

How to reduce idling time for commercial vehicles in Toronto

Haoda Li, Shujun Yan

April 5, 2019

I. Introduction

If all drivers avoided unnecessary idling for three minutes a day, we would save over \$630 million per year (assuming a fuel cost of \$1/litre). What's more collectively we would prevent 1.4 million tonnes of carbon dioxide (CO₂) from entering the atmosphere daily and contributing to climate change.

— quoted from Natural Resources Canada. *Why do Canadians idle?*

Despite the effort from the government, idling is still a problem for cities. Among different kinds of vehicles, commercial vehicles are one of the big source of idling-related problems. Commercial vehicles spend much longer time on the road and the heavy duty trucks produce more emission than the small common cars. Therefore, with the Geotab data on commercial vehicles, we seek ways to reduce idling of commercial vehicles in Toronto. We will look at the areas, time, and type of vehicles to find reasons behind idling and make suggestions on the improvements.

II. Data Overview

This data report primarily uses data from Geotab's *Areas of Idling* and *Searching For Parking* data sets. The Geotab data set are derived at aggregate from the commercial vehicles in Geotab's data environment, which 70% are trucks, 10% are passenger cars, 16% are multi-purpose vehicles, and 4% are buses (1).

In addition to the Geotab's datasets, this report will use map data from City of Toronto Open Data, including the following:

- *Toronto Centreline*(2): Including linear features representing streets, walkways, rivers, railways, highways and administrative boundaries within the City of Toronto.
- *Business Improvement Area*(3): Including the boundaries of business improvement areas in city of Toronto.
- *Neighborhood*(4): Including the boundaries of all city of Toronto neighborhoods .

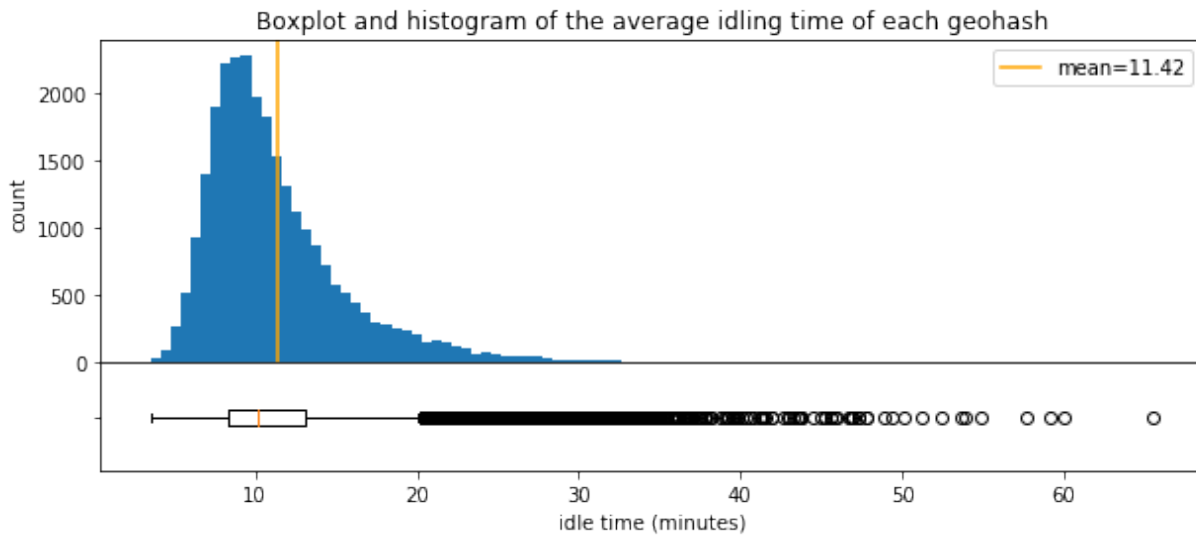
III. Identify idling problems and idling areas

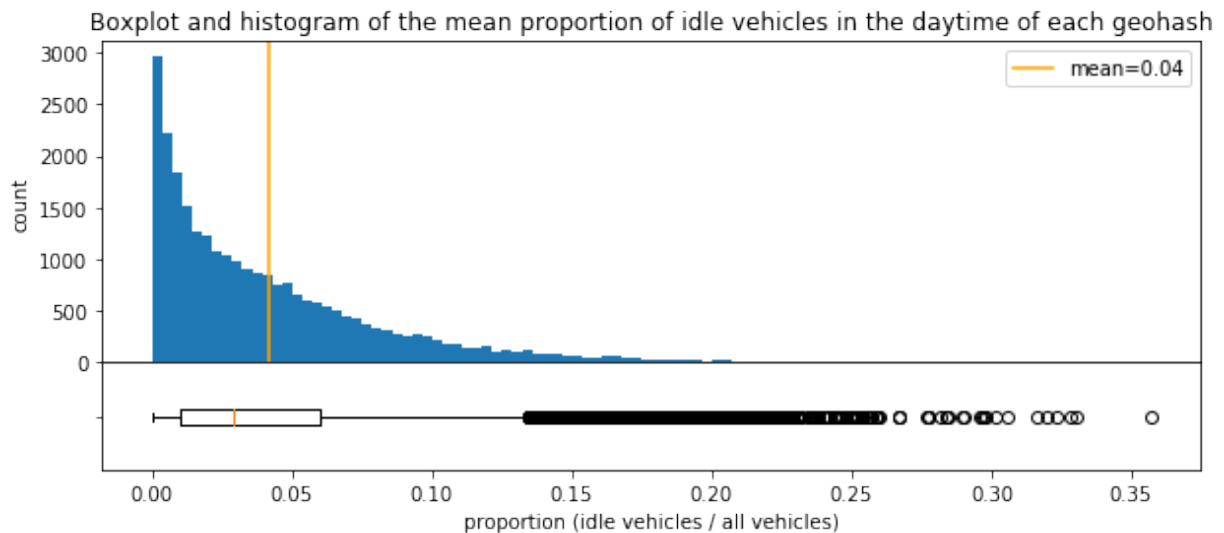
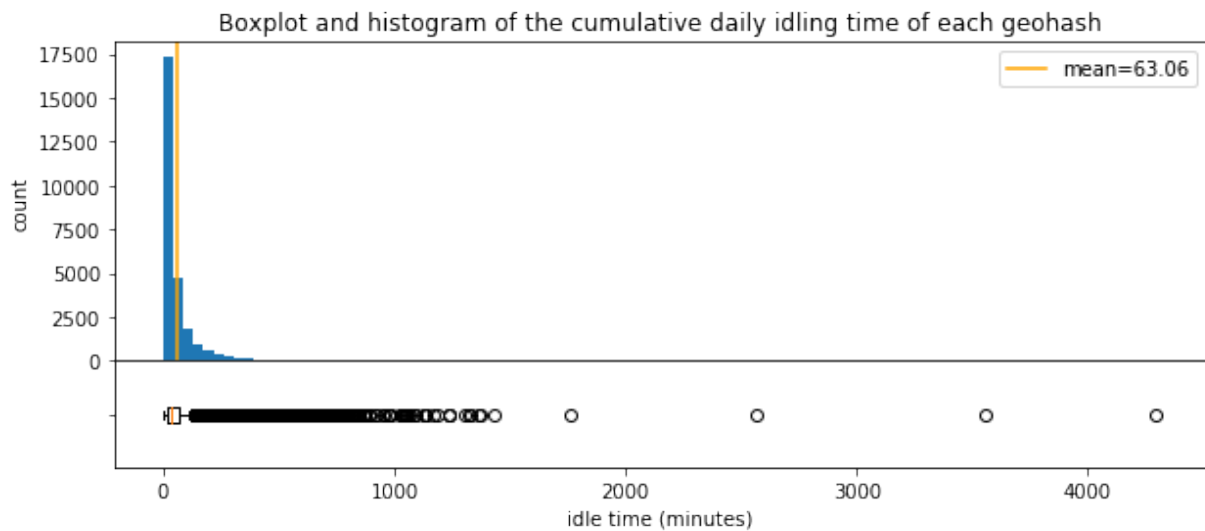
For the purpose of reducing idling time, cumulative idle time seems to be the most conclusive measurement. However, the total idling time is highly correlated with the traffic volume. Simply report the idling problem based on cumulative idle time is less meaningful. Therefore, I identify the following three idling problems.

problem of the area	data to look at	reasons for this measurement
long cumulative idling time	CumulativeDailyIdleTime	more intuitive and directive measurement of idling time
high proportion of idling (idling instances / all traffic)	IdlingByHour	identify areas where cars are more likely to idle
long idling time of idling instances	AverageIdleTime, MedianIdleTime	understand why cars idle longer in these areas

Notice that the data set only includes the hourly proportion of idling vehicles. To obtain a significant representation of the daily proportion of idling, I propose to take the mean proportion of the daytime (8:00-20:00). I choose to exclude the nighttime because there is much less traffic at night.

