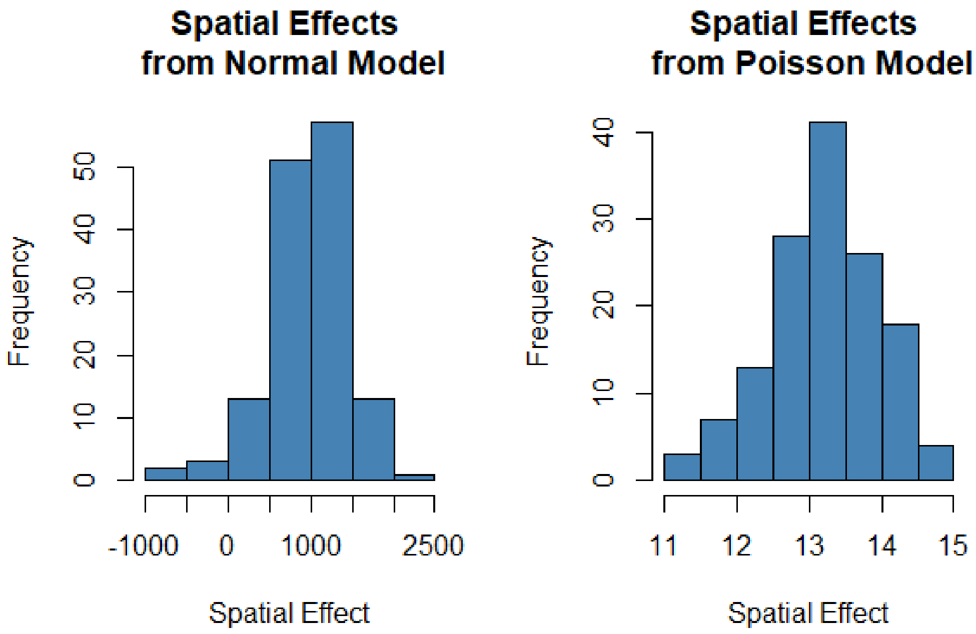

Figure 3: Spatial Effects for Bayesian Models

about the mean of 1000. For the poisson model, the effects are more spread out. Given that they are on different scales, the exact spatial effects are difficult to compare. However, it is clear that there is some level of dispersion in the posterior means of the spatial effects. This suggests that the spatial modelling approach is valid.

5 Discussion

From the results presented above, it seems that major bike lanes do not have a large impact on traffic. In the first model, they appear to have no impact. In the second model, they appear to have a very small negative impact. In the first model, minor bike lanes appeared to hinder traffic, whereas they reduced traffic in the second model. Their negative impact in the first model was also miniscule; reducing the hourly rate of cats by only around 4. This is very small compared to the average 1486 cars per hour in a neighbourhood. The models seem to suggest that the presence of bike lanes either does not impact traffic, or they impact traffic in very minor ways.

There were a few issues with the models which were fit. Both models had convergence issues for their parameters. In particular, the first model's parameter for the impact of population change on the cars per hour was very close to 30. In the model fitting process, the model was given a prior Uniform(-30,30) distribution, which explains the ceiling. Attempts to fit the model with higher ceilings ended with similar results.

Another issue the models faced was their computational complexity. The first model took almost an entire hour to compile. This was after the prior distributions were reduced in scope. The Poisson model had an estimated runtime of around 60 hours with all the covariates included. As a result, many of the covariates needed to be dropped, and the number of iterations run was also heavily decreased. In the future, the model should be run with all of its covariates and preferably with a better computer. This would provide more reliable results.

Finally, this report only serves as preliminary exploration of the topic of bike lanes and their relationship with traffic. A lot of preliminary work was done in order to collect demographic variables and the placement of bike lanes. However, further analysis should be done and variables such as speed limits of roads, small business revenues, and environmental impacts should be studied. Furthermore, the spatial effects should be made more sophisticated in the future. Neighbourhoods within a certain distance, or which are connected by major roads should also be able to impact each other's traffic.

One potential topic further study is the impact of the Bloor Street bike lane. This lane was placed recently, so there may traffic data available from before and after the bike lane's installation.

6 References

Harrison, Lane (2025, March 12). Ontario aware bike lane removals may not reduce congestion, could make people less safe: internal documents. CBC News. https://www.cbc.ca/news/canada/toronto/ontario-bike-lanes-internal-documents-1.7481729

TMC Data - https://open.toronto.ca/dataset/traffic-volumes-at-intersections-for-all-modes/
Neighbourhood Maps - https://open.toronto.ca/dataset /neighbourhoods/
Neighbourhood Profile Data - https://open.toronto.ca/dataset/neighbourhood-profiles/
Cycling Network Map - https://open.toronto.ca/dataset/cycling-network/