Analysis of Greenhouse Gas Emission from 2011 to 2018 in Toronto*

William Gerecke

06 February 2022

Abstract

Greenhouse gas emissions have a significant impact on the global climate and are a key factor in climate change. This paper aims to examine the greenhouse gas emissions in Toronto and their change between the years 2011 and 2018. I examine the data and find that the total reported emissions in Toronto are increasing yearly, but on average public agencies are reporting decreases in their greenhouse gas emissions.

Contents

*Code can be found at https://github.com/wlgfour/toronto-energy-consumption.git

1 Introduction

The world is starting to experience accelerating effects of climate change, such as melting ice caps, rising sea levels, global temperature increases, and a higher frequency of extreme weather events. Preventing climate change has become a major global issue with significant debate around steps that can or should be taken to maintain a habitable Earth. The Paris Agreement is an international treaty on climate change that created a framework for the economic and social changes required to minimize the effects of climate change (Parties to the United Nations Framework Convention on Climate Change 2015). Part of the Paris Agreement addressed greenhouse gasses (GHGs), which are strongly linked to global climate change.

In 2009, the city of Toronto published the Green Energy Act (Ontario 2009) which was taken into force in 2011. The Green energy act requires public agencies (or "operations") to report their GHG emissions every year to the city of Toronto. In this paper, I explore GHG emissions in Toronto between the years 2011 and 2018, and observe that the total amount of recorded year-to-year GHG emissions increases. However, I also find that individual operations have recorded yearly decreases in emissions on average between the years of 2012 and 2015, followed by a plateau in the rate of reduction.

In this paper, I discuss the data in Section 2 including the software used (Section 2.1), source of the data (Section 2.2), and methods used to clean the data (Section 2.3). In Section 4, I comment on the data (Section 4.1), discuss the broader impact of GHGs (Section 4.2), and discuss next steps (Section 4.3). In Appendix A, I note all R packages that were used to generate this report.

In Section 4, I suggest that stricter regulations on reporting would lead to more comprehensive results regarding GHG emissions in Toronto. Beyond that, data for GHG emissions in Toronto is only available through 2018 at the time of writing this paper, but the Paris Agreement came into force in 2016. In the future it will be important to assess the continued impact of the Paris Agreement on reported GHG emissions in Toronto when the data becomes available.

2 Data

2.1 Software

This project uses the statistical programming language R (R Core Team 2021) to generate graphics and process data. The data is managed using `tidyverse` and processing is done with significant use of `dplyr` verbs (Wickham et al. 2019, 2021). `ggplot2` and `patchwork` are used to generate graphics (Wickham 2016; Pedersen 2020). For geometry-specific processing, the `sf` package is used (Pebesma 2018). The data is downloaded from the Open Data Toronto Portal using the `opendatatoronto` package (Gelfand 2020). The `tidygeocoder` package is used to geocode addresses present in the downloaded dataset (Cambon et al. 2021). For a complete list of software that was used for this project, see Appendix A.

2.2 Data Source

This project used two datasets of interest, both of which come from the Open Data Toronto Portal. The first is the Annual Energy Consumption dataset, which contains columns for the energy consumption of individual buildings in Toronto that are required by Ontario Regulation 397/11, the Green Energy Act (2009) (Ontario 2009), to report their GHG emissions. Specifically, this dataset contains annual `xlsx` spreadsheets from 2011-2018 that have columns with the building: name; address; floor area; mega liters of water or sewage treated; amount of energy purchased in the form of electricity and natural gas; and GHG emissions. As seen in Figure 1, there are significantly more reports in and after the year 2014 and the range of values reported is fairly large with values for GHG emissions somewhat evenly distributed on the logarithmic scale, but more concentrated at the upper end. The energy consumption, floor area, and GHG emissions are plotted on a logarithmic scale because these features vary most significantly in their order of magnitude.