Then, I look at the distribution of the three factors through histograms and boxplot.





Although the distributions are quite different, the distributions are all heavily right-skewed. To identify the areas with idling problems, I propose to look at the areas from each measurement where the measurement is greater than the upper whisker (greater than $1.5IQR$ of the upper quartile). I choose this classification because such areas have measurements that are abnormally greater than the rest areas.

Areas with long average idling time in Toronto



Areas with long cumulative idle time in Toronto



Areas with high proportion of idle vehicles in Toronto



Areas with long average idle time and long cumulative idle time in Toronto



Areas with long cumulative idle time and high proportion of idling in Toronto



Areas with long average idle time and high proportion of idling in Toronto



Areas with all three problems in Toronto



As expected, most areas with high cumulative idle time are located in downtown and along major arterials, where the traffic volume are significantly greater than the other areas. However, the areas with high average idling time and areas with high idling proportion are spread through the city and has no obvious pattern. Also, there are no obvious patterns when I look at the areas with overlapped problems.

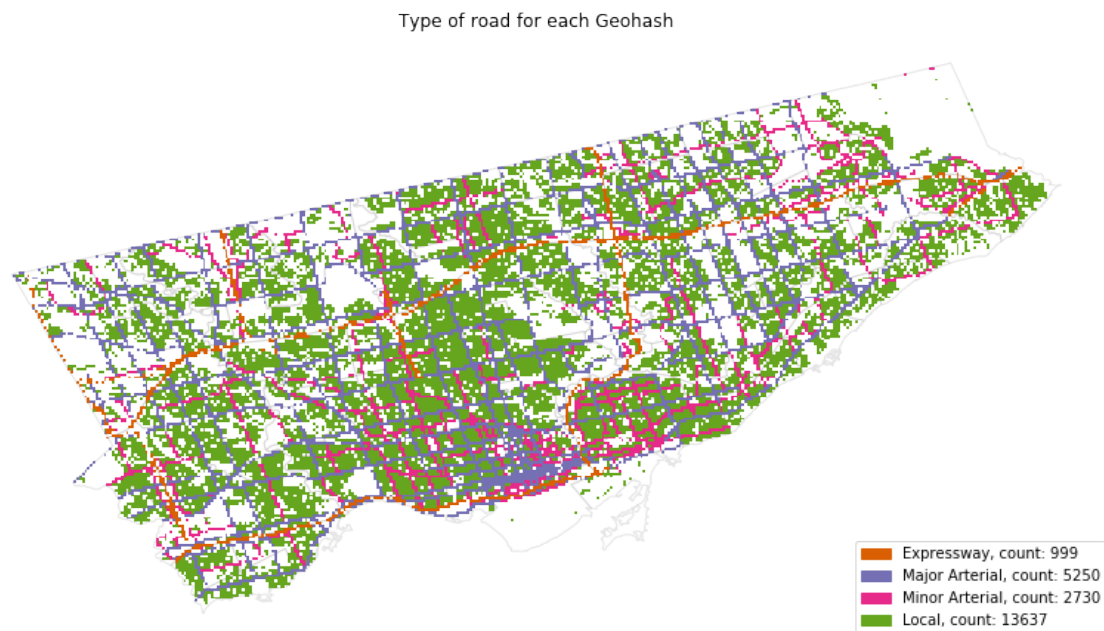
Therefore, we need further investigation with better categorizations.

IV. What is the relationship between idling problems and its locations.

4.1. Is there an association between the type of the road and the idling problems?

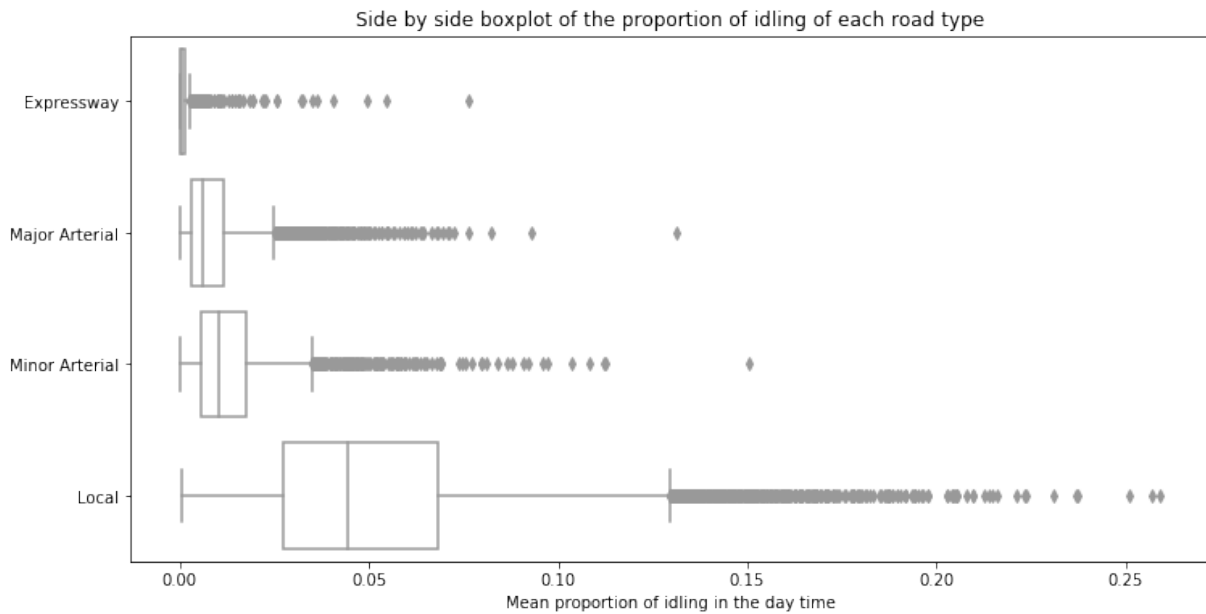
Since different road types handles different scopes of traffic within the city, they should have quite different idling issues. Therefore, I'm interested in the relationship between different types of road and their idling problems.

For the road types, I obtained road data from *Toronto centerlines*(TCL). TCL categorizes traffic paths and other linear features in Toronto into 26 groups. I'm interested in expressway, major arterial, minor arterial, and local roads. Then, I perform a spatial join between geohashes and interested roads and categorize geohashes by the most important type of road it intersects. The importance is defined from the most important to the least important as expressway, major arterial, minor arterial, local roads. The categorization is presented in the map below.



