Toronto's Average Recorded Car Speeds At Locations Where Speed Display Signs Are Present Imply That The City's Traffic Isn't That Bad*

Kevin Shao

September 27, 2024

Road congestions largely disrupt our everyday life by significantly increasing our everyday commuting time. To discover Toronto's overall road traffic severity, a data set of recorded car speeds at locations where speed display signs are present is used to conduct a study. With the use of histograms and tables of summary statistics such as mean and standard deviation, it can be inferred that there is no evidence that Toronto has comparatively bad traffic. However, limitations and bias still exist, and more data sets must be used in combination for getting a more accurate conclusion.

1 Introduction

Cars are one of our most frequently used transportation method, and are common for personal, military, and corporate use. There are still issues that disrupt people's everyday usage of cars, including traffic accidents, such as minor collisions as well as severe accidents with casualties, and traffic, which largely increases people's daily commuting time. In 2022, Toronto was ranked seventh in the world and first in Canada for bad traffic, and people in Toronto use 118 hours a year on Greater Toronto Area (GTA) roads due to congestion (O'Brien (2023)), which implicates the severe traffic in this Canadian metropolitan city. It is clear that improvements on Toronto's severe congestion is meaningful for improving the city's transportation efficiency. A medium for discovering Toronto's traffic is to get data of the city's driving speeds, as very low car speeds may reflect congestion on corresponding roads. Such a data set is also suitable for performing statistical analysis, since speeds in units of kilometers per hour (km/h) are

*Code and data used in this analysis can be found at: https://github.com/KevinShao1357/Paper1-Updated.git

numerical data that are easy for calculating summary statistics such as mean (average speed) and standard deviation (a measure of how spread-out numbers are for a data set).

Since the aim for this study is to analyze the extent of traffic in Toronto, the chosen data set should be most independent from other variables that may affect our results. Given that my chosen data is Toronto's driving speeds, I would like to eliminate the factor of dangerous driving at high speeds. Therefore, the final choice is a data set which records speeds at locations with existing speed display signs. At roads which have a speed display sign, there is an implication speeding here may lead to penalties, which may give drivers incentives to drive carefully and adhere to traffic rules. Therefore, this data set may avoid the disruptive variable of dangerous driving. After a thorough selection process on Open Data Toronto, the chosen data set is 'Mobile Watch Your Speed Program – Speed Summary' ("Mobile Watch Your Speed Program – Speed Summary" (2024)).

The entire study has the following structure. The 'Abstract' section provides the general summary of the paper's background, approaches, and findings. The 'Introduction' section introduces relevant background information, purpose of completing this study, and leads in to the 'Data' section. The 'Data' section dives in detail of the raw data, data processing, and data analysis, and finally gets an inference, also pointing out some relevant limitations.

After completing data processing and analysis, I can ultimately infer that there is no evidence that Toronto has comparatively bad (automobile) traffic, which is based on the provided figures and tables produced from processed data, as well as the summary statistics calculated such as mean and standard deviation (the 'Data' section gives a more detailed explanation of the raw data, data processing, and data analysis). 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

The process of choosing this specific data set is written in the previous 'Introduction' section, and this data set is one of the few provided by Open Data Toronto which directly provides numerical data points of recorded car speeds.

As mentioned in the Introduction, the data set chosen is 'Mobile Watch Your Speed Program – Speed Summary" on Open Data Toronto, which is published by Toronto Transportation Services and provides data of recorded speeds on locations where speed display signs are present. Speed display signs uses radar devices to measure speeds of oncoming cars passing through and uses LED screen to display their speeds, with the purpose of reminding drivers to obey speed limits. Specifically, the radar device equipped on the speed display signs detects a change in frequency of returned radar signals, which is in proportion to speed by the Doppler

Effect ("Frequency Modulated Radar" (1949)). These speed display signs are portable and are moved around various locations across Toronto, This data set records the location of each speed display sign, the unique id of each speed display sign location, the minimum speed when the devices starts to display oncoming vehicle's speeds, various percentiles of passing vehicle's measured speeds in units of kilometers-per-hour (km/h), as well as the number of vehicles with recorded speeds within various five kilometer-per-hour intervals ("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, note that the final data file prepared for analysis only consists the average recorded speeds for each location. This is because average recorded speeds is less affected by outliers, such as dangerous-driving drivers, and reflects the general speed of cars passing by, which is ideal for measuring the extent of traffic for entire Toronto. Also, the average recorded speeds for each location are recorded with units 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. I performed this data testing process by checking if there are any negative values, NA values, and if all elements are numeric. Note that in this case, if the speed is recorded to be 0, it is included in the data set from Open Data Toronto and would be considered valid. I confirmed that the elements are of 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

Recorded Mean Speeds (km/h)
15
13

2.3 Data Presentation and Analysis

My processed data is now ready for producing graphs and tables for analysis. To get a clearer image of of the data distribution, I will use the processed data to produce the histogram Figure 1.

