Toronto's Average Recorded Car Speeds Reflects That The City's Traffic Isn't That Bad As It Seems To Be*

Kevin Shao

September 27, 2024

Road traffic is a significant issue nowadays as many people rely on driving cars for work, school, and leisure purposes. Data of recorded car speeds at various locations with speed display signs present is used to discover Toronto's road traffic severity. After data processing and analysis with the use of R, it can be inferred that there is no evidence that Toronto has comparatively worse traffic compared with global statistics. However, a more thorough research with the use of more data sets should be conducted for more accurate results.

1 Introduction

Cars are one of our most frequently used transportation method, but automobiles also produce multiple issues, such as traffic accidents and traffic. In fact, with many vehicles driving on road networks, numerous metropolitan areas have been experiencing severe traffic, largely increasing people's everyday commuting times. In 2022, Toronto was ranked seventh in the world and first in Canada for bad traffic, and people in Toronto speed 118 hours on congested road (O'Brien (2023)). Toronto's severe traffic makes it meaning to improve the city's road network, and a useful data set may be a collection of recorded driving speeds, as low car speeds may reflect congestion on corresponding roads. Such numerical data is also suitable for completing statistical analysis, such as calculating mean (average speed) and standard deviation (a measure of how spread-out numbers, or data points, are).

The purpose of this study is to analyze the extent Toronto's road traffic in comparison with global performance, which is the first step for deciding how to improve Toronto's road network for less congestion. Here, I will measure the extent of traffic/congestion using average driving

*Code and data used in this analysis can be found at: https://github.com/KevinShao1357/Paper1-Updated.git speeds (higher speeds means less traffic) since it is a straightforward factor. Therefore, the chosen data set should be most independent from other confounding variables. I would like to eliminate the factor of dangerous driving at high speeds. A data set recording speeds at locations with present speed display signs is a reasonable solution, as speed display signs may give drivers incentives to drive carefully, as it implicates that devices which enforce speed limits may be present nearby. Therefore, the chosen data set for this study is 'Mobile Watch Your Speed Program – Speed Summary', one of the few provided by Open Data Toronto which presents recorded car speeds ("Mobile Watch Your Speed Program – Speed Summary" (2024)).

After completing data processing and analysis (the 'data' introduces more detail on this), I can ultimately infer that there is no evidence that Toronto has comparatively worse-off road traffic. This study may give Toronto Transportation Services a detailed image of Toronto's road traffic performance, and may also help Toronto choose more ideal locations to place their speed display signs.

2 Data

2.1 Raw Data

'Mobile Watch Your Speed Program – Speed Summary' on Open Data Toronto was published by Toronto Transportation Services and was last renewed on 5 September 2024. This data set provides data about recorded driving speeds of cars which pass by speed display signs. Speed display signs uses radar devices to measure speeds of oncoming cars and displays their speeds on LED screens, reminding drivers to obey speed limits. The equipped radar device detects a change in frequency of returned radar signals, in proportion to speed by the Doppler Effect ("Frequency Modulated Radar" (1949)). These speed display signs are portable and are moved around locations across Toronto in a monthly basis ("Mobile Watch Your Speed Program – Speed Summary" (2024)).

All data processing, including cleaning and analysis, are performed by programming in R language (R Core Team (2023)). The R code with detailed notes of data cleaning are included in the 'scripts' section of the GitHub repo link shown in the footnote, while other approaches related to data are explained in the following sections. In this study, the following R packages are used: tidyverse (Wickham et al. (2019)), readr (Wickham, Hester, and Bryan (2024)), dplyr (Wickham et al. (2023)), janitor (Firke (2023)), ggplot2 (Wickham (2016)), knitr (Xie (2014)), opendatatoronto (Gelfand (2022)), and tibble (Müller and Wickham (2023)).

2.2 Data Processing

First of all, the final data file prepared for analysis only consists the average recorded speeds for each location, since average recorded speeds is less affected by outliers, such as speeding vehicles. Average speeds take account of confounding factors, ideal for analyzing Toronto's overall road traffic. In this data set, there is one average speed per location. Units for average recorded speeds are in kilometers per hour (km/h).