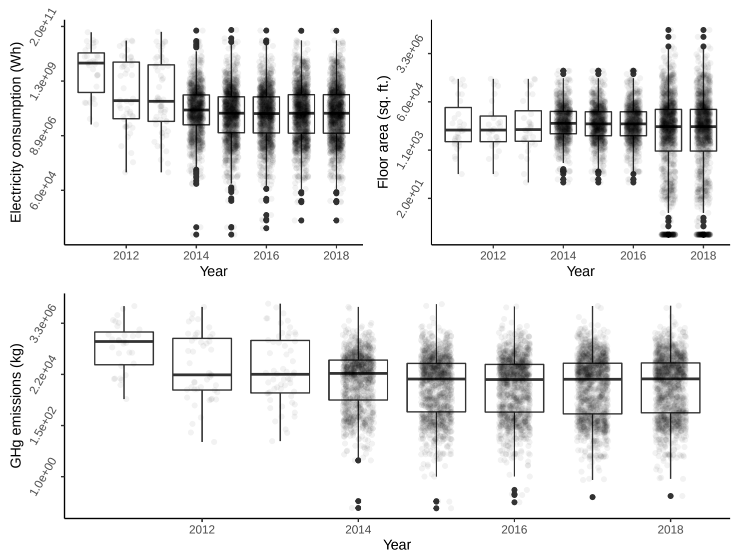

Figure 1: The figure shows reports for GHG emissions, electricity consumption, and floor area by year on a logarithmic scale. Each dot represents an operation that reported in the year specified on the x-axis.

Table 1: The table shows the number of successful and unsuccessful gecoded addresses for each year.

Geocode status	2011	2012	2013	2014	2015	2016	2017	2018	Total
failure	19	125	46	52	93	95	154	154	**738**
success	492	858	986	1,021	1,210	1,208	1,328	1,328	**8,431**
Total	511	983	1,032	1,073	1,303	1,303	1,482	1,482	**9,169**

The second dataset is called City Wards and it contains a shape file (`.shp` extension) that describes the 25 municipal wards of Toronto. The data contains the name of each ward as well as its geospatial information. The geospatial information was used to allocate a ward for each operation in the dataset.

In order to identify which ward each building belongs to, it was necessary to acquire the building coordinates (latitude and longitude). The coordinates were obtained by using the `tidygeocoder` package. The `tidygeocoder` package is free and open source, but the geocoding API has a rate limit of one request per second.

2.3 Data Cleaning

The Annual Energy Consumption dataset is provided in the form of several `xlsx` spreadsheets with a preamble at the top and grouped cells that describe column labels. The top rows (including column labels) were discarded and columns were manually assigned the correct labels. Additionally, there was one file containing spreadsheets for the years 2011-2014 which had to be separated. There were several paired columns where one contained a number and the other contained a unit. These were appropriately converted to the same unit using standard unit conversions. Finally, some postal codes contained a space in the middle (e.g. `M4Y 0A9`) and some did not. For the purpose of geocoding, these had to be converted to the format with a space[1].

In order to identify which wards each building belongs to, they were geocoded by address, postal code, city, and country using OpenStreetMap (OpenStreetMap contributors 2017). There were 2404 unique addresses,

[1]Regex: `[A-Z][0-9][A-Z] [0-9][A-Z][0-9]`

738 of which could not be geocoded for various reasons, such as `'Various addresses'` in the address column. In order to reduce the strain on the Nomination API, only unique addresses were geocoded which reduced the number of necessary calls to the API by 6765. The number of yearly reports and a summary of the geocoding process can be seen in Table 1.

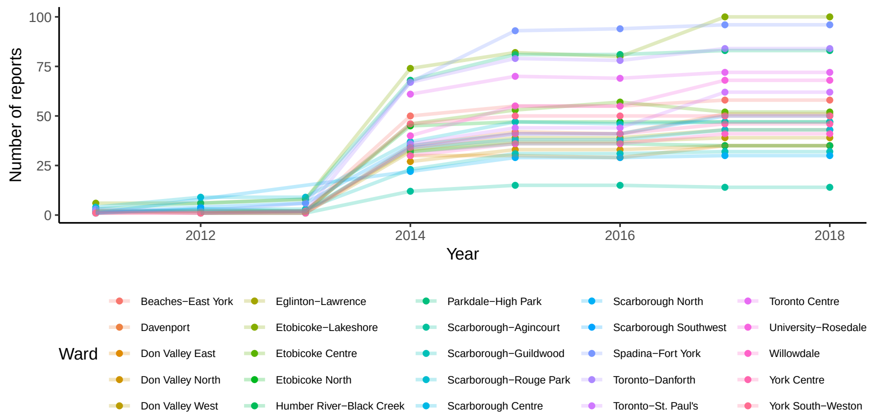

Figure 2: The figure shows the number of reports each year for each ward. There is a significant increase in reports across all wards between 2013 and 2014.

