Greenhouse gas emissions and its related factors for public agencies in Toronto*

Paper 1

Qiao Zhou

05 02 2022

Abstract

Annual energy consumption datasets are important tools for government to reduce GHG emissions and to make forcasting, however biases in the way data are collected may affect the reliability of this information.This report utilizes data on annual energy consumption in Toronto to:(1) discuss the poential bias and limitation if our data. (2)figure out which factors are strongly related to the GHG emissions. According to the data and the analysis, the usage of electricity and gas, the area, and the operation type are clearly related to the levels of GHG emissions with the energy (gas/electricity)consumption being the great GHG emissions contributors. However, as our data being not complete and representative, our conclusion may get less accurate and it will be tough for us to get a good eatimation using our data.

*Code and data are available at: https://github.com/kzhou1999/starter_folder.

1 Introduction

A new report from one of the world's most prestigious medical journals says Canada's failure to cut greenhouse-gas emissions isn't just killing the planet; it's killing Canadians.("Greenhouse-Gas Emissions Killing Canadians: Medical Journal" 2018) In 21 century, greenhouse gas pollution has become a very serious issue worldwide as it directly causes the climate change.And the scientists even suggests that successfully tackling greenhouse-gas emissions would be the single biggest thing governments can do to improve human health this century.("Greenhouse-Gas Emissions Killing Canadians: Medical Journal" 2018)

First, I would like to introduce what is greenhouse gas emissions and why we should take this issue seriously.Greenhouse gases are gases which are emitted primarily by burning fossil false, sold waste, and wood products.It will trap heat in Earth's atmosphere and prevent the heat that the sunlight brings from leaving the atmosphere. ("Climate Change Indicators: Greenhouse Gases" 2021)The greenhouse gas is harmful because it will cause worldwide climate change, and also increase people's death rate from heart disease and lung cancer. The high frequent heat waves will cause more forest fires as well.("Greenhouse-Gas Emissions Killing Canadians: Medical Journal" 2018) Canada has a high level of GHG emissions, therefore Canadian government takes this issue seriously. In 2009, the Ontario government promoted the green energy act. Starting from 2009, all the public agencies should publish their annual energy consumption and greenhouse gas emissions for their operations. The annual energy consumption data has become a important tool for government to reduce the amount of energy consumed by the public agency's operation and to do a forecast of the expected results of current and proposed measures as well.("Law Document English View" 2018)

Giving the importance of the annual energy consumption dataset, in this report, I would like to use the 2018 "Annual Energy consumption and GHG emissions" dataset collected by the Ontario government ("Annual Energy Consumption Data-2018" 2018) to analyze what are the largest contributors to the national GHG emissions.Based on a report posted in 2018, the largest source of GHG emission in Canada comes from the consumption of fossil fuels to make energy, including heat and electricity. Mining, oil and gas, and manufacturing are collectively responsible for the largest slice of this pie, followed closely by houses, shops, schools and other private and public buildings.("Where Do Canada's Greenhouse Gas Emissions Come From?" 2018)So my assumption is that the use of electricity and gas, the purpose for public agencies,the total floor area, the average opening hour are strongly related to GHG emissions. For the purpose of public agencies, I choose the six most common types in Canada, which are fire stations, public libraries, storage facilities, administrative offices,ambulance stations, and community centers.

In my report, I am going to introduce the data and the main methodology for analyzing the data and checking out my hypothesis. What's more, I will talk about the potential bias and the limitations for my data. Then, I will be showing some numerical/graphical analysis . The dataset will be processed and analyzed in R (R Core Team 2020) primarily using the tidyverse (Wickham et al. 2019) and dplyr packages(Wickham et al. 2021). Figures and tables will be created with ggplot2(Wickham 2016) . The packages knitr (Xie 2014), bookdown (Xie 2016) are used to generate the R markdown report.