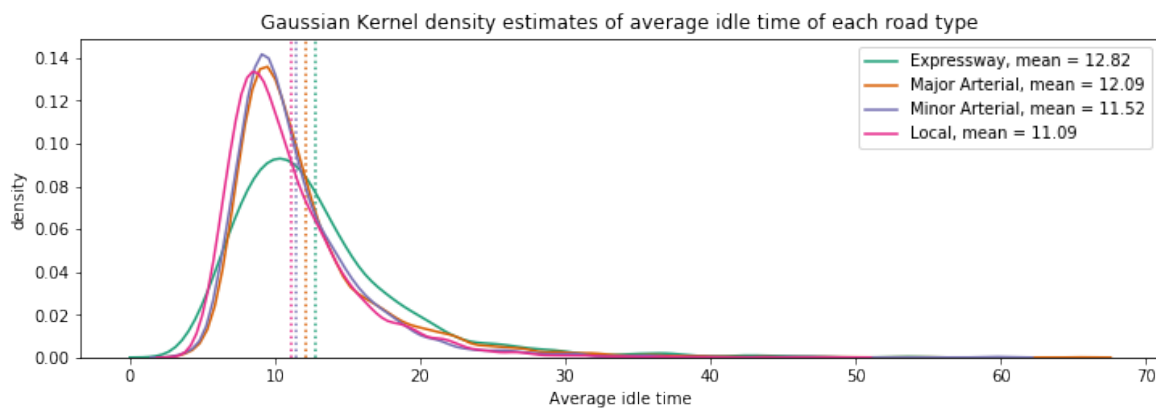
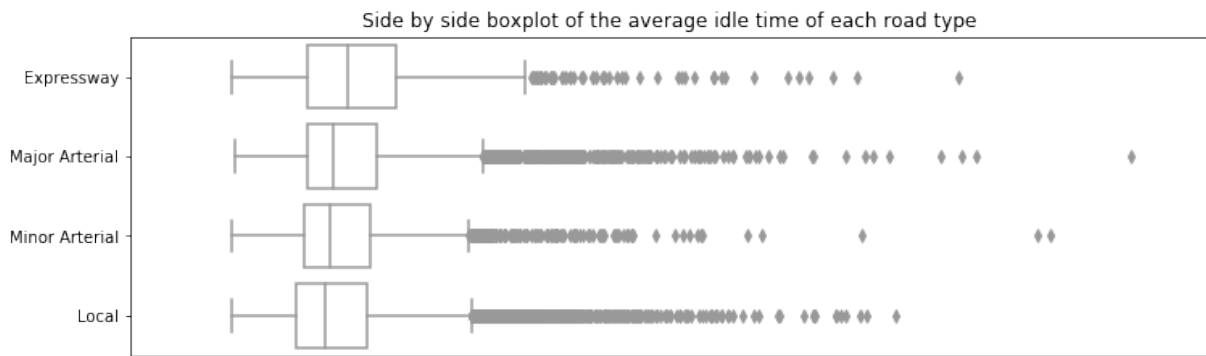
First, I look at the association between road types and proportion of idling.

I use side by side boxplots to investigate the distribution of the proportion of idling of each road type.



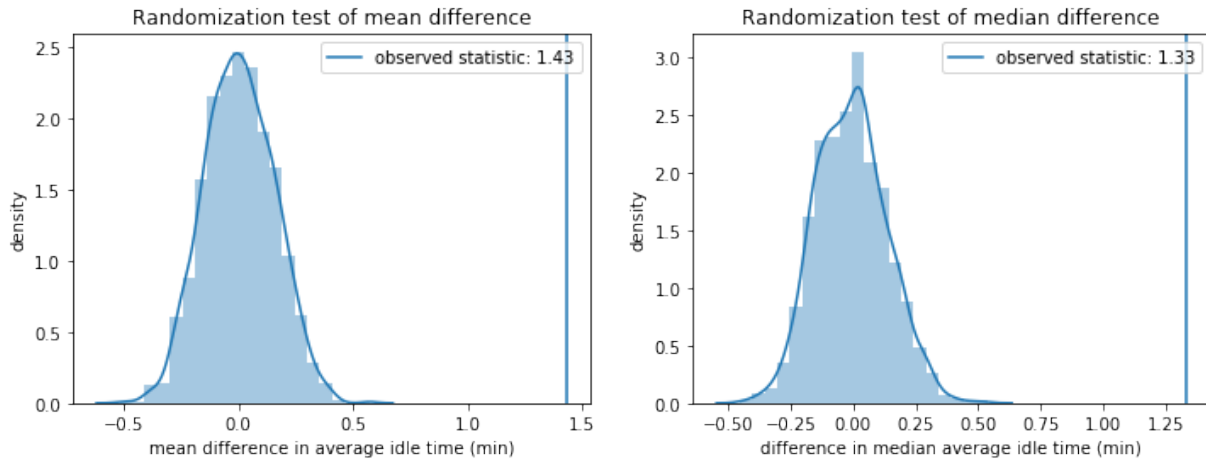
The difference is very obviously. The observations follow our common expectation, major roads have much smaller proportion of idling comparing to the local roads. Also, the variations of proportion for expressway is extremely small since idling is often unusual and dangerous on highways.

Then, I want to investigate the relationship between average idle time and the road types. I used side by side boxplot to examine the distribution of the average idle time for each road type.



There are only minor differences in mean and quartiles among different major arterial, minor arterial, and local roads. Also, the three road types have almost identical distributions. On the other hand, expressway has a different distribution, its mean and quartiles are also greater than the rest road types.

Therefore, I group the data into expressway and non-expressway and performed two-sample permutation tests. The null hypothesis is that there is no difference in mean and median in the two groups.

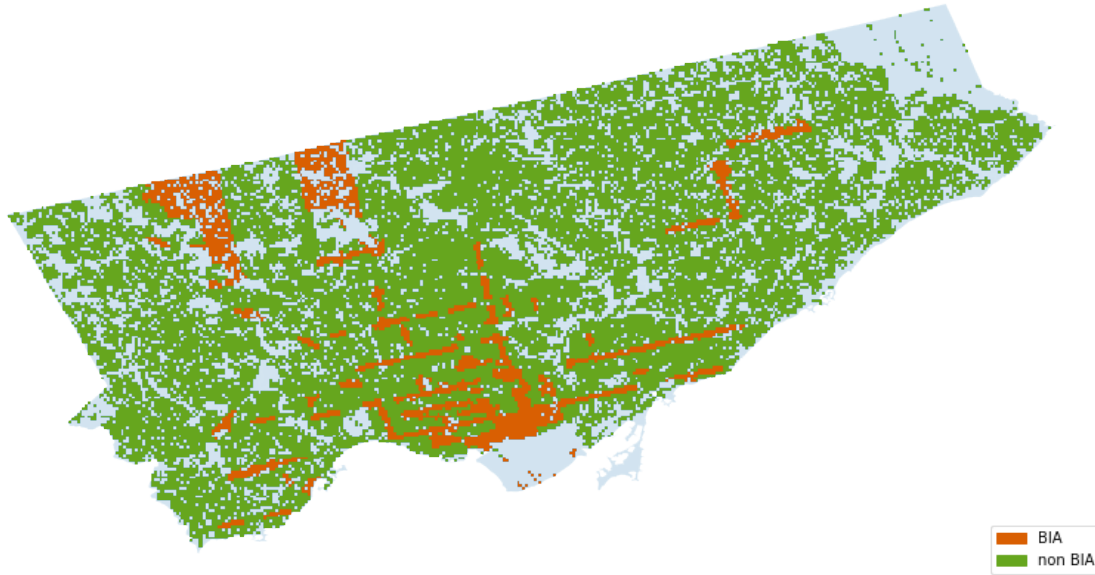


For both tests, the p-value is 0.0. We reject our null hypothesis. There is strong evidence that average idle time on the expressways is greater than the time on other road types.

4.2. Does idling happens more frequently in business areas or non-business areas?

Loading trucks, drop-in delivery, waiting for passengers are some common reasons for commercial vehicles to idle, and often these activities happen in industrial areas and business areas. Therefore, I'm interested in whether the idling problems is more common in business areas than the other areas.

To define business areas, I use business improvement areas from Toronto Open Data. These areas are often more concentrated with businesses, factories, and warehouses. Then, I join the geohashes with BIA boundaries, the geocoding is shown below.



First, looking at the issue of high proportion of idling. The areas with mean proportion of idling greater than the upper-whisker is considered to have idling problem. Then, I use the contingency table and perform a chi-square contingency table test. The null hypothesis is that having high idling proportion problem is independent of whether belonging to some BIA. The alternative hypothesis is that having high idling proportion problem is dependent of whether belonging to some BIA.