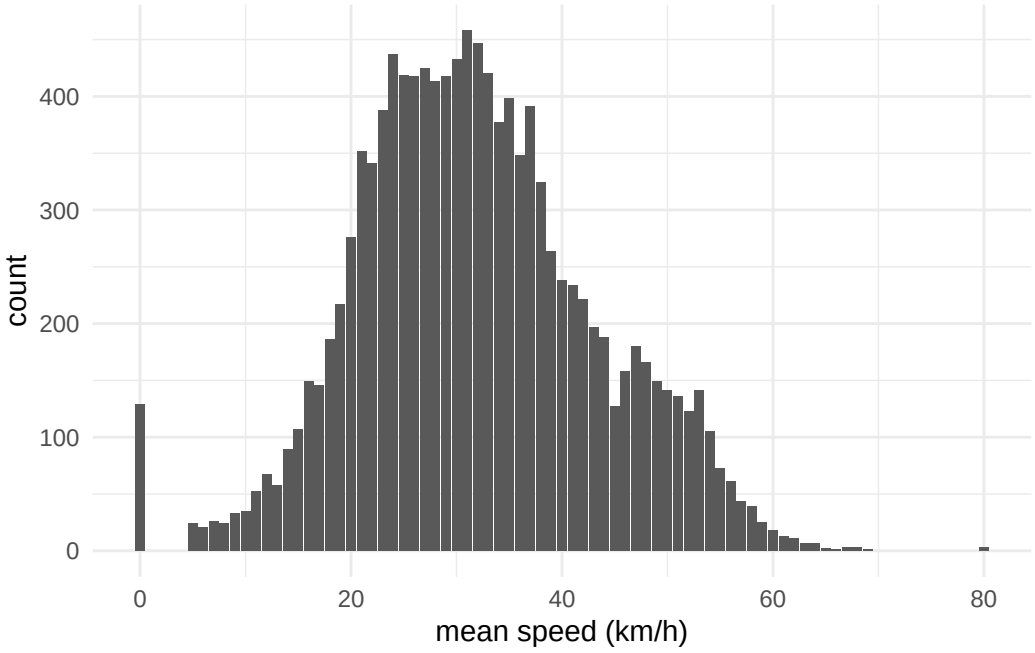

Figure 1: Histogram of Average Recorded Speeds

From Figure 1, I can infer that there are two outliers of the recorded average speeds, which are at roughly 0 and 80 kilometers per hour. The mean of the recorded average speeds is roughly around 30 kilometers per hour. However, I must also note that the outlier at 80 kilometers per hour has a much lower frequency than the outlier at 0 kilometers per hour. Here, I must also take account that there is no data points in the interval of 0 to 5 kilometers per hour, and with consideration of the high frequency of observations at 0 kilometers per hour, a reasonable hypothesis is that all values between 0 to 5 kilometers per hour have been automatically rounded to 0 kilometers per hour, and this prediction is very likely to be true, since there should not be no values between an interval of 5 kilometers per hour with nearly 12000 observations in total. If this educated guess holds true, a conclusion that there is a relatively high frequency of cars in Toronto that are experiencing congestion may be

concluded. If car has a recorded average speed of 0 kilometers per hour, it may in a parking spot and the radar recorded it by error, but if a car is driving at a very low speed of 0 to 5 kilometers per hour, it will only either be in traffic or just accelerating after a traffic light turns green. However, the first case is much more likely to occur. First of all, being in traffic means that cars tend to drive at low speeds for longer durations and more often. Secondly, if the speed display signs are located just beside a traffic light, its radar would most likely track either a car driving relatively fast when the traffic light is green or a car just accelerating when the traffic light, so the average recorded speeds would be less likely to be as low as 0 to 5 kilometers per hour.

To acquire more supporting data for a more thorough analysis, 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 the summary statistics displayed in Table 2, the mean of the average recorded speeds is 31.9 kilometers per hour, the standard deviation is 11.4 kilometers per hour, and the corresponding maximums and minimums are 80 and 0 kilometers per hour respectively. These values cohere to the hypotheses introduced in the previous paragraph, that there is a relatively high frequency that cars experience ongoing traffic at locations where speed display signs are present.

However, considering the calculated mean and standard deviation of average recorded car speeds in Toronto, these values are actually higher than the global average car speeds measured in 2022 of 30 kilometers per hour ("Statistics on Global Journeys & Markets (Mostly Pre-Covid)" (n.d.)). Surprisingly, these data have also improved compared to the average car speeds in Toronto two years ago, as measured in 2022 of 24 kilometers per hour, and is dramatically higher than the 20 kilometers per hour measured during peak traffic times (O'Brien (2023)). We must also account 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 than usual, implicating the possibility that the measured average speeds in Toronto may be lower than the actual value. Although it may be hard to prove that the Toronto's traffic situations are overall better-off compared to global circumstances just by this one data set, there is ultimately no evidence that Toronto has comparatively bad traffic given the relatively high data of average recorded speeds.

There are also a few limitations of the chosen data set that may affect the accuracy of my concluded results. First of all, as mentioned in the previous paragraph, the data set only contains measured car speeds at locations where speed display signs are present, so there is an inevitable possibility of bias that the speeds may be under-measured, given that the use of only using average data already minimized bias. Secondly, the data set does not measure speeds at every location in Toronto, so there is the possibility that locations where severe congestion often occurs may be ignored. This is because one of the main incentives Toronto Transportation Services chose to assemble speed display signs at these locations is because dangerous driving, such as speeding, often happen at these locations. This may conflict with the occurence of severe traffic, since drivers have no way to drive fast during congestion.

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

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.