2 Data

2.1 Data Source

The data frame I am using in this report is the "Annual Energy consumption and GHG emissions" in 2018.("Annual Energy Consumption Data-2018" 2018) This dataset is collected by the Ontario government. It was obtained in xlsx format from the City of Toronto Open Data Portal using the R package opendatatoronto (Gelfand 2020). Toronto Open Data is an open data initiative by the City of Toronto government in Toronto, Ontario, Canada.The dataset was last updated on May 12, 2020.

2.2 Methodology and Data Collection and Potential Limitations

The collection process of this data frame is straightforward. In the green energy act, the Ontario government list out the target population of their survey and expect responses from them. Here is all the target population: 1. Every municipality. 2. Every municipal service board. 3. Every post-secondary educational institution. 4. Every public hospital. 5. Every school board. ("Law Document English View" 2018) Then, these above public agencies should send back all the information to the government and also posted on their website. However, this dataset has a few limitations. One obvious limitations for this data is the sample size of our data frame is too small. This might due to the collection process of the data. Since data was collected through voluntary survey, those people or agencies who do not response to the survey won't have any penalties. So, in fact, our data is not complete and representative at all. Also, if we use this data frame to estimate the GHG emission of remote Canadian cities, the result may not be accurate as well. Because geographical location, climate, and technology differences will also lead to different levels of GHG emissionss, however, those factors are not included in our data. What's more, we know that the area of the public agencies and the average opening hour are closely related to their operation type. This may lead in misinterpreting of which factor is the actual GHG emissions contributor. For example, if we got a conclusion that the administrative offices has the greatest level of GHG emissions in overall.It's hard for us to determine if it is simply because of the operation type or because of the administrative offices on average have greater area or longer opening time.

2.3 Data Characteristics

The original data has 36 variables in total and has 1490 variables. This data frame recorded the energy usage of all public agencies in Toronto as well as the purpose of these facilities, floor space, and opening hours (per week) (Gelfand 2020) I did some data cleaning in order for me to only work with the data I need. First, I eliminate all of the missing values in the data. Then, I convert the variables that contain numbers into numerical class. For the categorical variable, I convert them into factors. I created a new dataset called "annual_energy_consump_2018." Then, I selected six target variables from the original data frame, renamed and stored them into this new data frame. Since I am only interested in 6 types of the public agencies, I use the filter function to select observations that has these operation types. Thus, the final clean dataframe has 390 observation and 6 variables;5 variables are numerical and 1 is categorical.

The defintion of variable is shown in the below table.

Table 1: Definition of Variables

Variables	type	Definition
area	num	The area of public agencies(in square feet)
use_of_elec	num	Consumption of electricity in one year (in kWh)
use_of_gas	num	Consumption of naurtal gas in one tear (in cube meter)
operation type	chr	The function of public agencies
average_hour	num	Average opening hours per week
GHG	num	The total amount of GHG emision in one year (in kg)

A sample view of the dataset is displayed below.

```
## Rows: 390
## Columns: 6
## $ area          <dbl> 46145, 10053, 39590, 43002, 18299, 34843, 780061, 8530,~
## $ use_of_elec   <dbl> 1259536.64, 70947.46, 480158.21, 105617.73, 181378.17, ~
## $ use_of_gas    <dbl> 34119.045, 5120.625, 46939.289, 26977.436, 21082.355, 4~
## $ average_hour  <dbl> 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70,~
```

```
## $ GHG            <dbl> 115197.074, 12565.499, 108376.427, 55473.475, 47305.066~
## $ operation_type <fct> Administrative Offices, Administrative Offices, Adminis~
```

The distribution of our numerical variables is shown in the below table.