-	not BIA	is BIA	total
Don't have problem	23065	942	24007
Have problem	2754	148	2902
total	25819	1090	26909

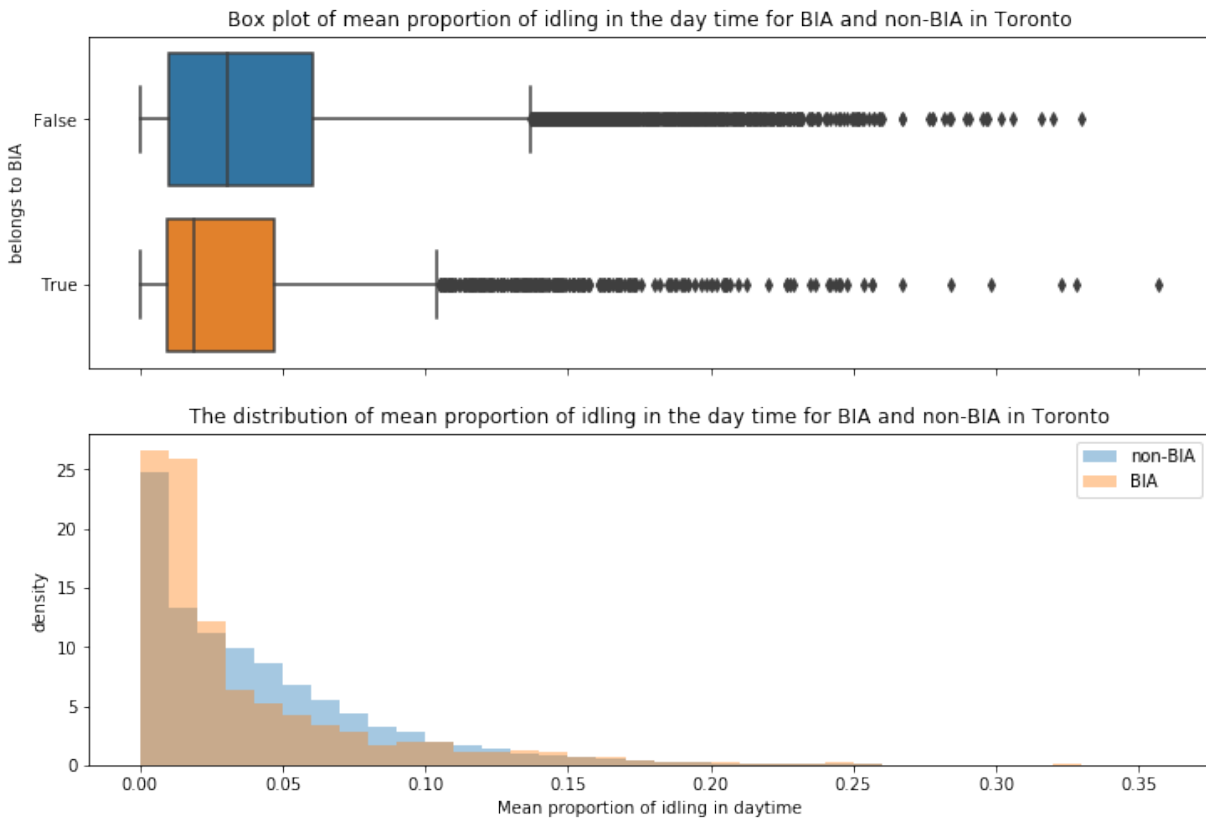
Number of geoshes for whether in BIA and whether having problem

-	not BIA	is BIA	total
Don't have problem	23035	972	24007
Have problem	2784	118	2902
total	25819	1090	26909

Expected values of the contingency table

The p-value is 2.40×10^{-2} , which is very small. Therefore, we can reject the null hypothesis. There is some evidence that having high idling proportion problem is dependent of whether belonging to some BIA.

Then, we look at the overall distribution of the proportion of idling in BIA and non-BIA geohashes.



Obviously, the distribution of the idling proportion for BIA is more concentrated to the left than the idling proportion for non-BIA. Overall, BIA geohashes has smaller proportion of idling than non-BIA geohashes.

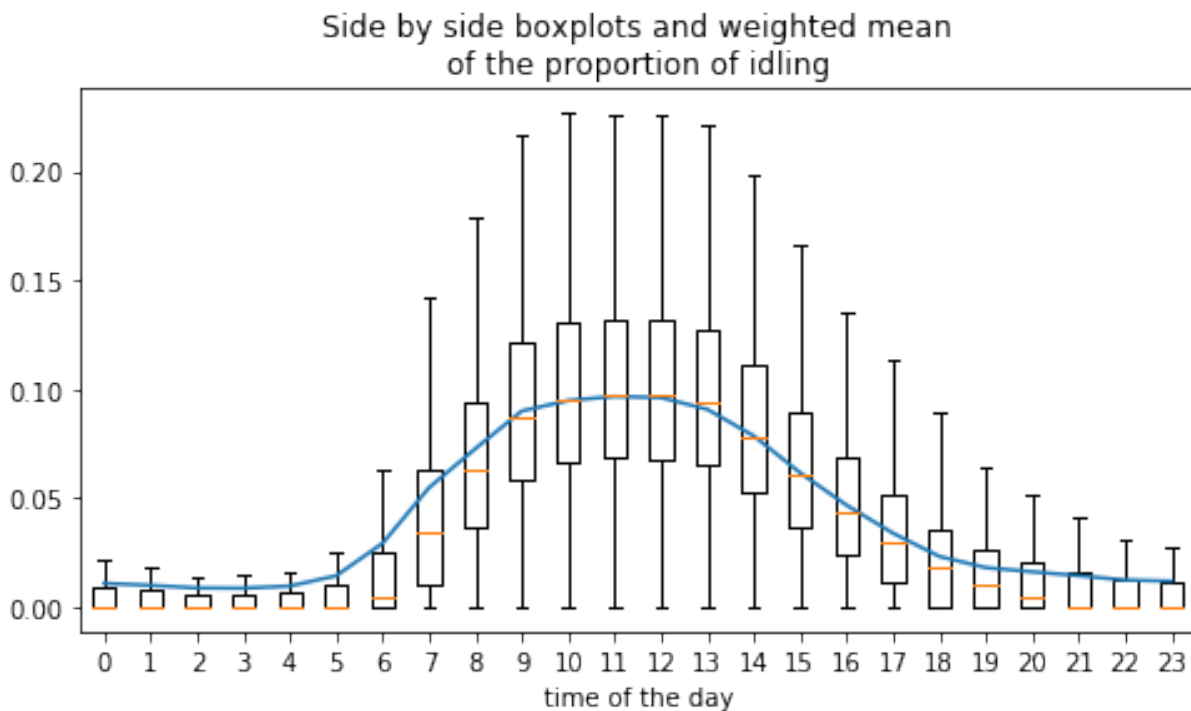
An interesting observation is that we observe significantly more problem areas in BIA than we expected (148 vs. 117.55), which the BIA geohashes are more likely to have high idling proportion problems. However, the overall distribution shows that BIA has overall smaller proportion of idling. One possible explanation is that the commercial areas in BIA have better regulations while the industrial areas in BIA have poor regulations. Therefore, the type of area need a more detailed categorization, rather than BIA vs. non-BIA.