Each building in the dataset was assigned a ward based on which ward's boundaries the geocoded point for that building intersected with. At the end of the cleaning process, excess columns were dropped, rows with invalid data were dropped, and the dataset was saved. The number of buildings that submitted reports in each ward year-over-year can be seen in Figure 2. In 2014 nearly all wards experienced a more-than-doubling of the number of reports received. Such a significant change in reporting is likely not due to an equally significant increase in the number of public agencies in Toronto, but potentially a phased-in requirement for public agencies to submit these reports[2], but there are also other reasons that this could have happened.

3 Results

3.1 Yearly emissions

In Figure 3, we can see that the total reported GHG emissions increases year after year with only some small decreases when the change in number of reports is low. Even between 2017 and 2018 where the number of reports does not change, the total reported amount of GHG emissions increases. It is likely that this is due to the large increase in the number of reports submitted by public agencies in Toronto rather than a change in underlying GHG emissions.

In order to get a clearer sense of the GHG emission trends in Toronto, we can look at the change in GHG emissions reported by each individual operation by looking at ΔGHG_y, or the amount that an operation's GHG emissions have changed in year y as seen in Equation (1) where GHG_y indicates an operation's GHG emissions in year y. If ΔGHG_y is aggregated over ward, as seen in Figure 4, we can see that operations who reported in 2011 and 2012 reported an increase in the total amount of GHG emissions released, but operations who reported from 2012 and 2014 saw a significant decrease in total GHG emissions, with a small rebound in 2015. Other than this, though, there are no significant trends in the total amount of change in GHG emissions reported per ward. That is, the amount of GHG emissions that operations in Toronto reported between the years 2016 and 2018 were relatively stable.

[2]This is not mentioned in the actual Green Energy Act, but it is possible that there is an arbitrary grace-period for Toronto regulations.

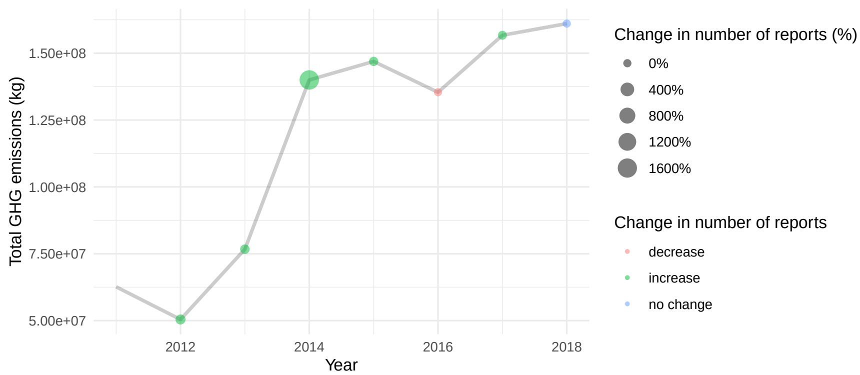

Figure 3: The figure shows the total reported GHG emissions by year. Point size represents the increase in the number of reports, and color indicates an increase, decrease, or no change with respect to the number of reports.

$$\Delta GHG_y = GHG_y - GHG_{y-1} \tag{1}$$

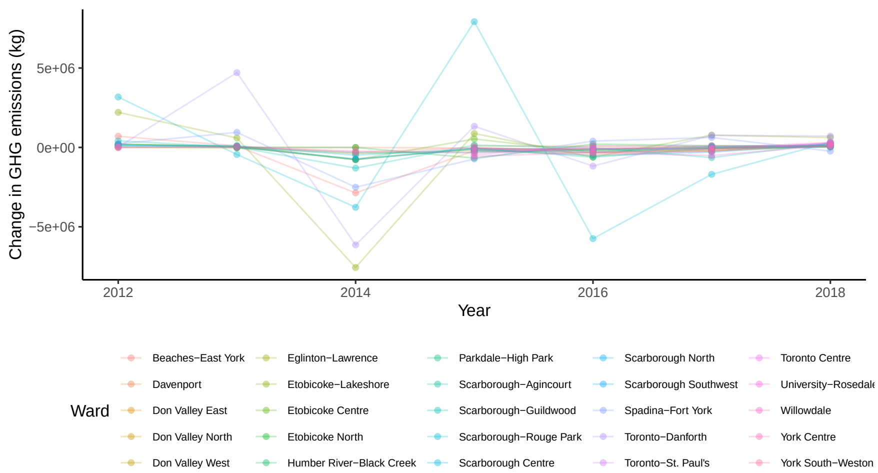

Figure 4: In the figure, ΔGHG_y is aggregated for each operation by ward and plotted against year. The figure gives insight into the change in total GHG emissions in each ward by year.