Table 2: Summary for Numerical variables

type of score	Min	1st Qu	Median	Mean	3rd Qu	Max	Sd
area	183	5713	11001	29874	24830	787186	78381.25
use_of_elec	5927	61117	127058	438070	342497	13448190	1157231
use_of_gas	0	7889	17697	37274	36563	692756	67429.4
average_hour	70	70	100	106.6	168	168	42.38682
GHG	270.8	20597.7	39869.7	100328.2	84768.3	2513373.2	223326

3 Discussion

3.1 Energy usgae(electricity and gas) and GHG emissions

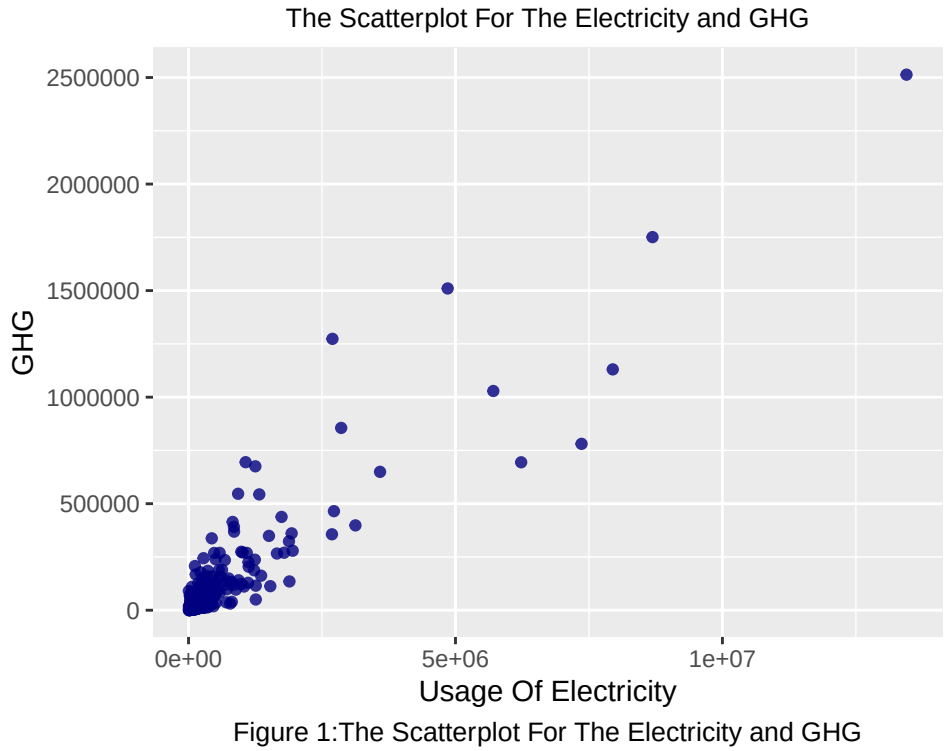

Figure 1: Figure 1:The Scatterplot For The Electricity and GHG

Let's take a look at the figures. The first plot (Figure1) shows the relationship between the use of electricity and the GHG emission level. The x axis represents the consumption level of electricity, and the y axis represents the GHG emissions. Overall,we can see these two variables are positively correlated; as the use of electricity increase,the GHG emission increase strongly. However, their correlation is not very strong. We can see the plots on the graph located closely at first. However ,the points spread wider as the electricity usage increased. The second plot (Figure 2) shows the relationship between the use of gas and