Before using the data for graphs and tables, I made sure that all the data elements of the data are valid speeds by checking if there are any negative values, NA values, and if all elements are numeric. In this case, if the speed is recorded to be 0, it is included in the data set and would be considered valid. I confirmed that the elements arein correct format, so there is no need to change anything in 'analysis_data'. A sample of the final processed data is shown in Table 1 below.

Table 1: Distribution of the number of doctor visits

Recorded Mean Speeds (km/h)
30
5
50
32
42
21
50
0
15
13

2.3 Data Presentation and Analysis

To get a clearer image of of the data distribution, Figure 1 is a histogram produced using the processed data.

From Figure 1, I can infer that there are two outliers at 0 and 80 km/h respectively, and The mean these speeds is roughly around 30 km/h. The outlier at 80 km/h has a much lower frequency than that at 0 km/h. There are no data points in the interval of 0 to 5 km/h. With consideration of the relatively high frequency of observations at 0 km/h, it can be inffered is that all values between 0 to 5 km/h are rounded to 0 km/h, since there should not be no values between a 5 km/h interval given a data set with roughly 12000 observations in total. If this hypothesis holds true, it can be concluded that there is a relatively high frequency of cars in Toronto that are experiencing congestion may be concluded. If the car speed is recorded as 0 km/h, it may be parking and the data is recorded by error, but if a car is driving at a very low speed of 0 to 5 km/h, it would be either in traffic or just accelerating after a traffic light turns green. However, the prior case is much more likely to occur, bas being in traffic means that cars tend to drive at low speeds for longer durations. Moreover, if the speed display signs

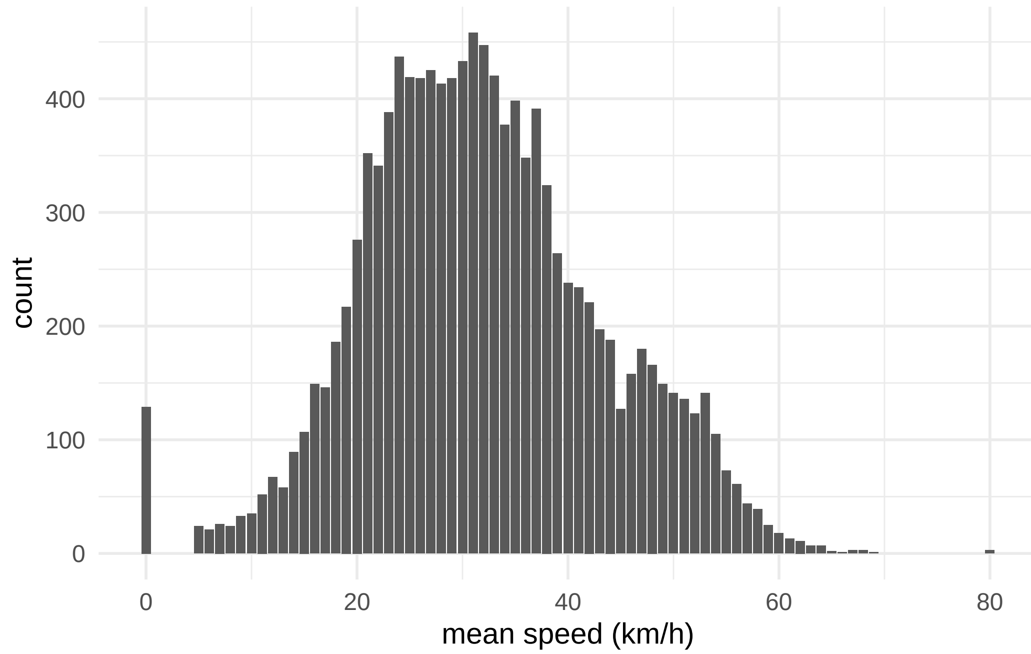

Figure 1: Histogram of Average Recorded Speeds

are located just beside a traffic light, its radar would track a car either driving fast with a green light or a car just accelerating when the traffic light turns green, so the average recorded speeds would be less likely to be as low as 0 to 5 km/h.