V. What is the relationship between area of idling and the time of a day?

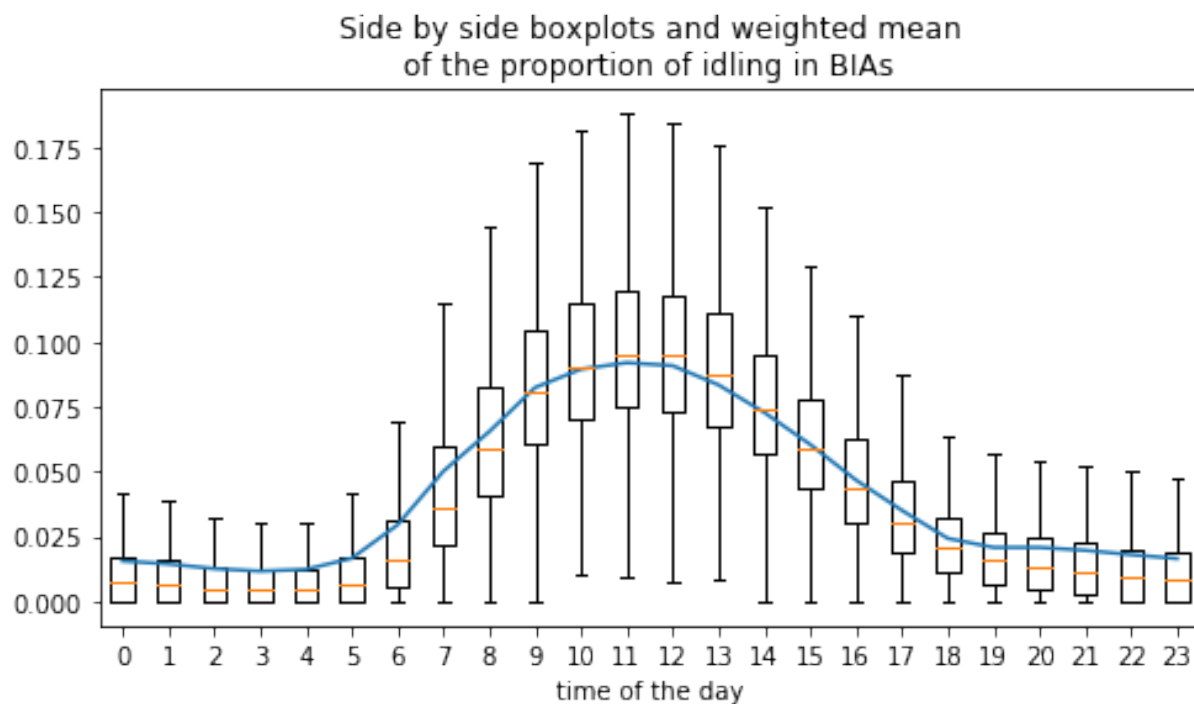
5.1. Does more idling happen in rush hours?

Because Geotab determines idling by checking whether the car moves, the traffic jam might be the reason, as lots of vehicles are clogged on the road. They may explain the areas with high cumulative idling time and low average idling time. We checked the traffic volume and average traveling time of different time in a day from *Congestion Trends in the City of Toronto*(5). Most traffic volume are in the morning (8:00 - 10:00) and evening (16:00 - 18:00) rush hours. I'm interested in whether hourly distribution of the traffic volume is correlated with the hourly distribution of the idling volume.

The first hypothesis is that the hourly proportion of idling should reach the peak on the rush hours. Since most of the vehicles will be clogged at that time, the proportion of idling should rise significantly. Therefore, I plot the boxplots for each hour and add the weighted mean proportion of idling, I use the cumulative idle time as the weight since the cumulative idle time is highly correlated with traffic volume, hence more likely to have a traffic jam.



Instead of two peaks at the rush hours, the peak is at the noon. Then, I suspected that the differences is caused by the large number of areas without much traffic congestions. Therefore, I narrow my sample to the BIA, which includes most of the commercial areas with higher traffic density.



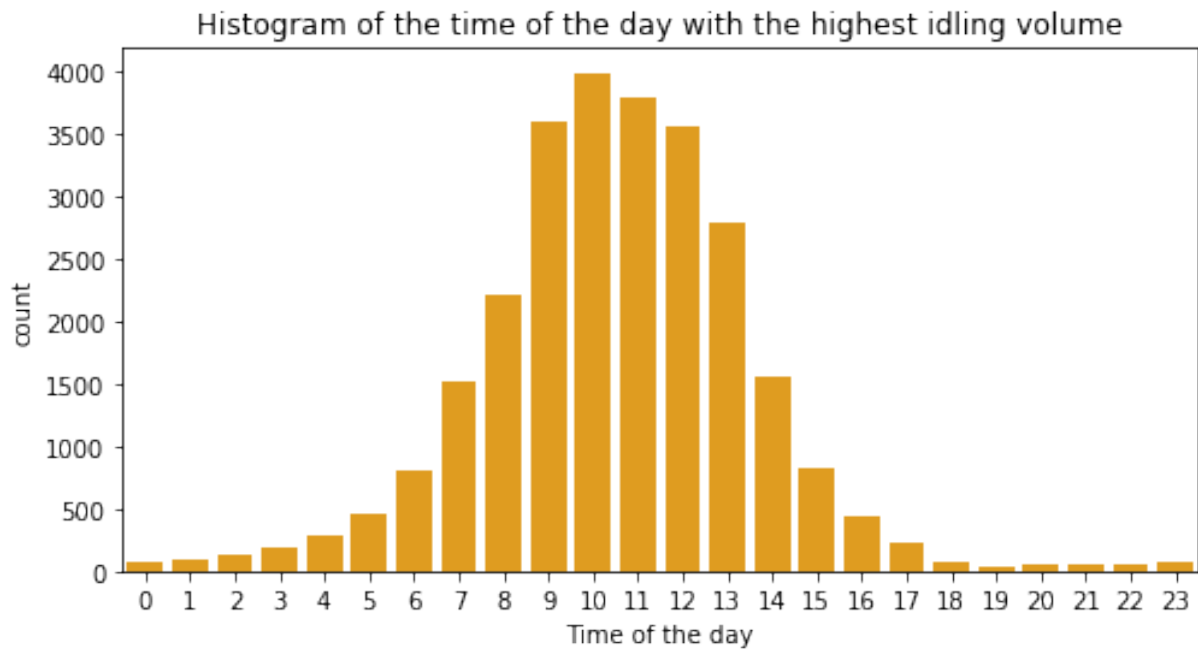
Although the peak in the BIA data is one hour prior to the peak in all data, the hourly distribution is still quite different from what we expected.

Therefore, there is no evidence that the idling problem is related to the traffic jams.

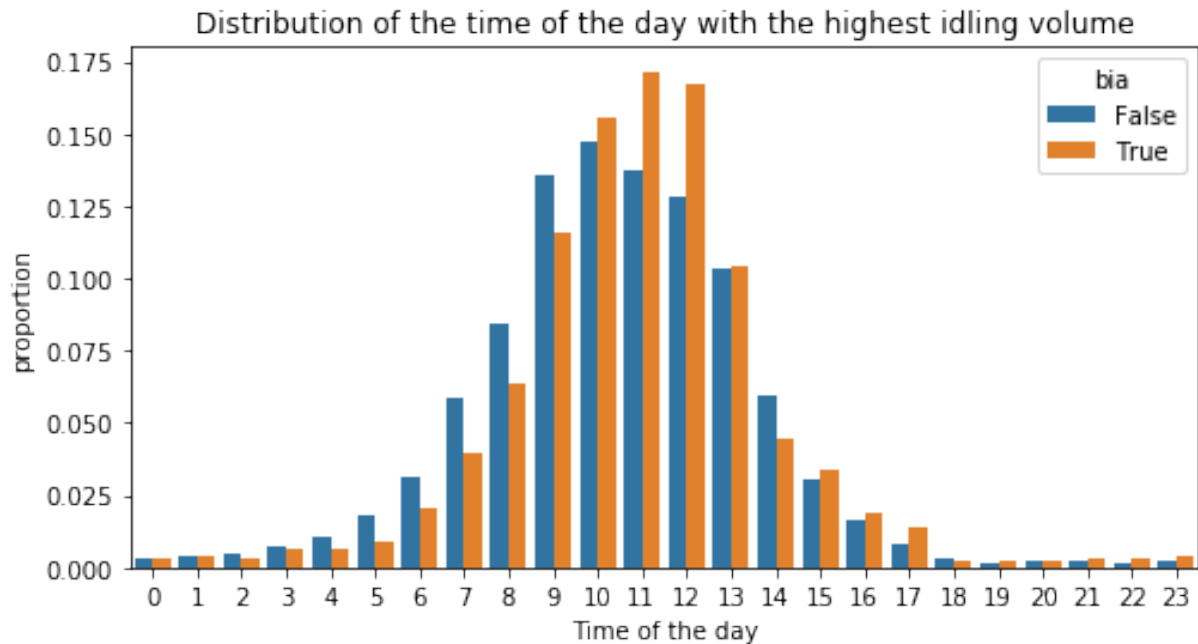
5.2. Is there a difference in hourly distribution of idling for different areas?