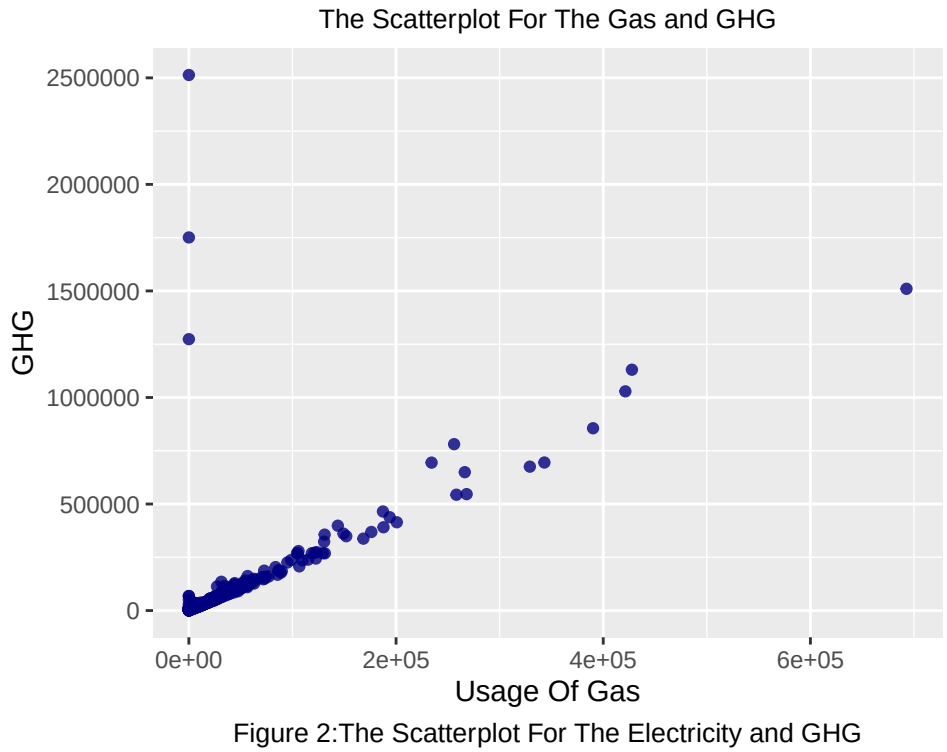

Figure 2: Figure 2:The Scatterplot For The Gas and GHG

the GHG emission level. By observing this graph, we can see these variables have a positive relationship and correlated strongly. All the points locate closely with others. In conclusion,the usage of both types of energy has a positive relationship with GHG emissions, which meets my assumption.

3.2 Purpose for agencies, area , average opening hour and GHG emissions

Now, let's take a look at the other figures. Figure 3 is a scatterplot for the area and the GHG emissions. It shows a positive relationship between the area and the GHG emissions. This outcome is not hard to understand at all, because larger area will naturally requires greater energy consumption, which will leads to higher level of GHG emissions. For Figure 4, it's a scatterplot for the average opening hour and the GHG emission. On the contrast, it does not show any clear relationship between the average opening hour and the GHG