To acquire more supporting data, Table 2 below is created to represent the processed data's summary statistics.

Table 2: Summary Statistics of Average Recorded Speeds

	Values of Corresponding Statistics (km/h, Rounded to Four Decimal Places)
mean	31.86784
standard_deviation	11.36143
maximum	80.00000
minimum	0.00000

According to Table 2, the mean, standard deviation, maximum, and minimum of the speeds are 31.9, 11.4, 80, and 0 km/h respectively, which cohere to the previously-introduced hypothesis that there is a relatively high frequency of cars experiencing ongoing traffic at locations with present speed display signs.

Considering the mean and standard deviation of average driving speeds in Toronto, these values are actually higher than the global average of 30 km/h measured in 2022 ("Statistics on Global Journeys & Markets (Mostly Pre-Covid)" (n.d.)). The average driving speeds in Toronto have also improved compared to that of Toronto two years ago, measured as 24 km/h in 2022, and is much higher than the 20 km/h speed during peak traffic (O'Brien (2023)). Note that the average recorded speeds in Toronto are measured where speed display signs are present, and such signs give drivers incentives to drive more carefully, and more slowly, than usual, meaning that the measured speeds in Toronto may be even lower than the actual value. This means that although there is a relatively higher frequency of cars in traffic compared to the number of cars that drive exceptionally faster as shown in the two outliers at 0 and 80 km/h respectively of the data set, these possible congestion occurences are not significant compared to global road traffic performances.

2.4 Data Conclusion and Limitations

Although it may be hard to conclude that Toronto's traffic are better-off compared to global circumstances just by this one data set, there is no evidence that Toronto has comparatively bad traffic given the high average recorded speeds displayed in this chosen data set.

There are also a few limitations of this chosen data set. First of all, the data set only contains measured car speeds at locations where speed display signs are present, so the speeds may be under-measured, as the use of only using average data already minimized the impact of confounding factors such as speeding vehicles. Secondly, the data set does not measure speeds at all location in Toronto, so locations where severe congestion often occur may be ignored. It can be inferred that Toronto Transportation Services chose to assemble speed display signs at these locations largely due to multiple cases of captured dangerous driving, which conflicts with cases of severe traffic, since drivers have no way to drive fast during severe traffic.

However, given such limitations, although more data sets must be combined to get a thorough conclusion, the inference that there is no evidence that Toronto has comparatively bad traffic can still be made.

3 LLMs

Statement on LLM usage: No LLMs have been used in this paper for any purpose.

References

Firke, Sam. 2023. "Janitor: Simple Tools for Examining and Cleaning Dirty Data." https://github.com/sfirke/janitor.

"Frequency Modulated Radar." 1949. McGraw Hill.

Gelfand, Sharla. 2022. "Opendatatoronto: Access the City of Toronto Open Data Portal." https://sharlagelfand.github.io/opendatatoronto/.

"Mobile Watch Your Speed Program – Speed Summary." 2024. Transportation Services. https://open.toronto.ca/dataset/mobile-watch-your-speed-program-speed-summary/.

Müller, Kirill, and Hadley Wickham. 2023. "Tibble: Simple Data Frames." https://tibble.tidyverse.org/.

O'Brien, Abby. 2023. "Toronto Ranked One of the Worst Cities in the World for Congestion." https://toronto.ctvnews.ca/toronto-ranked-one-of-the-worst-cities-in-the-world-for-traffic-1.6226070.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"Statistics on Global Journeys & Markets (Mostly Pre-Covid)." n.d. https://movotiv.com/statistics.

Wickham, Hadley. 2016. "Ggplot2: Elegant Graphics for Data Analysis." Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software*. 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://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. "Readr: Read Rectangular Text Data." https://readr.tidyverse.org.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.