I'm interested in knowing when most idling happen for different places. With these crucial information, the city can apply efficient regulations for specific areas.

I started by finding the hour with the greatest idling volume in each geohash. By plotting them as a histogram, we can explore when most idling happens.



Then, I looked at two categorizations: road type and whether BIA, as from the previous part.



The non-BIA areas has its peak one hour prior to BIAs. There is some differences between the two, while the difference is not very significant. Then, I categorize the day into “rush time”, “lunch time”, and “working time”. I use the contingency table and chi-square test to determine whether the idling volume in different time intervals are dependent on whether the area is in some BIA.

-	BIA	non BIA	total
rush time (8-10, 16-18)	616	5876	6492
lunch time (11-13)	801	5777	6578
work time (10-11, 13-15)	884	7443	8327
total	2301	19096	21397

Number of geohashes with the greatest hourly distributed hour for each time interval and whether in BIA

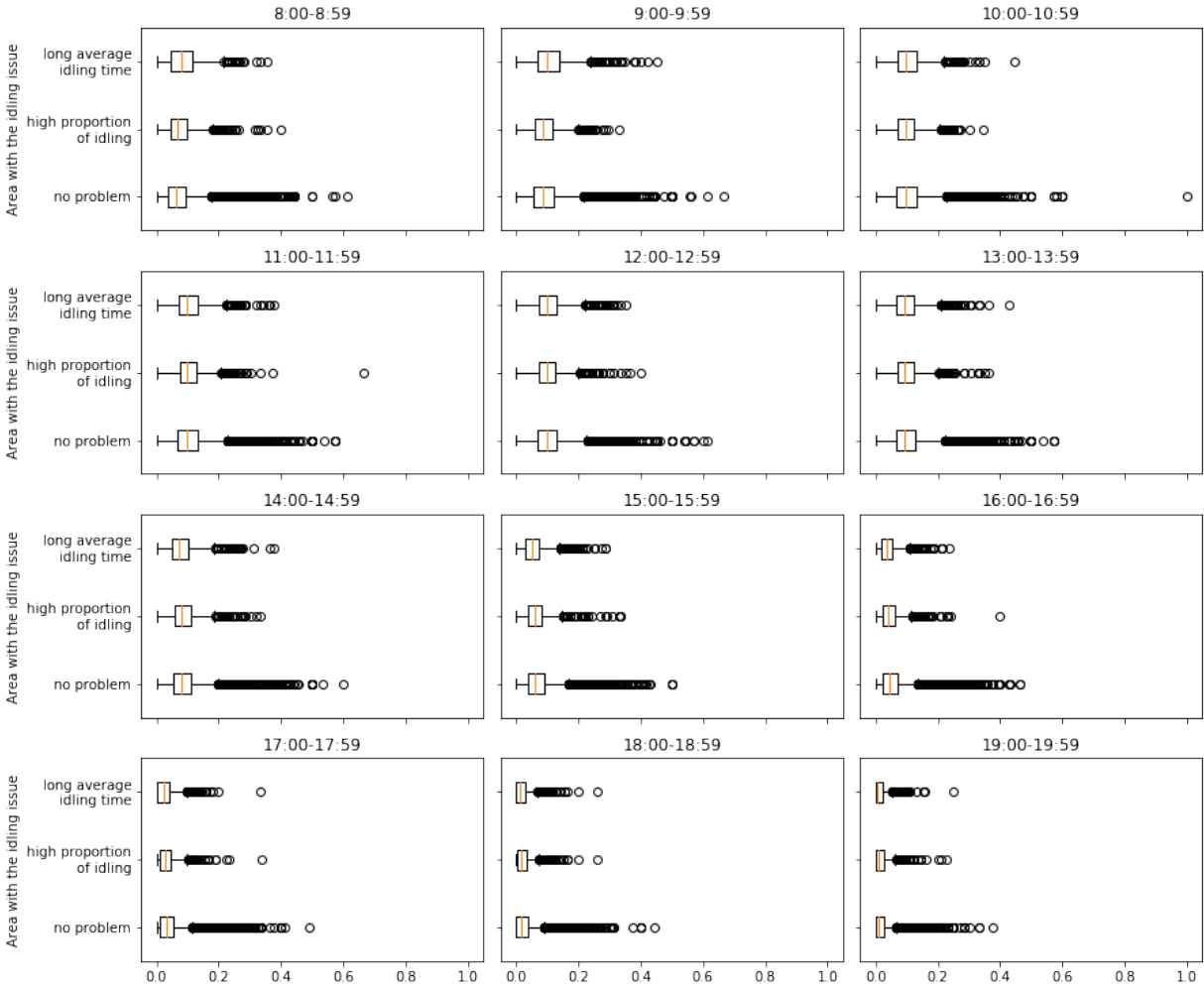
-	BIA	non BIA	total
rush time (8-10, 16-18)	698	5794	6492
lunch time (11-13)	707	5871	6578
work time (10-11, 13-15)	895	7432	8327
total	2301	19096	21397

Expected values of the contingency table

The p-value is 4.0×10^{-6} . Therefore, the idling volume in different time intervals are dependent on whether the area is in some BIA.

Then, I'm interested in whether the areas with idling problems have a distinct time distribution. For investigation, I plot the boxplot of the hourly proportion of idling volume from 8:00 to 20:00.

Boxplots of the hourly proportion of idling volume in areas with different problems



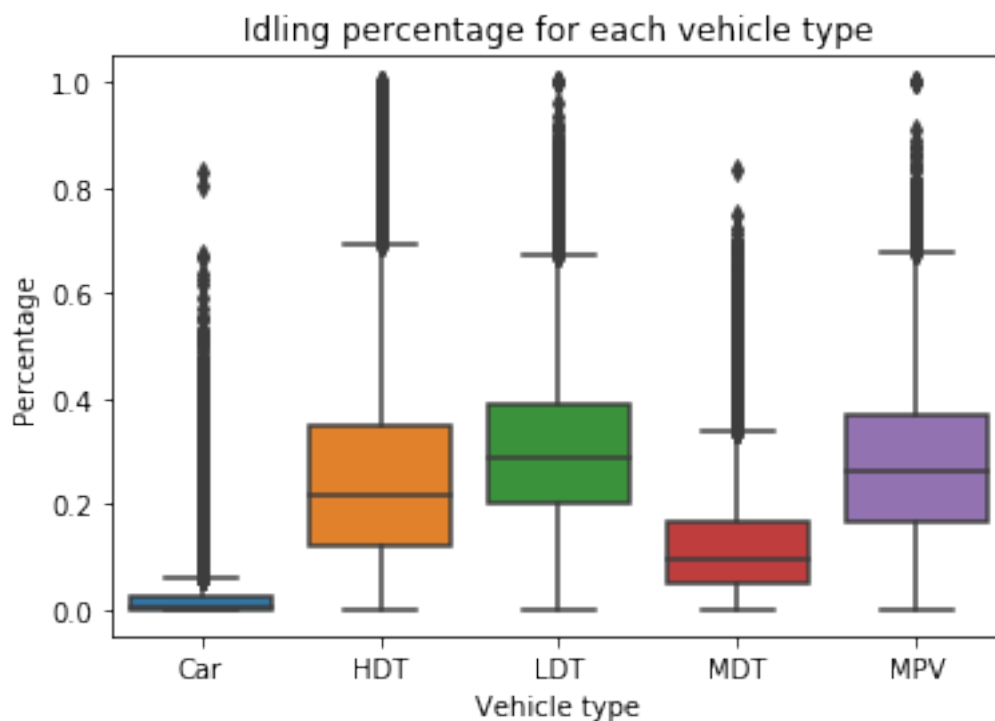
From the boxplots, the distribution for the three types of areas are almost identical for each hour. The summary statistics are so very close. Therefore, the observation suggests that the hourly distribution of idling volume is independent of whether there is some idling problems associated with the area.

VI. What is the relationship between the types of vehicle and idling issues?

6.1. Does the type of vehicles associated with different kinds of idling problems?

Since idling can be caused by various reasons, we figured that types of vehicle might play a role as well. In our dataset, there are two measurements for types of vehicle, the first one is the percentage of this type of vehicle in a specific geohash (note that this is `idling_type/idling_all`), and the second one is the average idle time for each type of vehicle. The percentage gives us a sense of whether a type of vehicle is a major issue in some areas, and the average time indicates the extent of idling issues.

To get a general idea of what the data look like, a boxplot of all types of vehicles with their percentage are showed below. From the plot we can see that, type car has relatively low percentage of idling in each geohash, and the amount of data for type car only compose a very small portion of our data. So we only will further consider the other four vehicle types in our discussion. Moreover, for each type of vehicle, there are some points with extreme high percentage, so first we want to explore these data points a little bit more.



And before we examine the pattern of those outliers, we want to take a closer look of the usage of each type of vehicles. - Heavy duty truck (HDT): transportation of large cargo - Light duty truck (LDT): light transportation such as small furniture, small equipments - Medium duty truck (MDT): carry moderate amount of cargo, some examples such as: delivery truck, garbage trucks - Multi-purpose passenger vehicle (MPV): vehicle that can be used in multiple ways but with up to 10 people

To have a better idea of those outliers, we categorized them based on road types they idle on and plot the outliers in our maps. We can see a pattern that west part of Toronto is mostly coloured in green, which represents HDT. And if we go to the east part of our map, more and more purple and pink dots are present, which indicate MDT and MPV respectively. On second thoughts, Etobicoke, the west part of Toronto, has lots of truck services companies, warehouses, truck driving schools, which might explains the high percentage of HDT. And for the east part of Toronto, most of the areas tend to have neighborhood

setting, and that might be an explanation of why daily life related vehicles have higher percentage there.

High idling on expressway for each type of vehicles



High idling on major arterial for each type of vehicles



High idling on minor arterial for each type of vehicles



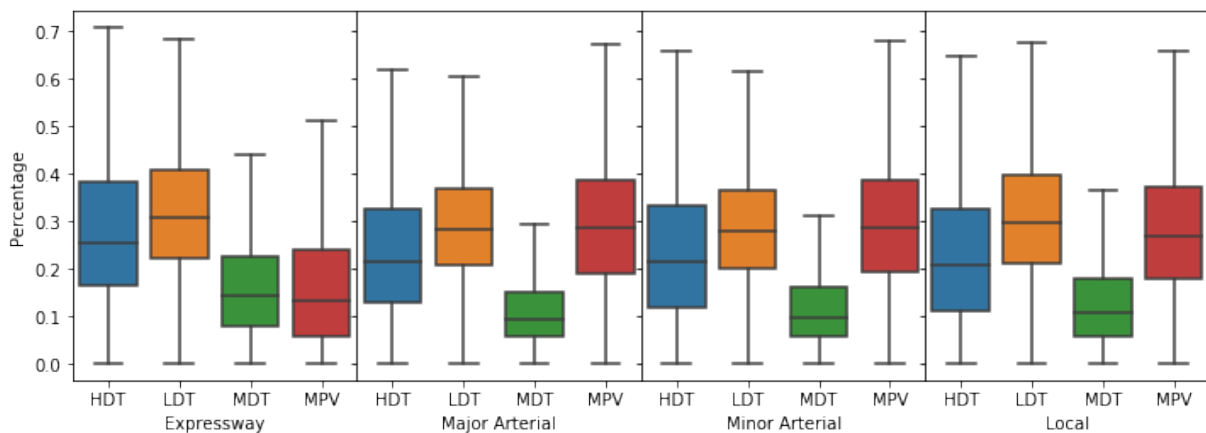
High idling on local for each type of vehicles



6.2. Does the effect of road types on idling problems differ with vehicle types?

After exploring the outliers of percentage, we plot the data in boxplot again, but in different road types to see if the distribution of percentage of vehicle types is independent of road types. And if that is the case, we do not need to consider road types factors later on.

Percentage of idling instances in each vehicle type on each type of roads



From the graph, we can tell that the distribution between major arterial, minor arterial and local areas are similar, while expressway seems a bit different. So we decided to count larger than 10% as a idling problem, and count the frequency of idling problem for each type. After that, we can do chi-square test.

-	HDT	LDT	MDT	MPV
expressway	892	961	662	611
local	10666	13071	7247	12591
major arterial	4363	5109	2463	4821
minor arterial	2214	2597	1340	2512

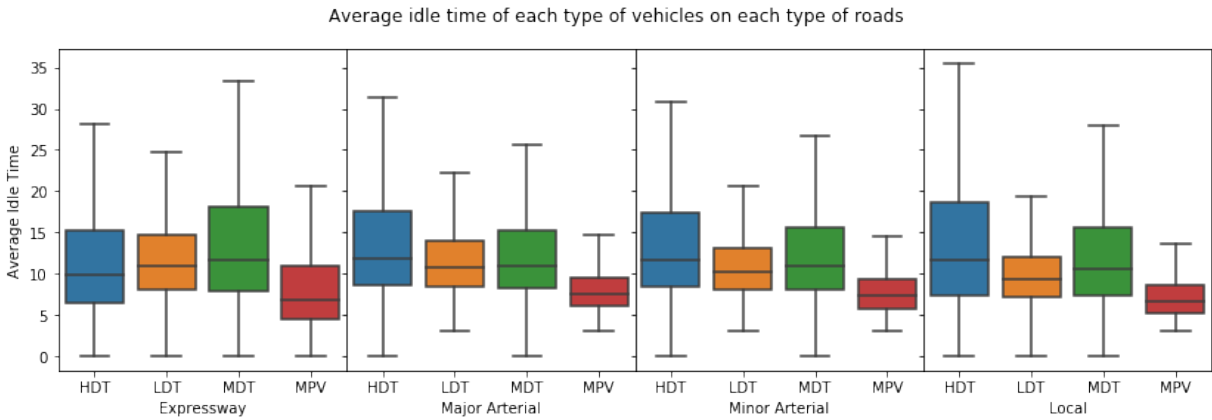
Number of areas with idling problems for each vehicle type on each road type

-	HDT	LDT	MDT	MPV
expressway	786	942	508	890
local	10957	13134	7076	12407
major arterial	4213	5051	2721	4770
minor arterial	2178	2611	14067	24667

Expected values of the contingency table

The p-value for this test is 0.0, so we reject the null hypothesis at 95% interval. And the result makes sense because each road type has different settings, and is designed for different usage. Therefore, there is some relationship between road types and vehicle types.

After examining the percentage data, we want to know if the average idle time distribution for each vehicle type is similar among different types of roads. Therefore, we plot the data in a boxplot again and try the chi-square test. But this time, indication of idling problem is time greater than or equal to 10 minutes.