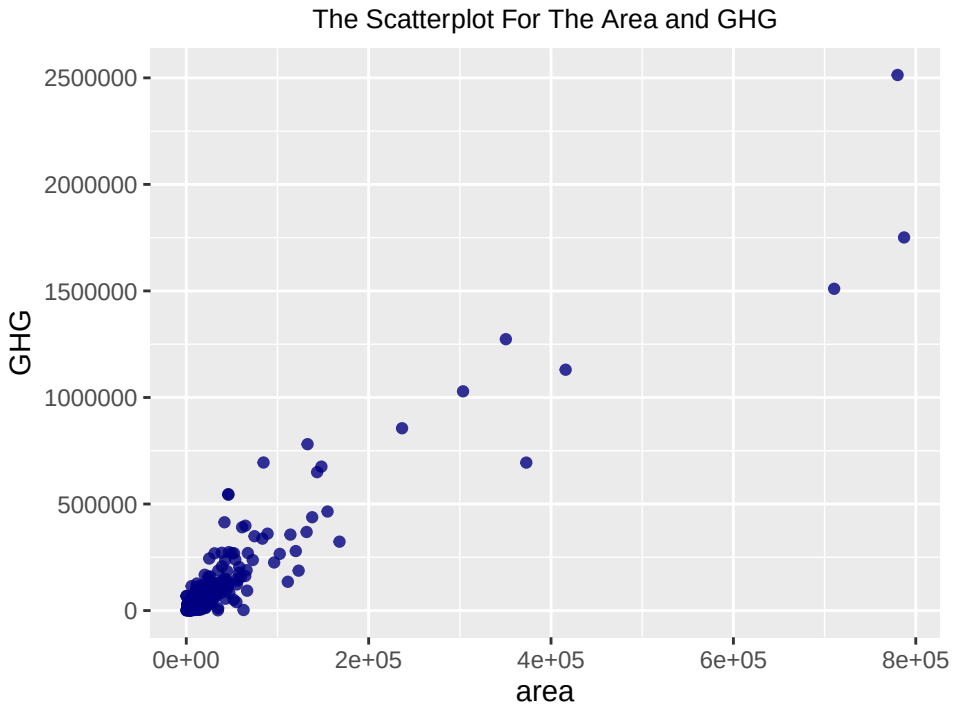

Figure 3:The Scatterplot For The Area and GHG

emissions as I assume./

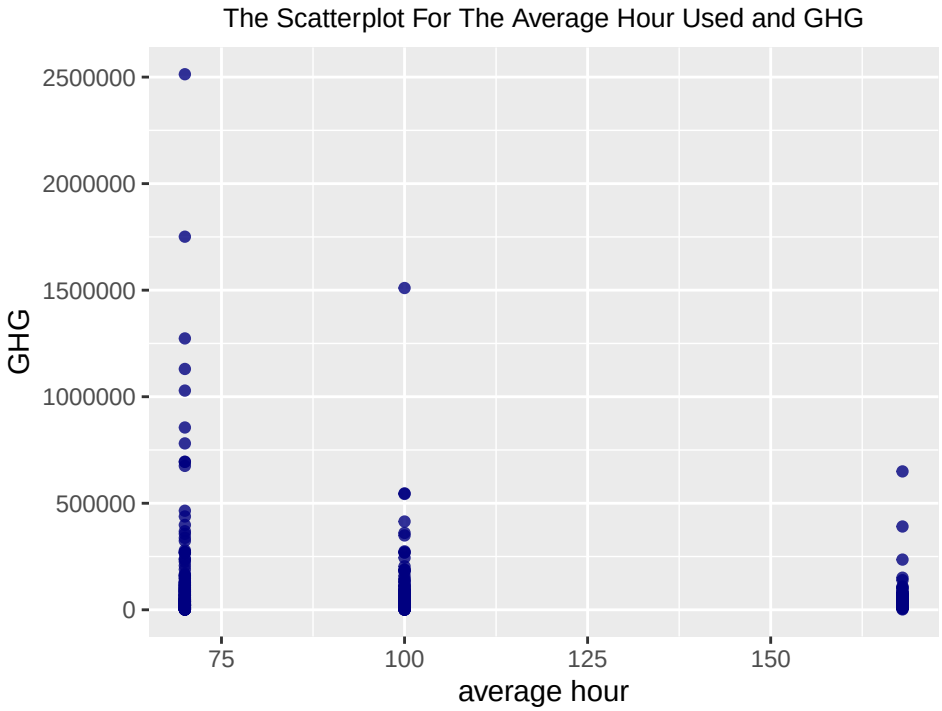

Figure 4:The Scatterplot For The Average Hour Used and GHG

However,as I have mentioned earlier that the area and the average opening hours are closely related to the purpose of the agencies. Since the average opening hour does not affect the GHG omissions much, we only consider area and the purpose of the agencies in this situation. In order to investigate how are these two factor related to each other and what does the relation between them impact the GHG emissions, I

draw another two plots in the following. Figure 5 shows the distribution of area grouped by the purpose of agencies. Figure 6 shows the GHG distribution for each operation type.