In Figure 4, we can see total yearly change in GHG emissions which is important to the overall environmental impact, but this still doesn't give much insight into the amount that individual operations are decreasing their GHG emissions. The problem is that the total amount of GHG emissions in a ward could easily be swayed by large operations significantly decreasing emissions. For example, Figure 1 indicates that there are some operations with at least 3 to 4 orders of magnitude more GHG emissions than other operations. If one of the larger operations went out of business, but the smaller operations on average increased their GHG emissions, it could easily appear as an overall decrease in emissions. This is bad, however, because there could be a significant number of operations increasing their environmental footprint rather than working to decrease it. Similarly, a large operation could make it seem that operations in a ward are increasing their emissions even if most of the operations are actually decreasing them.

In order to make it easier to examine the trends in GHG emissions from all operations, including smaller operations, we can look at $\frac{\Delta GHG_y}{Fl_y}$ where Fl_y is the operation's reported floor area (square feet) in year y. Again in Figure 5, we see that on average most operations were increasing their emissions from 2011 to 2012 when mandatory reporting began, but quickly decreased GHG emissions in 2013 and 2014. Beyond 2015 though, there is very little change in emissions across all wards.

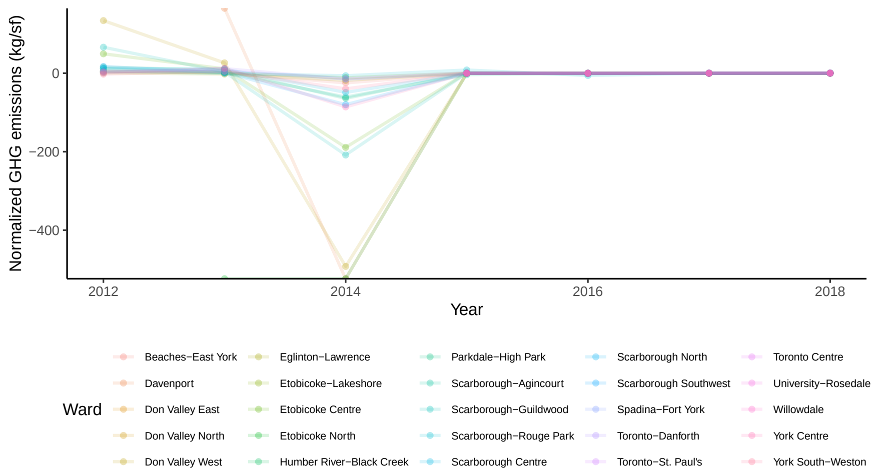

Figure 5: In the figure, ΔGHG_y is divided by the operation's floor area, aggregated for each operation by ward, and plotted against year. We can see more clearly the change in emissions for individual operations.

In Figure 3, we can see that year after year, there is a significant increase in GHG emissions in Toronto's wards. This is further supported by Figure 4 where we see that there was only a significant decrease in the amount of GHG emissions that each operation produced in 2014, and for the most part operations maintained constant levels of emissions. This means that any growth in the amount of operations in Toronto will necessarily translate to an increase in total GHG levels. Considering that the total GHG emissions in each ward could be dominated by large operations, I consider a standardized version of ΔGHG_y, which is plotted in Figure 5. Here we only see a change in GHG levels in the early 2010s, but this trend is suppressed by the increase in operations who logged GHG emissions in 2014 (Table 1). While the operations that decreased emissions from 2012 to 2014 may have continued to do so, the overall trend in the mid to late 2010s was that the GHG emissions of each operation stabilized. While this means that individual operations were not increasing their GHG emissions, it is important to note that this means that operations were not improving their emission levels.

3.2 Cumulative emissions

While it is important to understand the yearly change in GHG emissions in order to focus on progressing towards clean and sustainable practices, it is also relevant to examine how much progress has been made so far, as well as the impact on the environment that has already been inflicted. One of the reasons that this is important is that policy makers can adjust and form new plans with the purpose of reducing GHG emissions by looking at the total reduction in emissions so far. This also allows researchers to forecast the effects of GHG emissions on the plant and its population by looking at the cumulative release of GHGs into the atmosphere.