-	HDT	LDT	MDT	MPV
expressway	486	564	618	292
local	8046	5690	7320	2107
major arterial	3298	3009	3026	1113
minor arterial	1713	1423	1565	566

Number of idling problems for each vehicle type on each road type

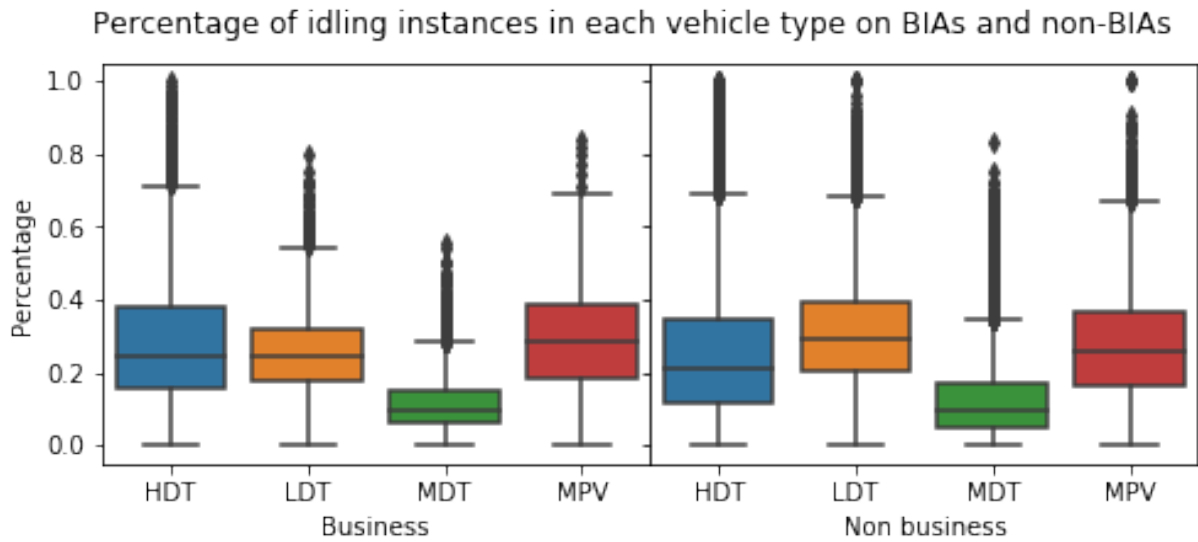
-	HDT	LDT	MDT	MPV
expressway	650	513	601	196
local	7682	6061	7107	2313
major arterial	3464	2734	3205	1043
minor arterial	1747	1378	1616	526

Expected values of the contingency table

P-value for this question is also 0, which indicates that we reject our null hypothesis at 95% confidence interval. From the expected values of the contingency table, we can see that all types of cars have lower than expected idling on expressway, and this makes sense because it is dangerous for cars to idle on expressway.

6.3. Does the effect of BIA on idling problems differ with vehicle types?

After considering road types, we want to see if business area category also plays a role when it comes to vehicle types. We first consider the percentage and do the chi-square test.



-	HDT	LDT	MDT	MPV
not-BIA	19027	22733	11768	21190
BIA	2608	2720	1397	2570

Number of idling problems for each vehicle type on different areas

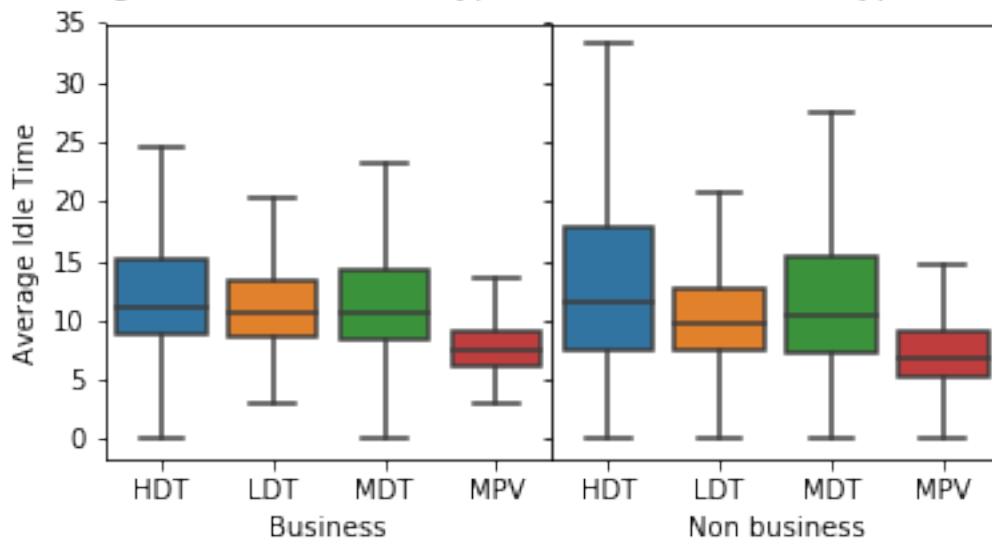
-	HDT	LDT	MDT	MPV
not-BIA	19241	22637	11708	21131
BIA	2394	2816	1457	2629

Expected values of the contingency table

1.77×10^{-6} as a p-value makes us reject our null hypothesis that percentage of each vehicle type on business and non-business area are independent. Notice how HDT idle more than we expected in BIA, and this can be explained by some parts of BIA contain lots of warehouses.

Same test is carried out with respect to average idle time as well.

Average idle time of each type of vehicles on each type of areas



-	HDT	LDT	MDT	MPV
not-BIA	14102	11199	12657	4407
BIA	1782	1634	1651	551

Number of idling problems for each vehicle type on different areas

-	HDT	LDT	MDT	MPV
not-BIA	14024	11330	12633	4378
BIA	1860	1503	1675	580

Expected values of the contingency table

This time our p-value is 0.0003, and we reject our null hypothesis at 95% confidence level.

VII. Conclusion

After the observation of several factors, we found some points that might be helpful when studying how to reduce idling time. First of all, idling happens mostly in local, small areas. This can be explained by the fact that a lot of local services are done with the help of vehicles and people usually do not turn off the engine while making a short stop. Secondly, although business areas have overall small proportion of idling, idling problems are more likely to happen in business areas. And whether or not a place is categorized as business area also affects the percentage of vehicle types that idle there. What is more, different areas and road types also play important roles in idling problems. Some specific type of commercial vehicles tend to idle at the areas that they are used (i.e., loading cargo, collecting garbage, distributing parcels) and different settings (i.e., capacity) of road types can also have different idling issues. Last but not the least, although we thought traffic might be one of the major factors that affect idling time, our observation result does not align with our hypothesis but vehicles do idle more at some particular time of a day - late morning and noon.

With these conclusion in hands, we think there are a couple of things that we can do to reduce the time of idling in Toronto, and they are all related to how to reinforce the regulations. Because in Toronto, idling more than one minute can be charged \$125 in fines. Now that we know some patterns about idling, we can assign inspection team to finish the task in a more efficient way. To start, more people can be assigned at business areas at late morning and noon to check if vehicles obey the rule. Meanwhile, truck services and driving schools definitely need more attention because truck has more pollution to the environment when it comes to idling. Moreover, as some of vehicles that idling belongs to city services, such as garbage trucks, government can take the lead to reduce idling time and reward companies that are willing to put effort in it, we believe more and more individuals will be aware of this issue.

VIII. Limitations

There are several limitations about this analysis. This dataset only contains commercial vehicles that use GEOTAB and that cannot be counted as a good representation of all commercial vehicles in Toronto. And the amount of data that are provided for different types of vehicles are not balance. Car has relatively a very small portion of data in the dataset, so we cannot actually study the trend for type car. On the other hand, idling can be affected by lots of factors, and different factors play different roles in idling issues. Thus, the number of factors we considered in this analysis is far from enough for us to generate a regression model to predict the idling time of certain type of vehicle at certain area. And details classification of each category, for example, subcategories for non-business areas, might also be helpful in future study.

References

1. Geotab. *What is Geotab's Telematics Data?*.
<https://data.geotab.com/our-data>
2. City of Toronto. *Toronto Centreline (TCL)*.
<https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#e4ec3384-056f-aa59-70f7-9ad7706f31a3>
3. City of Toronto. *Business Improvement Areas*.
<https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#2492fc58-6ae0-fa22-aecf-6e7b5392494b>
4. City of Toronto. *Neighbourhoods*.
<https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/locations-and-mapping/#a45bd45a-ed8-730e-1abc-93105b2c439f>
5. McMaster Institute for Transportation and Logistics. *Congestion Trends in the City of Toronto (2011-2014)*.
https://www.toronto.ca/wp-content/uploads/2017/11/91f0-city_congestion_trends_09-15-2015.pdf