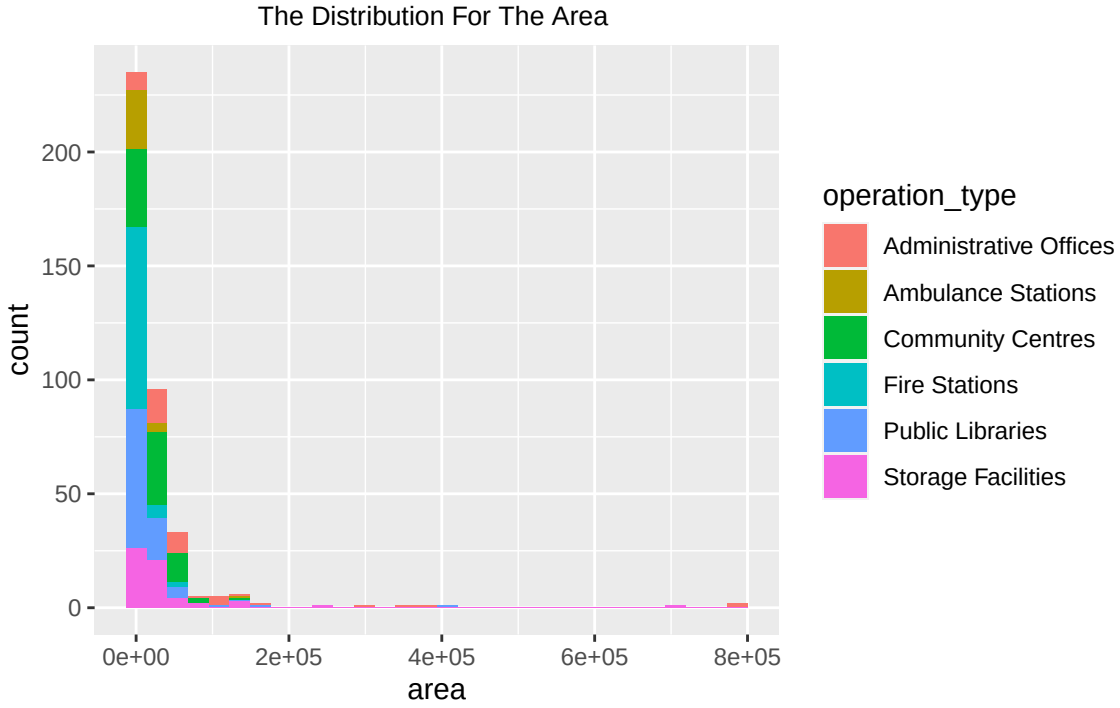

Figure 5:The Distribution For The Area

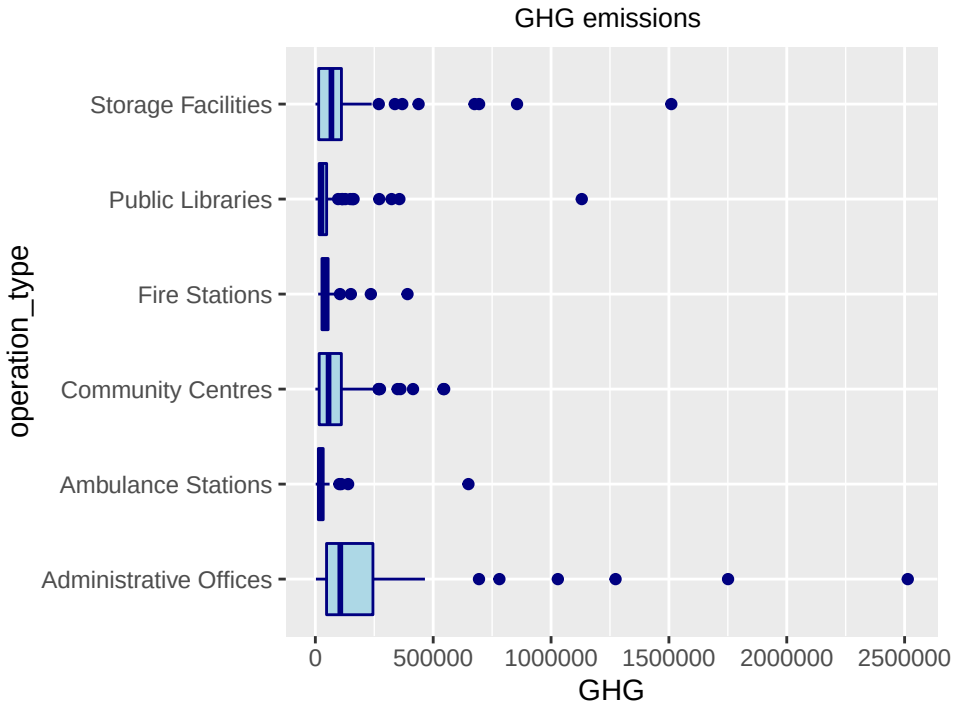

Figure 6:GHG distribution for each oepration type

First, let's look at the mean value of GHG in Figure 6. Although the ambulance stations have few outliers that have extremely high GHG, the mean value of ambulance stations is the lowest. In contrast, the

administrative offices have the highest mean value. As for the fire station, public libraries, and ambulance stations, they distribute closely. The distribution of GHG of public libraries and fire stations is left-skewed; the distribution of GHG of ambulance stations is symmetric. Notice that the Administrative Office has the widest distribution and has many outliers. Some administrative offices have less than 5000 GHG emissions, but some administrative offices still produce more than 1500000 GHG each year.If we look back to the Figure 5 for administrative offices, we can find explanation for this situation. In Figure5, we can find out that some administrative offices have quite large area. Since greater area has greater GHG emissions as shown in Figure5, those administrative offices with large area"s GHG emissions become the outliers in Figure6.
Overall, we can conclude the administrative office normally have high GHG emissions, and the ambulance station produces the smallest amount of GHG. And I would say area plays an very important role in operation type's influence on GHG emissions. Comparing Figure 5 and Figure 6 carefully, you will find that those two figures matched to some extent. The operation type which have smaller area on average in Figure 5 ends with smaller mean GHG emissions in Figure 6. Similarly, the operation type which have greater area on average in Figure 5 ends with larger mean GHG emissions in Figure 6.

4 Conclusion