Figure 6 illustrates the cumulative reported GHG emissions from 2011 to 2018 in Toronto. Subplot A makes the compounding nature of the information shown in Figure 3 abundantly clear. That is, even seemingly small or large, but individual increases in emissions lead to significant increases in the total amount of emissions when compounded over several years. We can see this concretely in the 426% increase in emissions between 2011 and 2014, and another 182% increase in emissions between 2014 and 2018. While both increases are large, they are larger when compounded together as the increase between 2011 and 2018 is 1384%. We notice this significant change in emissions despite almost no change in operations' GHG emissions between 2014 and 2018 as seen in Figures 4 and 5, as well as very little change in yearly emissions as seen Figure 3.

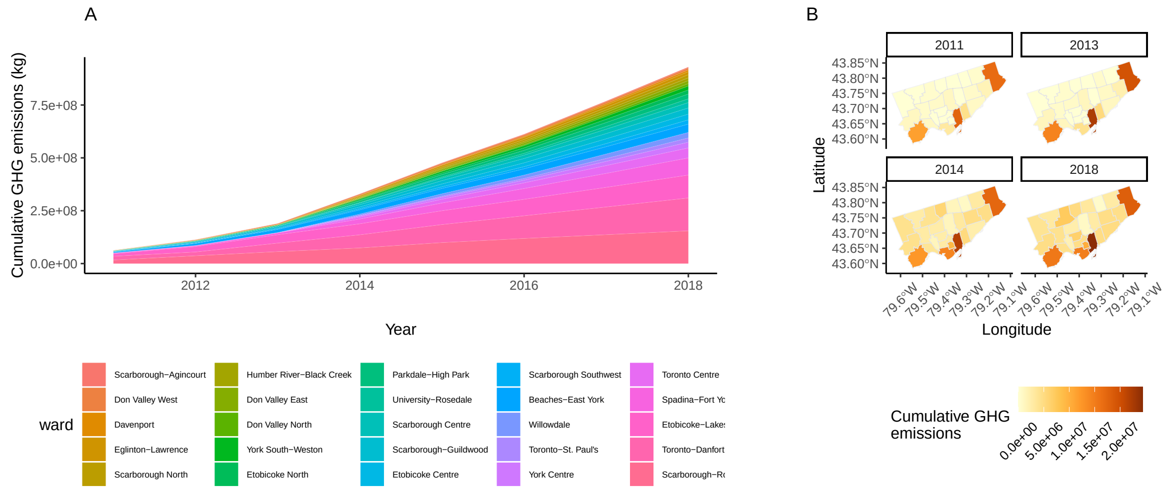

Figure 6: In Subplot A we can see the cumulative emissions for each ward stacked, so the total cumulative emissions are shown. In Subplot B we can see the cumulative emissions for each ward shown on a choropleth.

Figure 6.B shows the cumulative emissions in each ward starting at 2011 through to 2018. We can see that some wards have a significantly larger concentration of GHG emissions than other wards, which could indicate the need for research into the health and environmental impacts of GHGs in these wards. The effects of GHGs is at greater length in Section 4.

We can also look at the cumulative change in GHG emissions to assess the total improvement by ward. Figure 7 shows the cumulative change in GHG emissions as a percentage of that year's emissions. This way, we can see how much a ward has decreased their emissions relative to the total amount of emissions that year. For example, we can see that by 2018, operations in Scarborough-Agincourt had decreased their emissions by about as many kilograms of GHGs as they released in 2018. In other words, if operations in Scarborough-Agincourt had not decreased emissions between 2011 and 2018, Scarborough-Agincourt would have released at least twice as many kilograms of GHGs in 2018 as it did. Overall, we can see that after 2014, the amount that operations were improving their emissions could not catch up to the rate that new operations were reporting emissions, however, there was still a small net improvement in GHG emissions.

Between Figures 6 and 7, we can see that despite the fact that reported GHG emissions in Toronto have increased at an increasing rate between 2011 and 2018, there is still a net improvement in emissions on an operation-specific basis. If we make the naive and hopeful assumption that no new operations have been created and the only reason new operations are appearing is because they had not logged emissions in previous years, this actually indicates a decrease in the underlying emissions in Toronto. This is, however, most likely not strictly true, but the question remains: what is the net change in underlying emissions in Toronto?

With more strict reporting policies, the government of Toronto could require all operations to report emissions which would allow us to directly calculate the net improvement in emissions. As of now, with the Annual Energy Consumption dataset, we are left with the information that net reported GHG emissions have increased consistently (Figure 3), while operations who reported their emissions more than once have on average decreased their emissions (Figure 7). More simply: the total reported emissions are still getting worse, but on average operations are reporting improvements in their emissions.

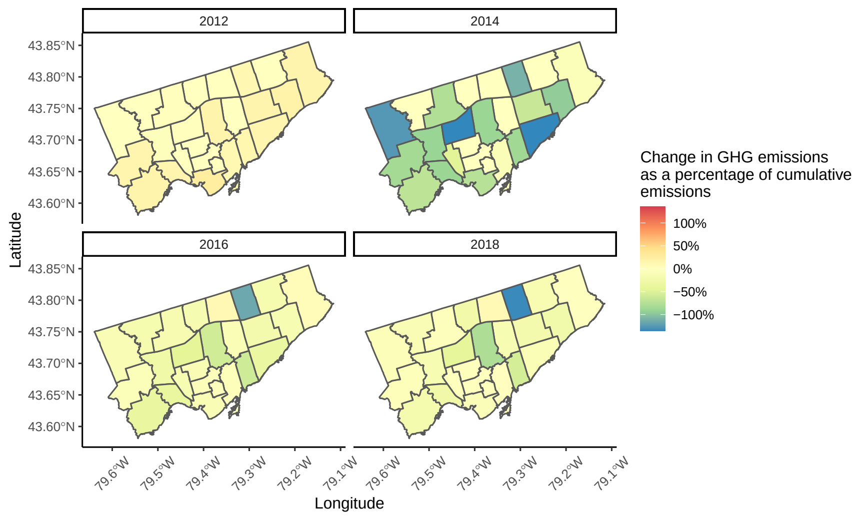

Figure 7: The figure shows a coropleth of the cumulative change in GHG emissions as a percentage of the cumulative emissions in that year. This shows how much each ward has improved their emissions relative to their emissions in the year of interest.

4 Discussion

4.1 Comments on the data

During the data cleaning process, there were many observations that I had to exclude. Some of these were due to the data points being out of the bounds that were being considered (Toronto's 25 wards). The next most significant reason that some values were dropped is that they had invalid data (ex: `N/A` under the floor area column). This could have introduced bias into the results by way of systematically excluding operations where certain fields do not apply, or operations that made the same error when recording their emissions each year. The remaining values that were dropped where values that could not be geocoded. Without having looked closely at the similarities within this subset of observations, this could have caused bias by excluding certain neighborhoods, or streets. There were also operations that submitted `Various Locations` as their address (ex: `WPC Service Chambers`), which excluded that operation from analysis in every year that they reported that address.

Beyond dropped values, the Green Energy Act only requires operations that meet certain requirements to report their results, but allows any operation to report results. This means that the dataset is inherently biased. For example, it is possible that a significant amount of the observations come from operations who are not required to report emissions, but do so because they have low emissions, while operations with higher amounts of emissions refrain from reporting. It is also possible that the operations who are required to report are under different legal restrictions than operations that are not required to report. This means that the population of operations which are required to report is completely different from that of the operations which are not required to report emissions.

4.2 Broader impact of GHG emissions

Beyond the immediate concerns of energy consumption and resource allocation, the operations described by the Green Energy Act have a significant environmental footprint by way of GHG emissions among other factors. GHG emissions specifically have been shown to have significant ramifications for the economy, agriculture, the environment, and the health of people and animals.

GHG emissions have been monitored as early as the 1990s and a study by Darwin (2004) looked at the effects of GHGs on the economy and its relationship to global agriculture. Darwin observed that various factors such as CO_2 fertilization and the relationship between GHGs and global temperature increase lead to increased crop production. This, in turn, leads to a positive effect on global economic conditions.

Other studies, however, have indicated that increased amounts of GHGs in the atmosphere lead to more negative outcomes. Xu et al. (2020) suggested that the significant increase in wild fires across the globe are synergistic with the increase in global temperature and GHGs. Xu et al. (2020) also examined the effects of GHGs and wild fires on human health, observing short-term as well as long-term effects on health such as decreased lung capacity, and increased risk for disease or death.

While there are many different effects of GHGs in the atmosphere that can have positive and negative impacts, global efforts have been taken to reduce the amount of GHG emissions in order to prevent catastrophic climate change. Mikhaylov et al. (2020) used sensitivity analysis to support the claim that "anthropogenic CO2 emissions from people (cars, households) would deescalate the consequences of [GHG emissions]," but "emissions are mostly associated with industries, which can be reduced if local Government will want to achieve the Paris Agreement goal." Toronto has taken steps to monitor and reduce GHG emissions with the Green Energy Act (Ontario 2009), but in recent years the data suggests that the reduction in emissions has stagnated.

4.3 Next Steps

In accordance with the Paris Agreement, it is crucial that global GHG emissions be reduced by approximately 50% before 2030 in order to combat and minimize the damages caused by climate change (Parties to the United Nations Framework Convention on Climate Change 2015). As such, continuing to monitor and find ways to reduce GHG emissions is paramount to keeping the global temperature increase below 2°C. In the coming years it will be crucial to monitor the effects of the Paris Agreement on GHG emissions in all parts of the world, especially centers for industry such as Toronto (Mikhaylov et al. 2020).

5 References

Cambon, Jesse, Diego Hernangómez, Christopher Belanger, and Daniel Possenriede. 2021. "Tidygeocoder: An r Package for Geocoding." *Journal of Open Source Software* 6 (65): 3544. https://doi.org/10.21105/joss.03544.

Darwin, Roy. 2004. "Effects of Greenhouse Gas Emissions on World Agriculture, Food Consumption, and Economic Welfare." *Climatic Change* 66 (1/2): 191–238. https://doi.org/10.1023/b:clim.0000043138.67784.27.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

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

Mikhaylov, Alexey, Nikita Moiseev, Kirill Aleshin, and Thomas Burkhardt. 2020. "Global Climate Change and Greenhouse Effect." *Entrepreneurship and Sustainability Issues* 7 (4): 2897.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

Ontario, Government of. 2009. *O. Reg. 397/11: ENERGY CONSERVATION AND DEMAND MANAGEMENT PLANS*.

OpenStreetMap contributors. 2017. "Planet dump retrieved from https://planet.osm.org ." https://www.openstreetmap.org.

Parties to the United Nations Framework Convention on Climate Change, Conference of the. 2015. "Paris Agreement." In *United Nations Treaty Series*, 7.d.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

———. 2021. *Forcats: Tools for Working with Categorical Variables (Factors)*. https://CRAN.R-project.org/package=forcats.

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.

Xu, Rongbin, Pei Yu, Michael J. Abramson, Fay H. Johnston, Jonathan M. Samet, Michelle L. Bell, Andy Haines, Kristie L. Ebi, Shanshan Li, and Yuming Guo. 2020. "Wildfires, Global Climate Change, and Human Health." *New England Journal of Medicine* 383 (22): 2173–81. https://doi.org/10.1056/nejmsr2028985.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.

A Packages

`knitr` and `bookdown` were used to generate the pdf output of this report (Xie 2014, 2016).

`dplyr` was used heavily to process dataframes using functions such as `mutate`, `summarise`, and `filter` (Wickham et al. 2021).

Functions and packages from `tidyverse` were used throughout the code (Wickham et al. 2019).

`tidygeocoder` was used to geocode addresses so that they could be allocated to a ward in Toronto (Cambon et al. 2021).

`opendatatoronto` was used to obtain both of the datasets of interest in this paper (Gelfand 2020).

`janitor` was used in preprocessing of the data to auto-format column names and some values (Firke 2021).

`reshape2` was used to reshape data for Table 1 (Wickham 2007).

`kableExtra` was used to format Table 1 as well as add the marginal distributions (Zhu 2021).

`sf` was used to read shapefiles for Toronto wards, to intersect ward geometries with address coordinates, as well as to plot the coropleths.

`ggplot2`, `patchwork`, and `RColorBrewer` were used to generate the plots (Wickham 2016; Pedersen 2020; Neuwirth 2014).

`forcats` was used to change the order of `levels` of a dataframe in a functional manner (Wickham 2021).

`stringr` were used in the data cleaning process for string operations with greater efficiency (Wickham 2019).