In my assumption, I list out five factors; Usage of electricity and gas, the area, the average opening hours, and the operation type. I think these factors will have an impact on GHG emissions. However, after analyzing the data through the figures, I found out only the usage of electricity and gas, the area, and the operation type are clearly related to the levels of GHG emissions. What's need to notice that is, some part of the operation types' influence on GHG emissions are due to the variation in their area. And the reason why greater area will have greater is mostly because greater area will need to consume more energy. Thus, the key factor which impact the GHG emissions is the consumption of energy according to our report, this confirms the result by reported by the Prarie Climate Center in 2018 ("Where Do Canada's Greenhouse Gas Emissions Come From?" 2018)as I mentioned in the introduction.

In my opinion, controlling global GHG emissions has become urgent; the greenhouse effect caused by GHG has seriously endangered our environment. As a university student, I hope my paper can help more people to understand the main reasons that affect GHG emissions. This report also tells us that we should try our best to reduce electricity and natural gas in our daily life. By reducing GHG emissions, we can saveour world and ourselves.

References

"Annual Energy Consumption Data-2018." 2018. *City of Toronto Open Data Portal*. Environment & Energy. https://open.toronto.ca/dataset/annual-energy-consumption/.

"Climate Change Indicators: Greenhouse Gases." 2021. *EPA*. Environmental Protection Agency. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

"Greenhouse-Gas Emissions Killing Canadians: Medical Journal." 2018. *CTVNews*. The Canadian Press. https://www.ctvnews.ca/health/greenhouse-gas-emissions-killing-canadians-medical-journal-1.4196858.

"Law Document English View." 2018. *Ontario.ca*. Government of Canada. https://www.ontario.ca/laws/regulation/r11397.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"Where Do Canada's Greenhouse Gas Emissions Come From?" 2018. *From Risk to Resilience*. Prairie Climate Center. https://prairieclimatecentre.ca/2018/03/where-do-canadas-greenhouse-gas-emissions-come-from/.

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* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown.